iHub Drishti: TIH on CV, AR and VR

"Trusted Vision for Secure and Better Living"

—

Indian Institute of Technology Jodhpur

Executive Summary

The TIH on Computer Vision and Augmented and Virtual Reality (CV and ARVR), named as iHub Drishti, will: (i) focus on advancing the research outcomes in core problems related to CV and, ARVR, (ii) augment imaging with additional (multimodal) sources of input such as haptics, language, and IoT to advance state-of-the-art in the domain areas, (iii) create technology solutions for socially relevant and industry-facing problems, (iv) support and nurture start-up ecosystems, (v) stimulate skilling and reskilling educational programs, and (v) to advise governments for appropriate policy-related matters in the domain of CV and ARVR.

The proposed TIH will focus on the core research areas of (i) Seeing and Sensing, (ii) Dependability, (iii) Real-time Computer Vision Systems, and (iv) Data Collection, Curation and Annotation for developing technologies in the following applications areas:

1. Computer Vision for Autonomous Systems
2. Computer Vision for Better Living: Healthcare and Biosphere
3. Imaging for Document Analysis
4. CV and VR for Industry 4.0
5. Dependable AR-VR for X

The current autonomous systems lack human intelligence, robustness and flexibility and may not be suitable for doing many tasks that humans can do. The TIH will steer a series of multidisciplinary research initiatives to harness transformational power of technologies to address the grand problem of building vision and AR-VR based dependable autonomous systems. This will pave the way for building durable solutions to national needs in the area of defense and security, healthcare, education, and many other applications such as disaster management, transportation, mining, and handling hazardous environments.

iHub Drishti, a Section 8 company, is hosted in the premises of IIT Jodhpur. The hub, steered by the Hub Governing Board and the Director of the TIH, will host teams of faculty members, research scholars, developers and scientists who will work towards developing the technologies. Further, four different kinds of partners are identified: (i) research collaborators from national and international institutions, (ii) domain expert partner institutions, (iii) hub and spoke model (mentor-mentee) based partner institutions, and (iv) research labs and industry including startups. The TIH envisions creating an ecosystem in and around the thematic areas and building collaborations with other TIHs to foster the overall growth and development of CPS technologies. The facility would include a mix of office and laboratory spaces with state-of-the-art infrastructure facilities that will cater to the collaborating institutions. The TIH will also build a network of researchers and developers, both externally hired and from the student body, such that in the next 5 years, iHub Drishti becomes the main hub of technologies related to computer vision, virtual and augmented reality in the entire country.

The proposed TIH will initially focus on fundamental research, developing Intellectual Properties and building prototypes (TRL 3-4). Based on the technologies generated, these IPs will be translated into dependable products and processes in association with startups, industry and it's own team of executives. The experts, through iHub Drishti, will also provide consulting services to other companies and government institutions. It is envisioned that iHub Drishti will be self-sustaining by the end of five years with continuous growth in research, technology and product portfolio.

TIH will aim to solve the following challenging open problems:

1. Building dependable vision systems for navigation and monitoring in the application area of defense, security, rescue operation, reconstruction operation, nuclear power plant operation, under-sea exploration.
2. Building dependable computer vision systems for healthcare in the application area of disease surveillance, bio-safety of the home and work environments, remote handling of patients minimizing exposure of health workers.
3. Immersive and experiential technologies in the application area of AR, VR, MR for medical, heritage and haptics.
4. Dependable vision and AR-VR for Industry 4.0 in the application area of smart manufacturing.
5. Image and video analytics platform for supporting education and research in the application area of multi-modal, multi-lingual, multi-document image analytics for deciphering Gandhian literature.
6. Computer vision for sustainable biosphere in the application area of wildlife, agriculture and plant life.

CERTIFICATE

Name of the TIH: IHUB DRISHTI FOUNDATION

Technology Vertical: **Computer Vision, Augmented and virtual reality**

1. This is to certify that the Detailed Project Report (DPR) on the Technology Vertical **Computer Vision, Augmented and Virtual reality** is prepared and submitted to Mission Office, NM-ICPS, DST is as part of implementation of Technology Innovation Hub (TIH) at Indian Institute of Technology, Jodhpur, N.H. 62, Nagaur Road, Karwar, Jodhpur 342037, Rajasthan (India) under National Mission on Interdisciplinary Cyber-Physical System (NM-ICPS).

2. This is to certify that this DPR has been checked for plagiarism and the contents are original and not copied/taken from any one or from any other sources. If some content was taken from certain sources, it is duly acknowledged and referenced accordingly.

3. The DPR will be implemented as per the Terms, Reference and Clauses stated in Tripartite Agreement signed on 27.11.2020 between Mission Office, DST, Indian Institute of Technology Jodhpur and IHUB DRISHTI FOUNDATION.

Date: Sep 20/2021

Place: Jodhpur

For iHub Drishti Foundation

Name and Signature of Project Director

Director / Authorized Signatory

Endorsement from the Head of the Institution

1. Certified that the Institute welcomes participation of Prof. Mayank Vatsa as the Project Director for the Technology Innovation Hub (TIH) and that in the unforeseen event of discontinuance by the Project Director, the Indian Institute of Technology, Jodhpur will identify and place a suitable faculty as Project Director for fruitful completion of the TIH activities.

2. Certified that the Host Institute shall provide basic facilities, faculty support and such other administrative facilities as per Terms and Conditions of the award of TIH, will be extended to TIH.

3. As per Tri-partite Agreement, the Host Institute (HI) shall play its role and fulfill its responsibilities for the success of TIH.

Date: 20|9|2021

Place: Jodhpur

Name and signature of Head of Institution

Santanu Chaudhury
Director
Indian Institute of Technology Jodhpur

Introduction

Cyber-physical systems (CPS) sense, process, and react to information from the physical world. They are created by combining computation, networking, and physical processes with intelligence. Human visual system can be perceived as a perfect example of a CPS. It sees, processes, connects to other body parts, and acts based on the information analyzed. In the entire process, it also ensures that the human must operate safely even in the presence of uncertainties and resource constraints. The TIH at IIT Jodhpur, named as iHub Drishti, will focus on building cyber-physical systems for computer vision, augmented reality (AR) and virtual reality (VR). Our mission is to build technologies and systems focusing on **trusted vision for a secure and better living**.

Computer vision involves acquiring, processing, analyzing and understanding a given image or scene. The fundamental concept has been there for more than six decades now, however, the application in unconstrained real world problems has started more recently. The omnipresent sensors and cameras provide the opportunity to integrate the core vision technology with the network of sensors to propose solutions for a host of problems, ranging from defense to medical to education. The success of autonomous cars which operates as a very complex integration of a multitude of sensors with computer vision and artificial intelligence algorithms is an astounding example of a cyber-physical system and the advancement and accessibility it can provide. On the contrary, augmented reality and virtual reality are not more than two decades old. AR and VR provide an interactive experience of a real-world or a digital world environment. With increasing sophistication in the performance of AR and VR, they are becoming very popular and their applications are being explored in a number of industries, for instance, retail, education, defense training, gaming, and entertainment. In 2017, the global market for computer vision stood valued over USD 9.2 billion and is expected to surpass a valuation of USD 48.3 billion by 2023[1], The augmented reality market was valued at USD 4.21 billion in 2017 and is expected to reach USD 60.55 billion by 2023, at a CAGR of 40.29%, between 2018 and 2023[2]. Similarly, the virtual reality market was valued at USD 5.12 billion in 2017 and is expected to reach USD 34.08 billion by 2023, at a CAGR of 33.95%, between 2018 and 2023.

Mission statement of iHub Drishti is

To become **the most coveted self-sustaining technology destination in CV, AR and VR** in the country, with a nationwide footprint, nurturing and supporting **cutting edge research and innovative technology deliveries through Startups, MSME's and direct execution of projects.**

To achieve this aim,

Mandate of iHub Drishti will be:

- Promote innovation in computer vision, AR and VR,
- Use of technology foresight to predict and develop new technologies,
- Support and nurture start-up ecosystems,
- To create patent portfolios and licensing of technology,
- Create technology solutions for problems of the user agencies and socially relevant problems,

---

[1] https://www.marketresearchfuture.com/reports/computer-vision-market-5496

[2] https://www.marketsandmarkets.com/Market-Reports/computer-vision-market-186494767.html

- To advise governments for appropriate policy-related matters,
- Knowledge dissemination through workshops, conferences, summer and winter schools,
- Participate and bid for sponsored research projects from governments and Industries,
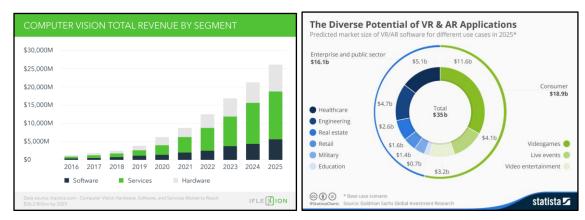- To become self-sustainable.

Figure: Projections of computer vision and virtual and augmented reality industry.

Horizontals and Verticals

iHub Drishti – the multidisciplinary technology innovation hub (TIH) of IIT Jodhpur, under the aegis of the Department of Science and Technology, Government of India, will be designing 'dependable' cyber-physical systems – focusing on the application of computer vision and AR-VR. There can be an innumerable amount of research problems and applications that can be pursued at iHub Drishti. The activities under iHub Drishti can, therefore, be visualized in the form of a 2D model of horizontals and verticals. **Verticals are the focus application areas whereas horizontals are the core technologies that cut across all or some of the verticals.** The list of horizontals and verticals of the hub are as follows.

Horizontals - Core technologies:

- Seeing and Sensing
- Dependable CV & AR,VR
- Real-time Computer Vision Systems
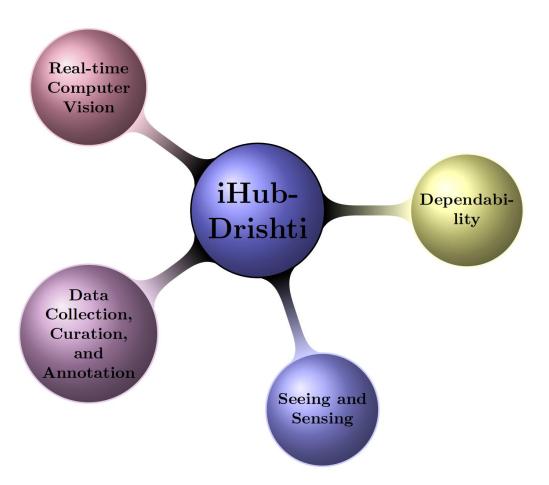- Data Collection, Curation, and Annotation

Verticals - Focus application areas:

- Vision for Autonomous Systems
- Computer Vision for Better Living: Healthcare & Biosphere
- Imaging for Document Analysis
- Computer Vision and Virtual Reality for Industry 4.0
- Dependable AR & VR for X

The Horizontals

Seeing and Sensing

Real-time Computer Vision

Dependability

Data Collection, Annotation, and Curation

Seeing and Sensing

Introduction and Motivation

Computer Vision is a vital enabling technology having applications in almost all the domains including autonomous systems, biodiversity, health care, document analysis, industry 4.0, augmented and virtual reality and others. All these domains require vision to see, therefore, cameras are the most crucial component for these systems. Further, the application of cameras or image sensors in these systems are manifold. The cameras can be used in security systems to detect enemy troops and enhance the targeting capabilities of guided missile systems. Military concepts such as situational awareness rely heavily on image sensors to deliver battlefield intelligence used for tactical decision-making. Another key Computer Vision application is the areas of autonomous vehicles, which need to navigate challenging terrain and detect adversaries. Computer Vision systems also support human drivers and pilots, enabling them to evade enemy fire. As is the case with many military innovations, technology transfer is now benefiting a broad range of industries. Further, owing to the increasing security concerns in the country and worldwide, governments are continuously increasing the area under surveillance.

Consider the problem of mass screening and early diagnosis of pandemic diseases such as Ebola outbreak or COVID-19. Fast and accurate mass screening of people for these diseases is a very critical step to be carried out in order to contain and prevent the deadly spread across the globe. The major challenge in screening the people infected with such outbreaks or pandemics is the transmission of the virus from the infected people to the staff while performing the screening. Therefore, a non-contact screening has become an important problem to solve. The infections caused by such viruses lead to various early symptoms in the suspected peoples which plays a major role in screening and early diagnosis of the infected people. For example, very early symptoms of the COVID-19 suspects are the raised body temperature, decreased oxygen saturation levels (SpO2) in the blood, etc. Now, to perform a screening of COVID-19 based on these symptoms, the first problem to solve is measuring such symptoms through non-contact devices. Vision-based sensors such as thermal cameras can be used to sense the body temperature and the images or videos acquired from the RGB or multispectral cameras can be used to sense the level of saturation of oxygen in the blood. The vision-based sensors (Multi-wavelength cameras, Thermal Cameras) plays a very critical role in sensing and measuring various physiological measurements of the human body, such as pulse rate, pulse transit time, ballistocardiogram motion, respiration rate, PPG signals, Cardiovascular age, and pain, for screening and diagnosis of various deadly diseases.

Consider the example of defense applications in harsh, unstructured, and dynamic environments. For example, the uncertain presence of dust storms in the deserts and ice storms in mountains, or fog in the forests, prevent the military from performing operations. The traditional RGB cameras do not see through such harsh conditions. The multispectral cameras are used for sensing the activities through such conditions. A very important aspect of the perception-action cycle from the defense perspective is that the perception-action cycle needs to be completed almost in real-time. Almost all of the tasks in the defense applications need to be solved in real-time for effective outcomes. For example, an autonomous system needs to reconstruct and recognize (whether it is an enemy object or not) a high-speed enemy missile and track it and perform an action before it hits the target. Therefore, the vision-based

sensors (asynchronous sensors) are required to sense the rapid activities quickly for performing actions. The problem of educational or military training and simulation, e-tourism, telesurgery, and teleoperations, using AR/VR techniques require the efficient sensing and reconstruction of the 3D models of the surrounding environments and the scene objects. Furthermore, the navigation of autonomous vehicles completely depends on the reconstruction of the 3D models of the surrounding.

Obtaining the appropriate signal of the scene is a key part of the applications which process this information to infer the various relevant aspects of the scene. But, the information may not be directly available in the presence of various types of occlusion in many scenarios. This determines the scope of the proposed approach for a given application as without information no processing can be performed. However, this information can be obtained from indirect means by utilizing the multimodal nature of the scene. We have classified these imaging mediums/occlusion into three categories i.e. Scattering, Reflecting and Refracting based on the nature of the interaction of visible range EM spectrum with them. The Scattering Medium or occlusion refers to those occlusion which causes the scattering of the EM spectrum utilized for the imaging, Haze, Fog, Water, Gases etc. are some commonly encountered examples of this type of occlusion. The Absorbing and Reflecting Medium causes loss of the reflected transmission of the EM spectrum. Walls, Sand Storm, Dust, Cloud, Metal, Bones etc. are some commonly encountered examples of this type of occlusions. The Refractive Medium causes the refraction of the EM spectrum utilized for the imaging. This creates distortion in the depth which can be uniform or non-uniform due to variation in refractive index based upon the homogeneity of the medium/occlusion. Mirages are some common examples of this type of occlusion.

These types of occlusions occur in various domains of imaging. Medical Imaging is one of the major application areas of such an imaging framework involving scattering, reflecting types of occlusions and hence, utilizes different imaging modalities and framework to capture the desired signal based on size of imaging target, its interaction with the selected spectrum and the medium. Similarly, seeing through haze, rain, dust, underwater, walls etc. are important aspects of the vision guided navigation system. Such techniques can also be employed for the purpose of the scanning, monitoring of the processes, infrastructure and the key equipment in case of Industry 4.0. This imaging framework can also be utilized in various applications of AR-VR such as telemedicine by employing the obtained information. Thus, seeing through occlusion is a fundamental imaging technique which has to be utilized for various applications across the all vertices of the TIH iHub Drishti.

Objectives

Objectives of the horizontal sensing and seeing are as follows:

1. Development of computational systems for capability enhancement of targeted imaging modalities
2. Design of advanced image based dependable diagnostic and prognostic tools
3. Development of computational techniques and image based analysis tools
4. Creating an AI based platform for responding to national emergencies

5. To perform cutting edge research and generate new knowledge through research and development for evolving vision based technologies.

6. The main objective of seeing through occlusion is to provide the optimal signal acquisition framework based on the accuracy, cost, processing time etc. requirements of the applications and possible scopes. These applications can belong to a variety of the areas such as Healthcare, Navigation, Industry 4.0, AR-VR etc.

Research Questions

Research on solving classical computer vision applications using the visual data acquired using the asynchronous sensing is still in its infant stage and mostly done in the lab scenarios. However, real life problems, for example 3D reconstruction, recognition and navigation, using the asynchronous are still open and require a lot of effort. Furthermore, the existing navigation solutions using the asynchronous sensors are proposed under the very controlled environments (such as indoor scenes where the surface and objects are rigid in nature) [Mueggler et al. 2018, Rebecq et al. 2016, Kim et al. 2016].

Another research question is related to the design of high dynamic range CMOS image sensors. The main challenge in high dynamic range sensors is noise reduction which is mainly limited by the noise floor of readout circuitry. This work is mainly focused on reduction in noise floor of the readout chain along with increment in signal at the pixel level. The dynamic range will be enhanced making the camera suitable for low light imaging applications.

Sensor fusion is another line of thought where the outputs of multiple sensors are used for improving the decision accuracy. Present autonomous systems incorporate many sensors including Camera, LiDAR, RADAR, Inertial sensors, GPS etc. However, the accuracy of all the sensors are limited and there is an equivalent probability of generating wrong decisions. In these systems, the entire vehicle control is governed by the sensors itself, therefore, the functionality is governed by multiple sensors altogether.

Similarly, accurate complex scene recovery of the scene contains objects of different reflecting properties. Currently, state of the art methods are focused upon detection of only a single object behind the occluder which is not a practical scenario. Presence of multiple objects results in mixing of various signals. Utilization of multiple sensors to capture the different modalities of the scene and characterizing them using a deep-learning network can be useful in such cases.

Contributions

1. An adaptive real-time method for image enhancement in scattering mediums such as water, haze, fog, smoke etc. with respect to variation in density which is robust against non-uniform atmospheric scattering, varying illumination conditions including saturations and high-speed operations.

2. A robust real-time NLOS imaging system utilizing conventional sensors under deep-learning framework which can be utilized for real-world applications such as imaging of a general scene under sand, wall etc

3. A general robust detection and correction system for the mirage, refraction under varying optical properties.

4. Design of high dynamic range and high frame rate CIS with inbuilt FPN cancellation for defense and autonomous systems.

5. Vision based autonomous systems with asynchronous vision sensors for defence application with the capabilities of automatic 3D reconstruction of the territories, obstacle avoidance, object detection.

6. Deep learning based approach for handling the data acquired through the event-based cameras and solving the problems such as 3D reconstruction, obstacle detection, object detection and classification, visual odometry, pose estimation, and tracking.

7. Vision-based sensing systems for reconstructing, recognition, and tracking objects present at a very large distance from the autonomous system.

References

Ju, C.C., Liu, T.M., Chang, Y.C., Wang, C.M., Tsai, C.H., Chen, Y.J., Wu, T.H., Lin, H.M., Chou, H.L., Chen, A. and Wang, A.H., 2020, February. 21.3 A 5.69 mm 2 0.98 nJ/Pixel Image-Processing SoC with 24b High-Dynamic-Range and Multiple Sensor Format Support for Automotive Applications. In 2020 IEEE International Solid-State Circuits Conference-(ISSCC) (pp. 326-328).

Kim, H., Leutenegger, S. and Davison, A.J., 2016, October. Real-time 3D reconstruction and 6-DoF tracking with an event camera. In European Conference on Computer Vision (pp. 349-364). Springer, Cham.

Mueggler, Elias, Guillermo Gallego, Henri Rebecq, and Davide Scaramuzza. "Continuous-time visual-inertial odometry for event cameras." IEEE Transactions on Robotics34, no. 6 (2018): 1425-1440.

Rebecq, H., Horstschäfer, T., Gallego, G. and Scaramuzza, D., 2016. EVO: A geometric approach to event-based 6-DOF parallel tracking and mapping in real time. IEEE Robotics and Automation Letters, 2(2), pp.593-600.

Shimizu, R., Arimoto, M., Nakashima, H., Misawa, K., Ohno, T., Nose, Y., Watanabe, K. and Ohyama, T., 2009. A charge-multiplication CMOS image sensor suitable for low-light-level imaging. IEEE Journal of Solid-State Circuits, 44(12), pp.3603-3608.

Real-time Computer Vision Systems: Technologies & Platforms

Collaborators

- IIT Jodhpur
- CSIR-CEERI
- IIT Indore
- IIT Delhi
- Industry Partner: NVIDIA

Introduction

The computer vision technology landscape continues to undergo relentless transformation, concurrently creating new application opportunities. Market analysts forecast that the market for computer vision is expected to grow to USD 17.4 billion by 2024 - growing at a CAGR of 7.8%. Major factors driving this growth include increasing need for quality inspection and automation, growing demand for vision-guided autonomous systems, rising demand for computer-aided diagnosis of medical images, and emerging field of augmented/virtual reality. The major challenge and issue brought by these emerging applications areas is real-time image/video data analytics with very low latency. Therefore, real-time computer vision, which is viewed as a technology with high potential, is the need of an hour.

Most of these current generation computer vision applications require processing / analytics of incoming image/video data near the camera nodes or at edge. But, the resource inhibited environment of node/edge devices, owing to their limited energy budget and low compute capabilities, render them a challenging platform for deployment of advanced computer vision algorithms for desired video data analytics. This need is further emphasized, as transfer of real-time raw video data typically requires high-bandwidth connectivity, and thus, communication between the camera node/edge and the cloud/fog is constrained not only in terms of bandwidth (for low/normal-bandwidth connectivity scenarios) but also in terms of latency. Therefore, now a day's computation shifts from the data centres and the cloud, to the fog and on the edge. In order to support the paradigm shift of the connected world near-camera computation and near-camera intelligence are starting to emerge as necessities. But processing of video/image data near the camera requires solving issues related to real-time processing, minimum latency, low energy consumption, memory optimizations, and communication network bandwidth (due to limited connectivity).

Although, increased computational complexity of deep neural networks has provided a significant improvement in performance, the increased hardware requirements for training and inference (testing) is drifting these networks away from being deployed in real world applications involving resource constrained devices, especially for portable solutions. Importance of addressing these issues can be seen from many upcoming research papers and patents in this field.

Recently, with the increasing compute power of embedded platforms, there is a growing interest in performing deep learning inference tasks for many real-time computer vision applications at edge level where latency in inference is a critical issue and decision needs to be taken in real-time based on inference output. This is similar to the in-situ video data processing

scheme. This kind of processing requires design of a dedicated hardware platform (FPGA based) with real-time processing capabilities, low energy consumption, and low latency.

Therefore, to address these issues we propose to develop optimization strategies to reduce the computational complexity of deep networks at the cost of insignificant accuracy loss so that the state-of-the-art high accuracy networks can easily be ported to resource constrained devices, and to increase computational capabilities of the edge processing devices by developing FPGA based hardware accelerators/platform for porting of optimised computer vision algorithms. Also, we look at the problem of developing hardware for haptic devices to be used in VR and other applications. The proposed plan to develop a mid-air haptic device has the potential to develop into a touch free surface device. In the COVID-19 scenario such devices can revolutionise the technology for interactive public utilities like ATM, Vending machine, etc. because of hygiene and safety.

The consortium of IIT Jodhpur, IIT Delhi, CSIR-CEERI and IIT Indore will address the problem of optimising computer vision algorithms for the target platforms given time, energy and memory constraints of the target. The team has delivered embedded Computer Vision solutions for different applications to industries including SAMSUNG, LG and Cognizant Solutions. The team had developed a standalone FPGA based smart camera system with real-time motion detection and object tracking with automatic pan-tilt capabilities jointly under the supervision of the present director, IIT Jodhpur.

National Need

Recognising AI's potential to transform economies and the need for India to strategise its approach, NITI Aayog has created a national strategy discussion paper to establish the National Program on AI i.e. AIforAll. This document clearly highlights the various applications of computer vision for societal and strategic growth of the country. Some of these are industrial environment monitoring, autonomous systems, healthcare, etc. A lot of work is happening in the county to address these application areas of computer vision. But the major requirement for these application areas is the development of standalone real-time computer vision systems. The end to end solutions to build real-time computer vision systems are not available. The research and technology development in the direction is still at a preliminary stage and significant effort for development of optimization strategies to reduce the hardware complexity of computer vision algorithms, development of FPGA based hardware accelerators/platforms for porting of optimised computer vision algorithms, and design of complete software development environment are required to build standalone computer vision systems. The project proposal focuses to address these problems.

Objectives

Aim of the project is to develop technologies for the resource constrained edge level real-time computer vision algorithms. The major objectives of the project are:

- To develop computation resources, power, bandwidth, latency, and memory optimized algorithms for real-time computer vision systems.
- To develop specialized resource constrained dedicated hardware accelerated edge computing architectures for distributed intelligence in real-time computer vision systems.

- To develop a complete tool chain for embedded implementations of Optimized algorithms for computer vision applications in resource constrained environments.
- To establish stat-of-the-art Real-time Computer Vision Lab equipped with both software development tools and hardware platforms at IIT Jodhpur

Research Challenges

The project addresses major challenges in the development of resource constrained algorithms and hardware architectures for optimizing power, bandwidth, latency, and memory for real-time computer vision systems. Project activities specific research challenges are mentioned below.

1. Design and development of implementation schemes for the optimization of deep learning architectures for real-time computer vision systems

2. Edge AI framework based on Federated Learning for Resource Constrained edge level implementations of real-time computer vision applications

3. An FPGA based Framework for design and development of real-time computer vision systems

4. Optimization of Computationally Intensive Machine Learning models using device-edge co-inference in resource constrained environment

5. Development of mid-air haptic interface

Dependability
Collaborators

- IIT Jodhpur
- IIT Bombay
- IISc Bangalore
- User Agency: DRDO, MHA (tentative)
- Industry Partner: DependableVision.AI

Introduction

Dependability is a fundamental premise of any decision we take. In every walk of life, computer vision systems are playing a significant and increasing role. They are being employed for making mundane day to day decisions such as healthy food choices and dress choices from the wardrobe to match the occasion of the day as well as mission-critical and life-changing decisions such as diagnosis of diseases, detection of financial frauds, and selecting new employees. Many upcoming applications such as autonomous driving to automated cancer treatment recommendations has everyone worrying about the level of trust associated with vision systems today. The concerns are genuine as many weaker sides of modern vision systems have been exposed through adversarial attacks, bias, and lack of explainability in the current rapidly evolving vision systems. While these vision systems are reaping the advantage of the novel learning methods, they exhibit brittleness to minor changes in the input data and lack the capability to explain its decisions to a human. Furthermore, they are unable to address the bias in their training data and are often highly opaque in terms of revealing the lineage of the system and how they were trained and tested. It has been conjectured that the current use of AI and CV is based on only about 20% of the data the world has access to. Rest 80% of the data that can help the systems is not available because of regulations and compliance requirements around security and privacy. The present computer vision systems haven't demonstrated the ability to learn without compromising on the privacy and security of data. Nor can they even assign appropriate credit to the data sources.

The figure below shows some of the unanswered questions that emerge from the state-of-the-art systems prevalent today vs what is expected from the dependable systems. As the TIH aims to operate in mission critical areas like defence and medical, it is of extreme importance that the technologies we build are dependable, the algorithms are robust and fair, decisions are not biased and they can be explained, and the process followed is secure and preserves the privacy of users. With this motivation, the TIH aims to focus on cutting edge-research in dependability in India. We aim to build a consortium of researchers working in dependability across the country and also internationally, and imbibe this as a mandatory component of all the technologies generated in the iHub Drishti. This is one of the unique components of the hub and our collaborators and user agencies share the importance of this upcoming problem statement.
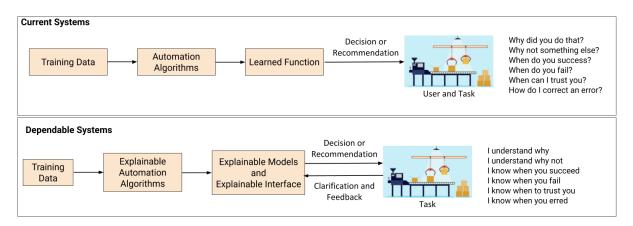
Figure: The figure below shows the overall structure of the solutions of the CV, AR and that we plan to build in the TIH[3].

Objectives

iHub Drishti aspires to obligate good practices of dependable system-design across all stages of technology development, irrespective of the application domain.

Core Research: We aim to design systems that follow the fundamental principles of dependability in CV, AR and VR systems.

- Bias and fairness,
- Explainability and interpretability,
- Adversarial robustness,
- Governability
- Trustworthiness and
- Privacy and security.

Evaluation Framework: We aim to design evaluation metrics and frameworks to determine whether a computer vision system follows the above dependability criterions.

Training and Outreach: Given the importance of dependability, one of our important objectives is to organize regular schools and workshops in this area to train the students, professionals, and users agencies and make them aware of the best practices and implications.

Research Questions

Although researchers are working in many areas related to bias, fairness, and trust, there has been limited attempt to comprehensively bring together all aspects of trust in Trustworthy computer vision under one umbrella and understand the inter-linkages between these topics. This TIH proposes to bring together experts from AI, CV, VR, security, and robust learning communities, to define the foundations of secure computer vision and design solutions to increase transparency and privacy. Some of the projects that we plan to undertake are:

- Design and development of bias invariant computer vision models,

---

[3] It is adapted from the DARPA program on Explainable AI, 2016.

- Design of evaluation metrics for bias estimation,
- Develop measures to evaluate the *biasness* of a database
- Design and develop de-anonymization algorithms that can de-anonymize images based on certain attributes,
- Privacy affirming frameworks to ensure secure and anonymized share of resources,
- Design and development of attack resistant computer vision models,
- Design and development of attack detection and mitigation algorithm,
- The design of explainable models,
- The design of the explanation interfaces for computer vision models

Contributions

- ***Foundational Research:*** Dependability is a very important horizontal of iHub Drishti which is universally present for all the verticals. Therefore, we will perform fundamental research in designed dependable and trustable models that can be used across all the verticals currently identified, and upcoming.
- ***Collaborative Research:*** The research and activities in this TIH are currently in conjunction with IISc Bangalore and IIT Bombay. We aim to build a consortium of academic institutions, industries, and user agencies that share the common goal of building dependable systems.
- ***Infusion in Curriculum:*** It is our strong belief that building dependable systems is of utmost importance to ensure that the automation systems are reliable and serve mankind. It is also important that the importance of this is taught to students and integrated in the curriculum. Therefore, we will incorporate these in the academic activities under the TIH and we will also organize short term and long term educational activities to increase the awareness and knowledge.
- ***Industrial Collaboration:*** The implementation of computer vision systems is mostly undertaken by the industries and there is growing interest in all the entities about this. We will partner with industries to foster strong linkages and promote inclusion of these concepts in the automotive systems being implemented for all the tasks.

Data Collection, Annotation, and Curation

Introduction

Data is one of the most important components of designing a CV, AR or VR system. According to the advisory firm Cognilytica[4], out of the total time spent in designing a computer vision or machine learning system, more than 70% of their time is spent in collecting, annotating and curating data for training the system. In order to learn and understand variabilities in the environment, it is important to identify the kind of data characteristics and distribution required. The data then has to be appropriately curated, annotated and labeled to answer all relevant questions. In addition, traceability/lineage[5] of data sources is a critical consideration towards ensuring dependable systems. Given the importance of this task towards ensuring the efficacy and reliability of entire cyber-physical systems, data collection, annotation and curation will be one of the key horizontals (core) for the Hub. It will be associated with the lab for Data Collection, Annotation, and Curation. Activities along this horizontal will involve interactions with domain experts particular to the different verticals of iHub Drishti.

Data collection is the process of gathering and measuring information on variables of interest, in an established systematic fashion that enables one to objectively answer research questions, test hypotheses, and evaluate outcomes. Data collection is common to all fields of study. While processes of collection may vary with respect to the intended purpose, the emphasis on honest practices is omnipresent. There are a number of reasons for data collection becoming a critical issue, when it comes to computer vision based research. First, as machine learning based CV is becoming commonplace across a number of application domains from surveillance to diagnosis, there is a splurge of new areas that do not necessarily have enough labeled data. Second, unlike traditional machine learning, deep learning techniques depend on large amounts of labeled data. Third, given the recent upheaval concerning biases in datasets and consequent prejudicial inferences by CV algorithms, it is essential to build all-inclusive datasets - in addition to techniques for bias-mitigation.

Under iHub Drishti, we aspire to mandate honest and ethical data collection practices and contribute to solutions for the three challenges highlighted. Given the myriad and unique application verticals of the TIH, our datasets will in all likelihood be one-of-a-kind. Depending on the application and the problem at hand, data could be crowdsourced, physically collected, or artificially generated. For instance, the data for training computer vision systems of autonomous cars would be collected by physically driving the car for thousands of miles, whereas data for scene recognition could be collected from the world wide web.

Many organizations use data platforms to publish their datasets, documents, associated services, tools and applications to meet requirements. For example, the Government of India is currently using the Open Government Data (OGD) Platform to collate data from different ministries and departments. The platform serves as a single point for hosting various datasets/apps across a variety of sources. This helps the government to make better decisions by analyzing recent and past data in the relevant domain. Currently, data.gov.in the OGD platform has close to 270 thousand resources. Despite a lot of advancements in the technology

---

[4] https://www.cognilytica.com/2019/03/06/report-data-engineering-preparation-and-labeling-for-ai-2019/

[5] AI Principles: Recommendations on the Ethical Use of Artificial Intelligence by the Department of Defense, *Defense Innovation Board*, USA, 2019.

domain, the existing platform is not suitable for computer vision and machine learning research without rigorous efforts to analyze the datasets. Moreover, the datasets are often incompatible with other popular datasets, and it is a trying task to extract deeper insights and build a utility out of it. Given the predicament, it is imperative that iHub Drishti designs its very own CV-AR/VR datasets - which would not only be indigenous but also be of immense value to research and researchers across the globe.
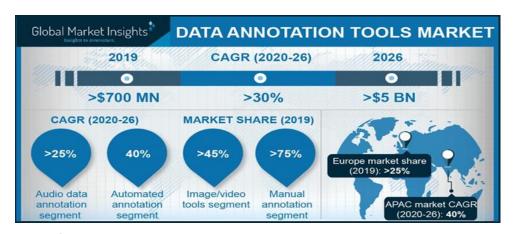

Figure: Projection[6] of data annotation tools market.

Given the increasing use of machine learning over the past decade, the size of the Data Annotation Tools Market exceeded USD 700 million in 2019 and is estimated to grow at a CAGR of over 30% between 2020 and 2026 (refer Figure above). Further, machine learning coupled with human assistance helps seamlessly turn unstructured and raw data into structured and highly useful Business Intelligence (BI) solutions by deploying data annotation tools, lending additional thrust to market growth.

Data annotation is followed by **Data curation,** which stands for the extraction of relevant details from scientific sources (journals, databases, spreadsheets, documents) and converting these into standard electronic formats. In order to be effective, data curation requires active management of information throughout the duration of its utility or usefulness. This facilitates data discovery and retrieval, quality assurance, value-addition, and re-use and preservation of information. For certain critical application areas (like medical imaging), it is imperative that strict-standards for data collection, curation and annotation are adhered to. Data Annotation and curation typically involves:
- Understanding the purpose of data collection
- Development of plans for data extraction, parsing, annotation and validation
- Identification and application of data standards or other such metrics to facilitate uniform comprehension and re-use
- Generating data-sheets or metadata (highlighting assumptions about the data or collection process; properties of the data collected; information about the sampling group; the collection timeframe; data-expiry details; methods of preprocessing, cleaning and labeling of data; and the potential uses and off label uses of the data) for the data collected. This is envisaged to encourage traceability, requirement monitoring, responsible data-use and helps mitigate data dependencies during system-design.

---

[6] Source: https://www.gminsights.com/assets/img/data-annotation-tools-market.png

Drawing from the exceptional data-sets and subsequent annotations that would be generated by iHub Drishti, data curation would unquestionably be a fundamental activity.
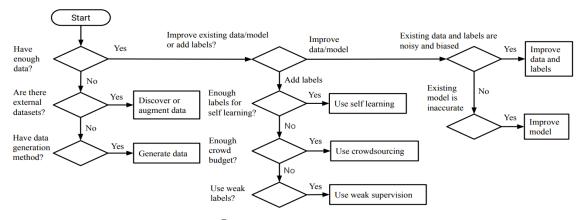

Figure: Graphical representation[7] of data collection process and subsequent curation.

The next step is to build a **data repository**. To build a Medical data bank which will be a nationally and globally recognized resource repository for datasets and algorithms from India with the aim of being a national AI/DL development platform where all clinical data of patients would be stored and curated and eventually given to researchers to enable the rapid development of insights, tools and also aid in bridging the gap of data scarcity in medical domain which is required for development of robust AI applications.

Objectives

Given the diverse range of activities to be executed under iHub Drishti; some of which are unique and hold the potential of leading to major paradigm shifts in the domain of CV-AR/VR research; the objectives underlying data collection, annotation, curation and repository by the TIH can be briefly summarized as:

- To devise novel mechanisms for automatic correlation of new datasets with existing datasets for subsequently processing them.
- To develop ingenious approaches for automatic identification of suitable data mining and machine learning techniques for computer vision, augmented reality and virtual reality applications.
- To reduce the effort and skill-set requirements for the extraction of meaningful information from datasets.
- To develop a plug-and-play architecture for data cloud platforms to add support for new tools and techniques.
- To design a workflow-based approach to add new knowledge to existing platforms.
- To develop a Cloud Platform containing data to address challenges important for the country. The repository will also provide facilities for collection and annotating high quality medical data which will enable expert teams to seek solutions (diagnosis, drug-discovery and public-health) and Provide anonymized data to research groups, developers and data scientists.

---

[7] Source: Roh, Y., Heo, G., & Whang, S. E. (2019). A survey on data collection for machine learning: a big data-ai integration perspective. IEEE Transactions on Knowledge and Data Engineering.

- Feedback mechanisms can be incorporated for the annotation of data further improving data annotation quality and hence would lead to development of more robust AI based applications in areas such as healthcare.

Research Questions

Drawing from the objectives pertaining to data collection, annotation and curation of the TIH, some pertinent research questions are:

- Process Automation: Can collection, annotation and curation of the data be done automatically? Each of these phases is independent of each other. Under the TIH, we would aim to identify a set of tools and techniques to adopt, optimize, combine together and provide it as a platform-as-a-service for data-cleaning and data-pre-processing as per the requirements of specific target problems under various verticals.
- Identification of Dataset Connections: As a new dataset becomes part of the datacloud, it is critical to identify connections between the new and the existing datasets - irrespective of type-heterogeneity. One envisions a system where, as soon as a new dataset (X) is included on a data platform, it will generate the subsequent metadata and explore that of existing datasets to extract a set of candidate datasets that could be processed in combination with X. For example, a dataset from the healthcare domain could be useful in the design of future autonomous vehicles to have emergency first-aid and point-of-care facilities. Under the TIH, techniques will be developed and converted into easy-to-use tools and frameworks to reuse methodologies across heterogeneous datasets.
- Smart Data Processing Engine: Considering data-heterogeneity and incompatible data formats and protocols, it is necessary to design and develop a smart data processing engine that is agnostic to these parameters. The engine should support conversion from one format to another on-demand. It should also be able to fuse data from multiple sources of information to create a more valuable dataset and make it available for vision and AR/VR processing.
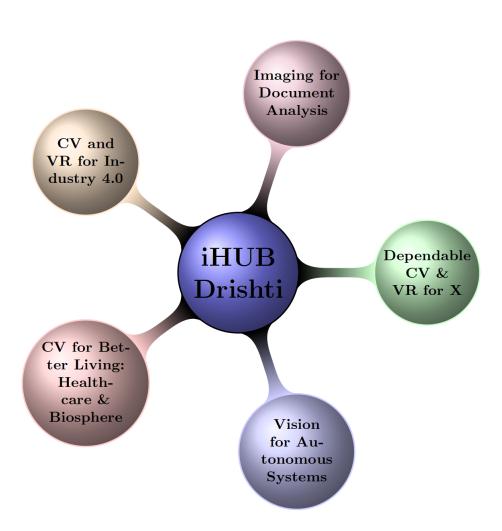
Contributions

This horizontal of iHub Drishti will not only support activities of various verticals but will also help the Indian Industry by providing a common data collection, annotation and curation facility. The major advantage of having an indigineous support for this purpose is that the data remains on servers within the territorial boundary of the country, but is yet available to researchers across the globe. On one side, while India grows to become a major source of data for the rest of the world, on the other side we presently have a handful of startups and small companies who are working on converting unstructured and unlabelled data to their structured and meaningful equivalents for consumption by various autonomous processes. Despite the huge data market in the country, it is unfortunate that we do not have any research laboratories in any of the academic institutions that focus on data collection, annotation and curation.

As part of the iHub Dristhi, we propose to build a world class research and development facility for Data Collection, Annotation and Curation for satisfying the growing demand in the research

and application fields of Computer Vision, Augmented Reality and Virtual Reality. Our key contributions will be:

- A centralized advanced data collection facility to efficiently acquire and store heterogeneous data from a variety of sources, in strict compliance with governmental rules and regulations.
- Development of advanced tools and frameworks for easy and quick annotation of data across different domains.
- Tools for automatic data identification and curation based on requirements.
- A training and skill-development facility to facilitate training in the domain.
- A national facility to annotate and curate all types of data.

The Verticals

- **Vision for Autonomous Systems**
- **Computer Vision for Better Living: Healthcare and Biosphere**
- **Imaging for Document Analysis**
- **CV and VR for Industry 4.0**
- **Dependable AR, VR for X**

COMPUTER VISION for AUTONOMOUS SYSTEMS

Collaborating   Organizations:
IIT Jodhpur, IIT Delhi, IIT Bombay, IISER Bhopal IIT Madras, IIT Kanpur
Industry   Partners: nVIDIA,   KritikalVision.AI,   DependableVision.AI
Foreign Collaborators: Columbia University, JHU, UMD, GaTech, USF, SUNY Buffalo
User Agency: DRDO (CAIR, DLJ, DTRL, DYSL- AT, DYSL-AI)

1. Introduction

With the advancements in technology and wider applications, there has been a significant interest in developing autonomous systems. It can play an important role in search and rescue, surveillance, etc. Initiatives have already been taken by several countries toward research and development of vision-based autonomous systems for defence and security applications. The development of such systems is made possible with advancements in technological developments of small size but efficient vision sensors, servo controls, responsive actuators, communication devices, artificial intelligence, etc. The motivation and driving force behind this effort is the need to fill up the surveillance blinds spots left by today's satellites and spy planes, especially in emergencies. Autonomous systems offer the advantage of portability; hence, it can be carried near the mission area for its deployment and information about the severity of causality can be easily gathered. Apart from the surveillance in emergency scenarios, these autonomous systems with small to large payload capacity will be useful in Autonomous cars, law enforcement, defence applications, agricultural survey, forestry, environmental monitoring etc.

There is also a major international push to design and build vision-based autonomous systems that should interact with each other and the physical world with cooperative objectives. This is achieved through advances in computer vision, wireless communication, the internet, intelligence and the introduction of new and cheap sensing devices. Such systems are usually referred to as cyber-physical systems (CPS) and are the cornerstone of the present social and technical infrastructure. Autonomous systems, for instance, unmanned aerial or ground vehicles and remotely operated underwater vehicles, is one paradigm of such systems that have recently emerged as an enabling technology to facilitate a diverse range of applications. Representative examples include, but are not limited to, inspection and monitoring, border patrol missions, search and rescue missions, bomb disposal, forest fire detection and monitoring, surveillance, data collection, agriculture, and crowd monitoring.

As a part of this Vertical on vision-based autonomous system focus would be laid down on research and development of vision based technologies for achieving semi to full autonomy

for mobility systems. This will allow autonomous systems to navigate and interact in unknown and unstructured environments. Primarily Developmental activity will be taken up in the area of vision based perception to provide ability to see and recognize things around mobility systems and to develop reasoning ability based on vision feedback to navigate and act on the environment. These developmental efforts will be made jointly with stakeholders in loop. A state of the art laboratory on vision based mobility systems will also be developed which will help in implementing technology developed as a part of this vertical.

1.1 Summary

Augmentation of semi-autonomous and autonomous Mobility Systems with computer vision is the focus of this vertical. It is typically desired that the autonomous system should have the capability to move in the unstructured and unknown environment to complete the overall objective effectively, by using their onboard sensing. Additionally, it is of prime importance in such operations to have an understanding of the surroundings so as to avoid obstacles or any other unwanted circumstances while performing a cooperative task. The scene understanding and creation of the map is a crucial step for autonomous navigation for path planning, obstacle avoidance. The importance of the visual information-based depth-sensing arises due to erroneous GPS information or the absence of it. The reliable availability of such a framework will result in the irrelevance of visual information. However, improvement in precision and availability of GPS information cannot be guaranteed due to occlusion. Also, an autonomous system requires a failover mechanism as a non-functional or erroneous operation of the agent should not affect the functioning of the entire system. Furthermore, the transmission of real-time images at high data rates and in a secure manner is a critical role of autonomous systems. In summary, Achieving full autonomy to semi-autonomous operation is a critical aspect, and computer vision will play an important role in achieving the same. Hence Research and development in vision based autonomous systems will play a crucial role in the evolution of next-generation mobility systems and is identified as the main theme of the proposed vertical.
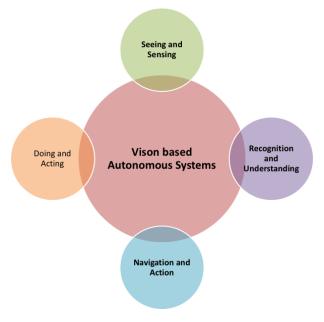

Figure: Thrust areas of Autonomous Navigation

1.2 Objective

Core Research: One of the salient objectives of this vertical is to perform cutting edge research and generate new knowledge through research and development for evolving vision-based technologies for autonomous systems to augment the capability of operating in unknown and unstructured environments. The technology development will address problems in the following four areas:

1. Seeing and Sensing

2. Recognition and understanding

3. Navigation

4. Doing and Acting

Key research and technological issues proposed to be addressed in the above four areas are highlighted in figure below:

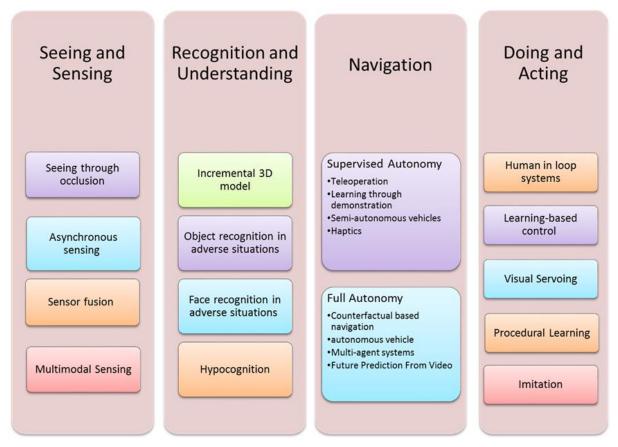Figure: Key research and technological issues proposed to be addressed

Technology Development for Applications: Autonomous systems are applicable in almost all areas of operation, for instance defense, law enforcement, transportation, and railways. We have currently partnered with DRDO labs to design and deliver technology solutions for the following areas:

1. Efficient sensing and sensing through visual clutter.

2. 3D Reconstruction and recognition of objects and movement in unstructured and unknown environments.

3. Biometric solutions for person authentication in unconstrained environments.

4. Develop understanding of the surrounding environment

5. Vision-based navigation of unknown terrains

6. Vision guided coordination of swarm of autonomous systems

7. Visual input guided execution of complex manipulation tasks

The core application technologies will be adapted and translated for meeting target deliveries for end users and stakeholders. In summary, our objective will be to develop and deliver next generation technologies for vision based autonomy. We propose to address problems from a variety of application domains like defense, disaster management, transportation, mining, and handling hazardous environments.

Training and Outreach: It goes without saying that we will generate manpower trained in different aspects of autonomous system development. It is not sufficient to only generate systems and deploy, rather for scalability and efficient usage, technology can be best utilized if the users have the knowhow of the system as well as the technology. Therefore, we will organize training sessions for the DRDO and law enforcement officers in different kinds of technology areas.

1.3 National Need and Impact

Autonomous Systems comes with both an opportunity and challenges for military operations. Defence forces which can explode their benefits will be able to deploy forces with persistence, greater range and speed on the battlefield. The current autonomous systems lack human intelligence, robustness and flexibility and may not be suitable for doing many tasks that humans can do. Many countries have initiated activities in this area and countries which can solve these issues will have an edge in future's warfare. For countries like India, research and development of such autonomous systems is crucial due to strategic location in South Asia.

Furthermore unlike earlier game changing technology like stealth or guided weapons which was majorly developed by military labs much of research and development in autonomous systems is driven by the commercial sector worldwide. Hence, there is urgency to move quickly. Global spending on military autonomous systems was estimated to be $7.5 billion a year in 2018. In the same timeframe, global spending on commercial and industrial robotics systems estimated to be $43 billion a year (Horowitz 2014). Furthermore, elements of technology that give autonomy, for example, better sensors, improved processing, cost efficient inertial measurement units, and miniaturization, is driven by a revolution in Computer vision and Artificial Intelligence.

In order to handle any dangerous and dirty environment, vision based autonomy is an important requirement. In the Indian context, these technologies are critically needed for handling natural disasters for rescue and reconstruction operations. Operations in nuclear power plants, mining, under-sea explorations etc. can be made safer by deployment of vision guided autonomous systems. Natural fall out of these technologies are for consumer applications like driver assistance systems, home security and surveillance, etc. In the present time challenges of COVID-19 have given rise to newer areas of applications of vision for

autonomous systems. These systems are needed for disease surveillance, bio-safety of the home and work environments and remote handling of patients minimising exposure of health workers. Therefore it is envisaged that the focused Technology Innovation Hub on CV and AR VR proposed at IIT Jodhpur jointly with end users as collaborators can pave a way for development and deployment of critical technologies for vision based autonomous systems.

1.4 Background

The requirements of vision based autonomous systems is finding a significant increase in multiple sectors including defense and law enforcement. Enhancing and automating the capabilities of the different sectors is of paramount importance for the efficient and effective operations in the country. Globally, several countries have projects and systems to enhance the capabilities of defense forces and law enforcement agencies. One of the major objectives of this vertical is to improve the technical capabilities and support of our defense forces. We have developed a strong consortium with researchers, startups and multiple DRDO labs are partnering with us on almost all the focus areas in this vertical.

- The previous experience includes design and development of various integrated circuits, readout electronics and complete camera systems. The camera systems are designed for various specifications like high speed, low power, low noise and others. The experience in camera design will help in any customised application required for defense. In addition to this, the experience in on-chip compression algorithms and machine learning algorithms will help in optimization of resources and improves the battery life of the systems. Our team is focused on design and development of state of the art CMOS image sensors for high dynamic range, high speed and time of flight 3D cameras. Our team also has developed working prototypes for multimodal resection and SpO2 measurement and gained experience in designing of an image sensor with on-chip metal grid polarizer to develop an integrated solution for SpO2 measurement.

- The members in the consortium have worked on multiple projects and have demonstrated expertise of delivery. The technologies developed by our members are also being used by agencies in different countries. The team has developed and delivered solutions for face recognition in constrained and semi-constrained environments. Team members have designed and delivered multimodal biometric fusion and quality assessment algorithms of biometric modalities in unconstrained environments. The system also incorporates context switching to determine the best suited modality of authentication, on a case by case basis. Two of our team members have contributed in building the biometric standards for Aadhaar as well as for e-Governance Applications for the Government of India. The team has also designed face based authentication systems for autonomous golf carts.

- Team members have extensive experience of design, modelling, navigation and control of autonomous systems/vehicles. Several projects have also been carried by the team members in the area with funding supported from DRDO, HAL, ARDB, SERB and DST in the area of autonomous vehicles.

- Some of the important projects include Optimal Control of Multi mobile manipulators where the problem of path planning through tight spaces, for the task of spatial payload transportation, using a formation of mobile manipulators was addressed. Team members have also worked on design of distributed control algorithms for a group of multi-agent systems, modeled as unicycle kinematics, to achieve difference formation patterns, while restricting their motion to a given workspace during stabilization. A novel compliant robot was also designed for traversing on unstructured terrains. A novel and generic reactionless visual servo controller for redundant robots was developed using a task function approach.

- Design and development of autonomous or unmanned rotary wing vehicles has been initiated at IIT Jodhpur to set up a laboratory for design and development of an autonomous mini-helicopter. The major focus of this research activity was to design and integrate the sensors and communication systems in a model helicopter, and provide the autonomous control and navigation capability.

2. Seeing and Sensing

For modern armies, Computer Vision is a vital enabling technology that helps security systems detect enemy troops or saboteurs and enhances the targeting capabilities of guided missile systems. Military concepts such as situational awareness rely heavily on image sensors to deliver battlefield intelligence used for tactical decision-making. Another key Computer Vision application in the areas of autonomous vehicles, which need to navigate challenging terrain and detect adversaries. Computer Vision systems also support human drivers and pilots, enabling them to evade enemy fire. As is the case with many military innovations, technology transfer is now benefiting a broad range of industries. Further, owing to the increasing security concerns in the country and worldwide, governments are continuously increasing the area under surveillance. Most public spaces including airports, railway stations, government buildings, and bus terminals are equipped with CCTV cameras in order to monitor daily movements. The availability of sophisticated and cost-effective surveillance devices along with the ease of installation and invigilation has provided the impetus to the video surveillance infrastructure at large. Further, it is expected that the majority of the areas in major cities will be monitored via surveillance cameras, schools, stores, and public places. The surveillance not only includes the monitoring of individuals, but it can be used for monitoring everything in the view of these cameras. Defence applications also require similar activities to be performed by UAVs and drones. The set of challenges vary in that case because the images and videos captured by drones are captured in unconstrained environments by moving cameras, with unconstrained individuals and/or objects. Applications like identifying friend and foe in marine environments and intense weather conditions are some of the challenges.

Globally, there are several companies and research labs working towards individual problems discussed earlier in the seeing and integrating technologies to design unified solutions. In India, most of these activities are performed manually, which requires trained and experienced manpower, and is difficult to scale across the country. The proposed TIH aims to collaborate with law enforcement and defence organizations to perform research and design technologies in the area of sensing that can assist with respect to advanced surveillance requirements. Some of the projects that we plan to undertake can be segregated into different categories such as:

- Seeing through occlusion
- High Dynamic Range Image Sensor
- Asynchronous sensing
- Sensor fusion

2.1 Seeing Through Occlusion

Occlusion is one of the main limiting factors of sensing in various applications. Sensing through occlusion has been an emerging area of research. The sensing technique varies with the nature of occlusion. The Sensing part of TIH iHub Drishti plans to contribute towards the development of effective technology that can address the sensing modalities in various types of imaging mediums. We have classified these imaging mediums into three categories i.e. Scattering, Reflecting and Refracting. The nature of the interaction of EM radiation with matter changes across the spectrum with a change in wavelength as shown in the figure below. Details of each type of occluding medium and seeing-through strategy has been described below.
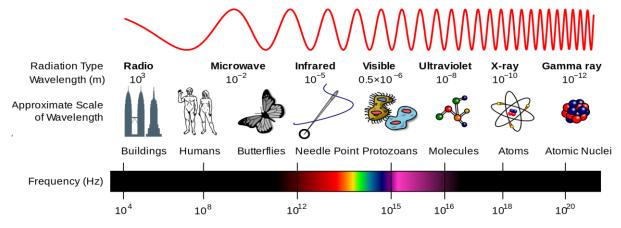

Figure: EM Spectrum and Scale of Wavelengths.

Scattering Medium (Haze, Fog, Rain, Water) - The wavelength and size of the particles in the imaging medium play a crucial role. Comparable size of imaging wavelength of EM spectrum to medium particle causes scattering of the illumination. The nature of scattering also changes with dependent from wavelength-dependent scattering (Rayleigh Scattering) to wavelength-independent scattering (Mie Scattering). The wavelength-dependent scattering is useful in applications such as underwater imaging. The wavelength-independent scattering in mediums such as dust, water droplets, smoke etc. results in such as Haze, Fog, SMOG etc. [Nayar et al. 1999] proposed a simple scattering model for dehazing based on Atmospheric Light (A), transmittance (t) and Radiance (J) as shown in the figure below. This simple though effective model has been utilized extensively since inception to enhance the degraded image not only due to haze but also images belonging to underwater, smoke, medical etc. domains.
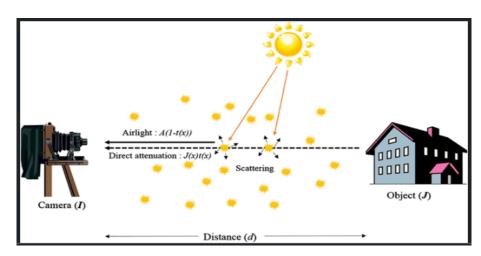
Figure: Imaging in Scattering Medium.

Research challenges:

o   An adaptive real-time method for image enhancement in scattering mediums such as water, haze, fog, smoke etc. with respect to variation in density. The properties of the medium can be adaptively selected for such a purpose.

o   A robust method/prior/principle for estimation of non-uniform atmospheric scattering parameter A to address the non-idealities in the model. Since, accurate estimation of A plays a crucial role in accuracy of recovery of the scene radiance and dynamic nature of the medium may result in different A in different locations of the scene.

o   A robust framework for recovery of the scene radiance J under varying illumination conditions including saturations. Utilization of other parts of the EM spectrum can be useful under such cases.

o   An image enhancement framework in scattering medium under high-speed operations.

Absorbing and Reflecting Medium (Walls, Sand Storm, dust, Cloud) -   Absorbing and Reflecting Mediums are another main types of occlusions encountered by imaging systems. Suitable EM spectrum can be utilized for a target imaging modalities e.g. high energy EM spectrum can be utilized for the detection of metal, bones etc. The size and thickness of the target object plays a crucial role in determining these aspects of imaging. Thus, appropriate imaging modalities are utilized for desired application based on properties of the target such as terahertz imazing [Redo-Sanchez et al. 2016], microwave imaging etc. The Non-Line Of Sight (NLOS) imaging techniques are another emerging area of research in this domain which employ Time of Flight (ToF) imaging techniques to image through occluders as shown in figure below.
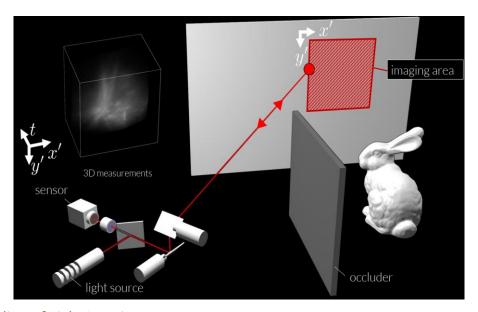
Figure: Non-line of Sight Imaging.

Research challenges:

- o A real-time NLOS imaging system by utilizing conventional sensors such as thermal camera, radar under a deep-learning framework such that they can be utilized for real-world applications such as sand, wall, cloud etc.

- o Accurate complex scene recovery of the scene containing objects of different reflecting properties. Currently, state of the art methods are focused upon detection of only a single object behind the occluder which is not a practical scenario. Presence of multiple objects results in mixing of various signals. Utilization of multiple sensors to capture the different modalities of the scene and characterizing them using a deep-learning network can be useful in such cases.

- o Generalization of such a characterization is also very important for a general purpose solution as opposed to state of the art methods which targets a specific scene condition.

Refractive Medium (Mirage) - Refracting medium can be another invisible type of occlusion which results in the illusion of the object at different location than original due to Total Internal Reflection (TIR) caused by variation in refractive index due to temperature gradient as shown in figure below. This natural occurring phenomenon is referred to as Mirage.

Research challenges:

- o A general robust system for detection of presence of mirage, refraction under varying optical properties as variation of the temperature, turbidity etc will alter the refractive index which will change the position of the Mirage appearance.

- o An estimation framework for a measure of the extent of the refraction and a corrective system for shifting of the object in the image plane due to refraction in phenomena such as mirages etc under variable conditions such as temperature.
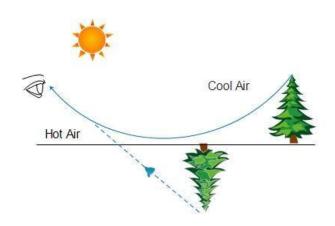
Figure: Imaging in Refractive Medium (Mirage).

Outcome: In this Sensing part of TIH iHub Drishti, we would explore various specific imaging modalities for a given set of target and imaging medium. The underlying optics and the deep network framework can be crucial in this regard. Development of sensors designed specifically for corresponding properties of the scene medium such as refracting, reflecting or scattering to possibly invert the degradation process itself or possibly inverse the effect by post-processing the captured data. In this TIH iHub Drishti, we would address the current technological and practical challenges as discussed in the subsections above to develop a solution to see-through occlusion technology for various applications such as mobility technologies for assisted and autonomous navigation under adverse conditions, Surveillance, Military applications, disaster management, Industry 4.0 etc.

2.2. High Dynamic Range Image Sensors

Comparatively, little attention has been given to the more difficult problem of recognising an object in the adverse conditions of the real world. Human visual system is extremely robust, even in conditions of dynamic, non-uniform lighting, poor visibility, and occlusion. Further, there are many instances where the human visual system can correctly interpret an image, even when that image possesses insufficient information to enable such interpretation. It is our belief that such a system can only function by making use of some prior knowledge of the scene". Drones can be used in multiple settings such as active surveillance for a particular person/group of people both in indoor and outdoor settings or for passively monitoring region/event from a height covering a large field of view with greater mobility than the usual CCTV cameras. Drones can also be used to provide surveillance aerially by performing event detection or activity detection. This framework can be utilized to overcome fundamental limitations of sensing to achieve high accuracy in respective applications. High Dynamic Range (HDR) image sensors are one of the targeted design problems.
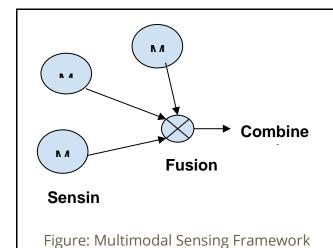
Figure: Multimodal Sensing Framework

Figure: As image showing camera with HDR mode OFF and ON.

CMOS cameras are commonly used in imaging applications like in autonomous systems where the cameras act as the eyes of the systems. The availability of sophisticated and cost-effective surveillance devices along with the ease of installation has provided impetus to the video surveillance infrastructure at large. Defence organizations require activities to be performed by UAVs and drones. The images and videos captured by drones are captured in unconstrained environments by moving cameras. Applications like identifying friend and foe in marine environments and adverse weather conditions are some of the challenges. To cater the defence applications, a high dynamic range CMOS image sensor (CIS) will be designed.

Research challenge: This work is mainly focused on reduction in noise floor of the readout chain along with increment in signal at the pixel level. The dynamic range will be enhanced making the camera suitable for low light imaging applications. In addition to this, the image sensor will be designed with column-parallel readout to enhance the speed. The column readout adds its own noise called the fixed pattern noise (FPN). The proposed image sensor will be designed with inbuilt FPN cancellation.

Outcomes:

1. Design, analysis, pre-layout, post-layout simulations and fabrication of high dynamic range and high frame rate CMOS image sensor

2. Working camera module targeted for autonomous systems.

2.3 Asynchronous Sensing

Introduction: In the vision-based defence applications, the vision sensors are used for the purpose of reconstructing a 3D map of the scene, localization, navigation of the autonomous system or the manned vehicles. The mainly used vision sensors for this purpose are the mono or stereo colour cameras, point cloud scanners, etc. The main demand of the vision-based defence applications is the high temporal resolution (low latency), low power requirement, and higher per-pixel bandwidth. However, the vision sensors such as colour cameras are synchronous sensors that capture the data at a predefined clock frequency. For example, a video color camera captures frames at a prespecified frame-rate. They further capture intensities for all the pixel locations irrespective of whether there is motion at a pixel or not.

The majority of the defence application scenes are the desert regions where only a few pixels exhibit motion. Therefore, capturing intensities for all the pixels unnecessarily increase the power consumption and per pixel bandwidth, and the latency. Therefore, the main idea is to capture only the motion present in the scene that alleviates all these limitations and perfectly fulfils the requirement of the defence application systems. The asynchronous vision sensors (Event-based Cameras or Single-photon avalanche diode (SPAD) cameras) exactly achieve this task which only records the motion present in the scene at each pixel location independently and asynchronously. In the below figure, we show the overall problems to be solved using the asynchronous sensing techniques.
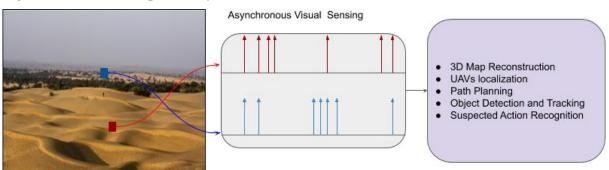

Figure.: Asynchronous visual sensing for navigation and planning in the warzones. (Desert image credit: Wikipedia)

Research Challenges: Research on solving classical computer vision applications using the visual data acquired using the asynchronous sensing is still in its infant stage and mostly done in the lab scenarios. However, real life problems, for example 3D reconstruction, recognition and navigation, using the asynchronous are still open and require a lot of effort. Furthermore, the existing navigation solutions using the asynchronous sensors are proposed under the very controlled environments (such as indoor scenes where the surface and objects are rigid in nature) [Mueggler et al. 2018, Rebecq et al. 2016, Kim et al. 2016].

Research Problems: In order to build the computationally and power consumption efficient defence sensing system, we will use the asynchronous vision sensors. We will further use these asynchronous systems for the real-time 3D reconstruction of the war scene, obstacle detection and avoidance, and automatic navigation of the robots towards the salient motions in the scene. There exist a few existing event-based cameras for asynchronous visual sensing such as DAVIS camera [Brandli et al. 2014] and dynamic vision sensor (DVS) [Lichsteiner et al. 2008]. We will further develop deep learning-based algorithms for efficient processing and solve the classical computer vision problems such as 3D Reconstruction, Obstacle Detection, Object Detection and classification, Visual Odometry, and Pose estimation on the data acquired through the event-based cameras. In summary, we will solve the following research problems.

1. Vision based autonomous systems with asynchronous vision sensors for defence application with the capabilities of automatic 3D reconstruction of the territories, obstacle avoidance, object detection.

2. Deep learning based approach for handling the data acquired through the event-based cameras and solving the problems such as 3D reconstruction, obstacle detection, object detection and classification, visual odometry, pose estimation, and tracking.

2.4 Sensor Fusion

Introduction: The military typically operates in demanding, dynamic, semi-structured and large-scale environments. This reality makes it difficult to detect, track, recognize/classify, and respond to all entities within the volume of interest, thus increasing the risk of late (or non-) response to the ones that pose actual threat. A key challenge facing the military operators, in these contexts, is the focus of attention and effort, that is, how to make the most effective use of the available but scarce sensing and processing resources to gather the most relevant information from the environment and fuse it in the most efficient way. Adaptive Data Fusion and Sensor Management can aid this information gathering and fusion processes by automatically allocating, controlling, and coordinating the sensing and the processing resources to meet mission requirements. This multimodal information can be combined in a fusion framework to obtain the combined information as shown in figure.

Research challenges

1. Raw data collection from different sensors followed by a data fusion algorithm.
2. Prototype development.
3. Technology solution (MEMS and CMOS in a single integrated platform).

2.5 Technology Fallouts

1. An adaptive real-time method for image enhancement in scattering mediums such as water, haze, fog, smoke etc. with respect to variation in density which is robust against non-uniform atmospheric scattering, varying illumination conditions including saturations and high-speed operations.

2. A robust real-time NLOS imaging system utilizing conventional sensors under deep-learning framework which can be utilized for real-world applications such as imaging of a general scene under sand, wall etc

3. A general robust detection and correction system for the mirage, refraction under varying optical properties.

4. Design of high dynamic range and high frame rate CIS with inbuilt FPN cancellation for defense and autonomous systems.

5. Vision based autonomous systems with asynchronous vision sensors for defence application with the capabilities of automatic 3D reconstruction of the territories, obstacle avoidance, object detection.

6. Deep learning based approach for handling the data acquired through the event-based cameras and solving the problems such as 3D reconstruction, obstacle detection, object detection and classification, visual odometry, pose estimation, and tracking.
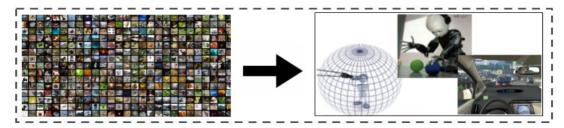
3. Recognition and Understanding

Figure: Computer vision is advancing from learning with a disembodied bag of labeled examples to learning in the context of moving and acting in a practical robotic environment also known as embodied visual learning. Considering the unstructured environment in the context of defence applications, we aim to develop recognition methods for challenging robotic environments.

Object recognition has been a highly studied research area in Computer Vision [Krizhevsky et al., 2012; He et al., 2016]. The community has achieved remarkable success in Imagenet Large Scale Visual Recognition Challenge (ILSVRC) recently where state-of-the-art methods are better than humans [Simonyan and Andrew Zisserman 2015; He et al., 2016]. However, most of these successes are limited to carefully curated datasets in a structured environment and a limited number of categories. Most state-of-the-method do not deliver well when tested on an unstructured and unknown environment. Consider scenarios of recognizing tanks in a war field by capturing videos from a drone, recognizing a suspicious action in an airport or unconstrained face identification, or in-the-wild object recognition in a robotic environment. Despite several applications associated with defence, such recognition problems in unstructured and unknown environments have not been addressed in a rigorous setting. We aim to target these problems under the recognition theme of iHub Drishti. Specifically, our goals are as follows:

- Developing systems for embodied visual recognition
- Devising methods for Incremental 3D reconstruction
- Building real-time suspicious action recognition systems
- Developing a biometric solution for security application
- Hypocognition

Typically, the defense application areas are unknown and unstructured. For example, rapidly changing surface structures and the presence of mirage in the desert areas. Unknown, cold, occlusions, and the highly variable environment in the mountains. Cluttered and unknown forests with occlusions due to trees. The problem of object detection, recognition, and tracking is a fundamental problem for effective defense applications. However, this problem is a challenging problem to solve in war-zones due to the reasons mentioned above. Furthermore, existing deep learning approaches for solving this problem in the warzone may not perform well due to the out of distribution test points. For example, the enemy's war tanks, army dress can be completely unknown and have different make and models. In order to solve this problem, we will propose techniques based on the visual attributes for handling unknown objects, seeing through occlusions to handle occlusions, causal models and uncertainties (Bayesian Neural Network) for making better decisions about the out of distribution examples,

and 3D models for adding more information. In the figure, we show the overall problems and applications we would like to work on.
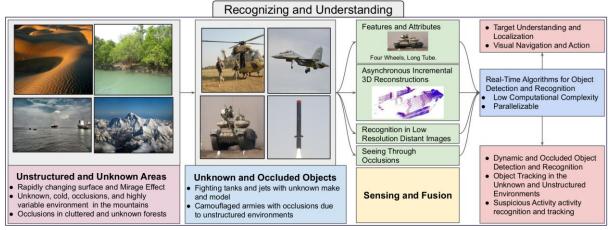
Figure: Planned visual recognition framework in unknown and unstructured environments. (sub-images credits: Wikipedia)

3.1 Embodied Visual Recognition

Modern learning-based recognition methods are often passive in nature and lack the ability to move in the environment and actively control the viewing angle to better understand object shapes and semantics. Such a recognition method is insufficient in object recognition in an unstructured and unknown environment. Therefore, we plan to develop methods for embodied visual recognition where an agent learns to move left or right for a better perception of objects. Further, we plan to extend this to an unknown environment by addressing zero-shot recognition in embodied visual learning setup.
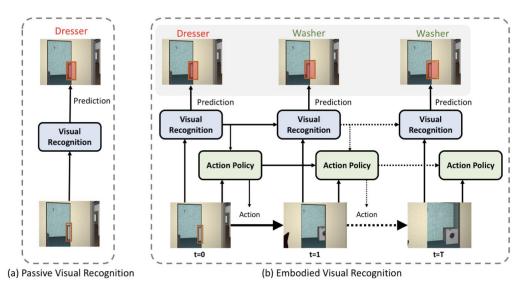
Figure (Figure taken from [Yang et al., 2019]) Traditional visual recognition (passive) attempts to recognize the object present in the image and often fails in cases where the object is heavily occluded. On the other hand, in embodied visual recognition the agent learns to move appropriately and adjust the view so that the recognition improves.

Research Challenges:

Defence and security-related applications often require real-time inference while dealing with unstructured and unknown environments. Addressing this challenge, we plan to develop recognition methods that leverage minimal supervision and can be integrated with robots and drones. Our previous works such as [Tripathy et al., 2020] where we build an object detector with hand-drawn sketch supervision can be an early direction to approach this problem. In [Tripathy et al., 2020], we have shown to detect objects in the scene by just drawing an abstract representation of the object (i.e. sketch). We plan to extend this work where the robot has never seen an image of an object, but it has access to sketches of the object.

3.2 Incremental 3D Reconstruction

Introduction: The fundamental problem of efficient and automatic navigation of autonomous systems in the war zone is mainly tackled through the 3D reconstruction of the war zone or territory. Furthermore, the problem of intelligent recognition of objects present in the reconstructed 3D model of the surrounding environment is a fundamental problem to solve for efficient navigation of the UAVs to avoid obstacles and taking the critical automatic decisions in the war zone. Therefore, the problem of 3D reconstruction of the surrounding environment should be solved accurately. The war zone surface layouts consistently keep changing, e.g. the surface nature of the desert areas changes very frequently and significantly due to the movement of the sand dunes by the high-speed winds. Therefore, we can not rely on the previously reconstructed 3D surface models and have to reconstruct real-time 3D surface models. Furthermore, the vision-based sensing devices, e.g. stereo cameras or LiDAR scanner, are installed on UAVs that always explore new areas. Therefore, we have to follow an incremental 3D reconstruction approach. Furthermore, UAVS for defence applications always move at a very high speed. Hence, we have to reconstruct the 3D model of the surrounding regions in real-time for better navigation of autonomous systems and recognition of the objects present in the surrounding region.

Research Challenges: These approaches heavily rely on image features and perform very well on architectural scenes, which are rich in features. However, the problem of reconstructing the 3D surface of scenes like the desert and mountains is a challenging problem due to the smooth nature of desert surfaces containing negligible features and nonrigid motion of trees and occlusions in the mountains. Therefore, existing approaches cannot be used directly to solve the incremental 3D reconstruction problem in regions like desert and mountains. Furthermore, most of the defense applications require performing perception-action cycle in almost real-time in the unknown and unstructured environments which is an important and open problem to solve.

Research Problems: Under this TIH, we will solve the problem of incremental 3D reconstruction of the surrounding region with very high accuracy and a very low computational time complexity. The major challenges in solving this problem are the high speed of UAVs, continuously changing surface structure and very smooth surfaces with no significant features. There exist many approaches that address the incremental 3D reconstruction problem using the structure from motion. However, they assume that the sensor data for the entire scene is available. Then, they find the best pair of images among all the pairs from the scene graph to start with to overcome the problem drift and error accumulation while solving the bundle adjustment problem. However, in the problem of 3D reconstruction for the defence application, we do not have access to the sensor data for the entire scene and which is acquired

sequentially. Therefore, under this area, our goal is to propose fast (log-linear in the number of points) and accurate algorithms for incremental 3D reconstruction from stereo image data, partial LiDAR 3D range maps, and motion data from event-based cameras. Furthermore, we will solve the problem of real-time object detection and recognition in the multimodal data acquired from these sensors. We will further propose vision systems for sensing and reconstructing the objects present very far away from the UAVs to anticipate the movements and actions. In the figure, we show the overall of the incremental 3D reconstruction problem. In summary, we would like to solve the below research problems.

1. Fast and accurate techniques for real-time incremental 3D surface reconstruction of the defense application zones from images, range scans, and asynchronous vision sensors data.

2. Fast and accurate techniques for real-time detection and recognition of objects present in the reconstructed 3D surface models.

3. Vision-based sensing systems for reconstructing, recognition, and tracking objects present at a very large distance from the autonomous system.
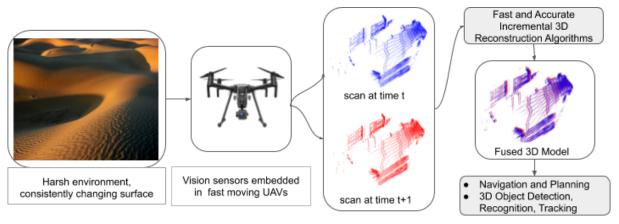
Figure Incremental 3D reconstruction for fast and accurate navigation and planning. (Images Credits: Wikipedia, DJI, A2D2).

3.3 Intrusion Detection under Challenging Scenarios (Prof. Guha, IITG)

Introduction: Video surveillance and analytics of surveillance videos attempt to solve a large number of problems in Computer Vision. Surveillance systems generally focus on intrusion (possible target) detection, target (object of interest) recognition, (multiple) target tracking, and their activity analysis. Intrusion Detection is the first stage of most video surveillance systems. Also, a large number of video surveillance systems only focus on intrusion detection with intrusion triggered video recording. This proposal focuses on the development of Intrusion Detection Systems (IDS, henceforth) for challenging scenarios.

The North-Eastern region of India and her border has always witnessed army deployment on account of its border with multiple nations. This region has hilly terrain with dense forests and rivers. Video surveillance in such regions is naturally challenged by a number of scenarios. The proposed IDS aims to overcome a few hurdles of visual surveillance under adverse conditions. These are described next.

Challenging Intrusion Detection Scenarios: Partial or complete Occlusion of intruders is caused by the Thick Vegetation in dense forests containing elephant grasses, bushes, and tree trunks.

However, the intruders can still be sensed through their partial visibility (during their self-motion), generated noise, heat signature or induced (anomalous) motion in the vegetation.

This problem is further worsened by different kinds of Poor Visibility Conditions. Poor visibility can be attributed to the following conditions. First, Low Ambient Light in forests, around hills, or during the night. Second, Heavy Rain and Thunderstorms frequently occur in different parts of the North-Eastern states. Third, seeing through Fog, Smoke, Snowfall or Low Clouds is a hard problem for video surveillance applications. Such poor visibility conditions pose serious challenges to both near and far-field visual intrusion detection (be it by man or machine). Additionally, the intruders do not appear wearing contrasting clothes. In most cases involving genuine threat, the intruders will Camouflage themselves to avoid detection. It is especially proven to be hard to detect intruders camouflaged against backdrops of snow, forest, or barren hills.

Each of these individual problems and their combinations poses challenges to visual surveillance in regions with mountains and forests. The goal of this proposal is to design a system that operates under the challenging conditions of Occlusions, Poor Visibility, and Camouflage. We will also create a multi-modal dataset capturing various aspects of the challenging scenarios. The dataset will contain image sequences recorded using RGB Cameras, thermal imagers and LiDAR sensors

Figure. (Horizontal to vertical) Illustrating the challenging scenarios arising out of (a) occlusions in dense forest; (b) low light conditions; poor visibility due to (c) fog and (d) low clouds; potential detection failures due to camouflage against the backdrop of (e) hill or trees; and (f) snow.

3.4 Suspicious Event Recognition with UAVs

Introduction: The primary objective of this proposal is towards automated visual detection and recognition events in geographically challenging regions (like, dessart, mountain, forests etc.). Since building up a proper static camera-based network is not always feasible in those

regions, we plan to develop the system with unmanned aerial videos (UAV) surveillance. It is going to be primarily applicable to those areas where manual surveillance is quite challenging. We then aim to extend this work to the recognition of some suspicious activities or events. Different machine learning and deep learning techniques will be explored here to infer the unconstrained videos.

Figure: Example General Event-recognition from videos: (a) person running, (b) person standing, (c) car moving.

Research Challenges: The challenges those are going to be addressed in this project primarily include:

1. Recognition of events from videos captured with UAVs
2. Recognition of events in the geographically challenging regions (like dessart, mountains, forrests) and
3. Detection of suspicious events from those videos
4. Prediction of events

Outcome: Here in the proposed project, we planned to develop methods of event recognition from videos captured with cameras mounted to UAVs. Our aim is to develop vision-based automated surveillance systems for geographically challenging locations. Building up standard surveillance systems with static camera networks may not be possible in such places. That is why we planned to deploy event recognition methods for ego-centric videos.

Event recognition from ego-centric videos is the first task to be done here. As we know, the motion of the drone may not be uniform always, therefore, tracking movements of moving object(s) with relative velocity may fail. Besides, proper recognition of objects with aerial views needs to be performed accurately. Scalability is another challenge associated with this approach due to variations in drone height. We are going to address the challenges with newly developed methodologies in the project.

Figure: Different Geographically Challenging Regions: (a) Dessert, (b) Mountain, (c) Jungle; Solution? Drone-based Surveillance

Figure: Example Event Videos Captured with Drone Mounted Cameras in ERA: (a) Traffic Congestion, (b) Fire, and (c) Concert

We shall then focus on finding out the abnormal activities like left luggage for explosive attacks, theft, running crowds, fights, vandalism, and crossing borders. These are usually unusual or suspicious activities rarely performed by humans in public places. We aim to use multispectral video analysis and spectral reconstruction to come up with effective solutions in the proposed project.

Figure: Example Suspicious Events: (a) Baggage left, (b) Fight, (c) Theft

3.4 Biometrics in Adverse Conditions

Person authentication and recognition is a very important application for security. Biometrics is the field of study that deals with identifying humans based on their physiological and behavioural traits [Jain et al., 2007]. There are several different biometric modalities that have been explored in the literature, for instance, face [Zhao et al., 2003], fingerprint [Maltoni et al., 2009], iris [Daugman, 2004], and palmprint [Kumar et al., 2003]. Out of these, fingerprint and iris modalities are the most reliable but they require specialized hardware and expect a relatively higher degree of user cooperation, whereas, a face image can be collected by placing a commercially available imaging device within a distance of up to few meters of the subject. Further, it is possible to capture face images without restricting or causing discomfort to the

user, by utilizing devices such as CCTV cameras, as shown in the Figure below. This non-intrusiveness is an important characteristic in various application scenarios such as law enforcement and surveillance where the user may be free moving and is not necessarily cooperative. Thus, for adverse conditions, face recognition is one of the most suitable modality.

Figure: Illustrating the procedure for acquiring face, fingerprint, and iris samples. (a) Face image is acquired using the front-facing camera of a handheld mobile device. (b) Optical sensor based fingerprint acquisition device used as part of the US-VISIT (Visitor and Immigrant Status Indicator Technology) program. (c) Iris image is captured using a specialized handheld device.Image Sources: goo.gl/htrLxZ, https://goo.gl/tmH5vE, https://goo.gl/7riXQC

Figure: CCTV cameras can be installed for the purpose of surveillance. The images show CCTV cameras installed at the entrance of a premise to keep record of the visitors

Face recognition is a task that humans perform every day swiftly, accurately, and repeatedly. However, when automated algorithms are used to perform the same task, it becomes a challenging research problem that has received proportionate attention in the literature. Even after decades of active research in this area, face recognition algorithms struggle to achieve the consistently accurate performance of the human mind for familiar faces. The earlier research majorly focused on addressing each covariate, such as pose, illumination, and expression (PIE), individually. The research succeeded in demonstrating the potential of face recognition for various well-controlled scenarios. In the last few years, researchers have been broadening the scope of face recognition to increasingly uncontrolled scenarios. For example, law enforcement related application scenarios such as surveillance, assume very limited control over user or environment.

Security and defence applications not only require constrained face recognition, but a major requirement is driven from unconstrained scenarios, such as monitoring and identifying individuals in refugee camps, border crossings, and identifying persons of interest in surveillance videos. Most public and sensitive spaces including airports, railway stations, government buildings, and bus terminals are equipped with CCTV cameras in order to monitor

daily movements. The query face images in the application scenario of law enforcement are likely to be acquired in an unconstrained environment without user cooperation. For example, in case of surveillance in public places, such as metro stations, people may be moving, parts of some faces may be occluded, and there can be non-uniform illumination and various imaging artifacts present in the acquired sample. In such scenarios, along with face, other biometric attributes like gait can also be used, integrated with contextual attributes like kinship analysis, race, age and clothing.

Under this TIH, we plan to create a holistic person recognition system that extracts all the identifying cues of the person and combines them to make several predictions, like who the person is, what is he/she doing, and expression analysis. In the current literature, there is research ongoing towards individual aspects, for instance, independent research on face recognition, gait recognition and attribute analysis. However, there are very limited efforts towards combining these modalities in real-world situations, in a context-sensitive manner. Appending the recognition solutions along with attribute analysis is also not explored for real-world applications. The figure below provides an overview of what we plan to achieve in this TIH in terms of person recognition via biometrics and attribute analysis.

Research Challenges:

1. Face recognition in unconstrained scenarios
2. Identifying and tracking the person in surveillance scenarios
3. Designing multimodal biometric solution for person recognition in unconstrained scenarios
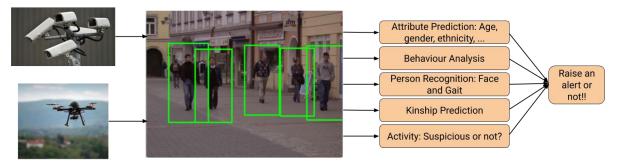
Figure: Illustration of activity recognition and attribute analysis in a surveillance environment

3.5 Hypocognition

Hypocognition refers to the lack of understanding a concept due to the absence of linguistic or cognitive representation of the concept. For example, cultural references to certain emotions, colors, and words are not often completely translatable into other languages, and hence create a barrier in full appreciation of the concept or the emotion. In terms of psychology, hypocognition could be understood as people's finite conceptual horizons that constrain the understanding of concepts for which a prior instance or schemata is not available. As for people and their understanding, this concept can be faithfully extended into AI-driven intervention for defence applications particularly in unknown terrain or circumstances.

Research Challenges

From a data-driven machine perspective, hypocognition complicates what it really means to have "knowledge" because it refers to where the thought or understanding goes even when

one fails to have a representation of the concept. A vision system in a completely unknown terrain may not be able to recognize the exact nature of the terrain, people, atmospheric conditions. Its decision-making capability in such an unknown and rather incomprehensible situation is an object of great interest for next-generation AI systems.

Assume an instance of international conflict or local surveillance where certain actions or objects are predetermined to denote actionable items, eg. act of aggression or trespassing, carrying weapons in unauthorized areas, etc. With extensive data analysis, training, modelling, and testing using state-of-the-art machine learning tools, answers to the following questions may be sought:

1. Given a set of circumstances and previously unknown consequences associated with the circumstances, how does the AI system understand the numerous subtle nuances that may not have had any individual previous associated distinct consequences, but as a cumulative effect denote an act of security breach?
2. As the AI makes decisions purely on the basis of prior information, it is able to identify its own hypocognition in an unknown environment? Or does it suffer from the 'totality illusion' that works around its existing knowledge to arrive at a decision?
3. Can the AI-based vision system be trained to identify hypocognition, and be trained to reverse it, especially in a defence/surveillance-based application? Eg. Can a humanoid robot soldier in combat be sensitized towards taking a deeper and more nuanced decision depending on multimodal perception, microexpressions, linguistic constructs (that may be culturally different), etc. to avoid unintentionally casualties?
4. Can the data, despite hypocognition, be compiled and organized (eg. satellite imagery, terrain information, multimodal sensor data) (Trepagnier et al., 2007) to create accurate situational awareness by applying deep learning, statistical analysis, and probabilistic algorithms (instead of next-best recommendations)? The same could be used during practice training of troops and in actions for peacekeeping
5. Can the system assist in cybersecurity against DeepFakes through new linguistic and multimodal constructs?

Potential outcomes include new super AI-models that use multimodal data to assist humans in making complex inferences on a wide range of defence applications ranging from surveillance, security on known terrains, training of peacekeeping forces, as well as international diplomatic relations, using multidimensional factors and data and possible decision paths into consideration.

3.6 Technology Fallouts

1. Person recognition system in constrained and unconstrained environments

2. Behavioral analysis of individuals in constrained and unconstrained environments

3. Person tracking, recognition and re-identification via camera sensor networks

4. System for describing unseen objects and scenes

5. Recognition system that leverages minimal supervision and can be integrated with robots and drones

6. Fast and accurate techniques for real-time incremental 3D surface reconstruction of the defence application zones from images, range scans, and asynchronous vision sensors data.

7. Fast and accurate techniques for real-time detection and recognition of objects present in the reconstructed 3D surface models.

8. Vision-based sensing systems for reconstructing, recognition, and tracking objects present at a very large distance from the autonomous system.

4. Navigation

With the advancement in technology and wider applications, there has been significant interest in developing autonomous systems with collision-free navigation in a cluttered environment. There are several other operational and environmental conditions such as the presence of varying weather and illumination, presence of static and dynamic obstacles, different shapes and orientation of the obstacles etc., that pose challenges towards a trusted and safe navigation. In recent years, the development of efficient navigation systems is made possible with advancements in technological developments of small size but efficient gyros, accelerometers, magnetic sensors, servo controls, communication equipment, GPS, power plant and payloads consisting of a camera, IR sensor, etc. The motivation and driving force behind this effort is the need to fill up the surveillance blinds spots left by today's satellites and spy planes, especially in emergencies. These systems offer the advantage of portability. Hence, it can be carried near the mission area for its deployment and information about the severity of causality can be easily gathered. Apart from the surveillance in emergency scenarios, this autonomous navigation with small payload capacity will be useful in law enforcement, defence applications, agricultural survey, forestry, environmental monitoring etc.

One of the key factors that lead to the effective application of robotic technology is robotic autonomy. Supervised autonomy and full autonomy are two such paradigms for robotic navigation and control that can be adapted depending upon the requirement in an application. The defense applications such as safe diffusion of a bomb and tracking and monitoring of sensitive regions where expert/domain knowledge needs to be leveraged, supervised autonomy is preferred. While repetitive tasks and dangerous missions may be accomplished using full autonomy among robots. However, achieving autonomy in the real dynamic world remains an extremely challenging problem because of the operating conditions that will require the robots to adapt to the environmental dynamicity. Thus, as far as applications in defense missions are concerned, both supervised and full autonomy are of equal importance depending upon the specific applications.

Once the team of ground robots reaches its destination, the second stage of operation will be taken over by the team. However, for many important applications in the defence sector as mentioned above, robots need human supervision for executing a teleoperation task. This is where teleoperation will help in guiding robots/agents from a distance towards accomplishing specific tasks. Such human-machine interaction will be enabled via multimodal sensing at the site followed by latency-free communication of the signals. Possible tasks benefiting from teleoperation include on-ground vehicle control, user-assisted navigation, surveillance/reconnaissance, detection and removal of landmines, and bomb diffusion, to name a few.

We note that the mentioned scenario falls under a very popular field named telepresence and teleaction (TPTA) systems. TPTA systems allow a user to be present and active in a remote environment. Such systems are very useful for defence applications, for example, in case of bomb diffusion and providing medical facilities to the defence personnel using telemedicine. A typical TPTA system will consist of a human system interface connected to an operator, a communication medium, and the teleoperator/robot, as shown in figure below. The human system interface has input devices for the position and orientation sensing, and output devices for displaying multiple modalities (i.e., vision, auditory, haptics, etc.). An operator is connected with the human-system interface and commands the position/velocity of the teleoperator while observing the remote environment through the multi-modal feedback. The communication medium is used to transport the multimedia streams bidirectionally.
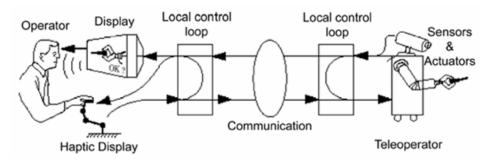
Figure: A Telepresence and Teleaction Scenario. Figure source: Hinterseer et. al 2008.

Additionally, many defense applications require the transmission of real-time images at high data rates and in a secure manner. Wireless communication in an urban environment is quite challenging due to the frequent obstructions and occlusions by buildings and other man-made obstacles. UAVs can enhance communication reliability in such an environment by acting as relays to assist the existing communication systems. Unlike traditional fixed ground relay, UAV relaying can adjust its location to cope with the change of communication environment. This advantage motivates many research works on UAV-assisted relaying systems in an urban scenario with potential defence applications.

Providing theoretical as well as experimentally tested solutions to the problems pertaining to autonomous navigation via single and multi-robot systems, comprising various types of ground and aerial robots, communication, teleoperation, embedding computer vision is one of the prime technical activities planned under this track. The thrust areas of autonomous navigation are illustrated in the Figure below.
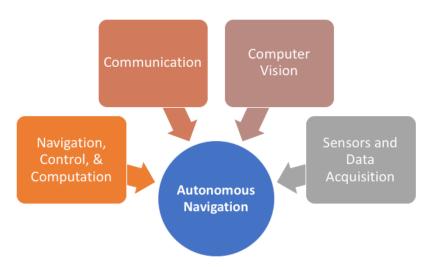
Figure: Thrust areas of autonomous navigation

4.2 Visual Sensing based Autonomous Navigation using Single and Multiple Agents

The navigation by a team of multiple robots is preferred over a single robot because of the robustness and redundancy of the multi-robot systems. The cooperation among robots is established by their onboard sensing. Multi-agent or multi-robot systems, for instance unmanned aerial vehicles (UAVs), is one paradigm of such systems that have recently emerged as an enabling technology to facilitate a diverse range of applications such as inspection and monitoring, border patrol missions, forest fire detection and monitoring, surveillance, data collection, and crowd monitoring.

It is of prime importance in such operations to have an understanding of the surroundings so as to avoid obstacles or any other unwanted circumstances while performing a cooperative task. The scene understanding and creation of the map is a crucial step for autonomous navigation for path planning, obstacle avoidance. The importance of the visual information-based depth-sensing arises due to erroneous GPS information or absence of it. The reliable availability of such a framework will result in the irrelevance of visual information. However, improvement in precision and availability of GPS information cannot be guaranteed due to occlusion. Also, a multi-agent system requires a failover mechanism as a non-functional or erroneous operation of the agent should not affect the functioning of the entire system.

The key aspect of navigation is precise localization, which is one of the primary requirements for autonomous robotic systems. In both indoor and outdoor environments, autonomous robots should have their exact location and orientation to perform required tasks. For effective navigation, an autonomous system must be able to quickly and accurately determine its location. There are various techniques to solve localization using different sensors. The key idea of most of the current localization schemes is to combine different measurement data to obtain better positional information. In general, most of the approach has been developed for the case of a single robot. However, for multi-agent systems, solving a collective localization problem is quite challenging.

Navigation of defense force vehicles in poor visibility weather conditions including that of dust/sand-storms is also a significant problem in this context. We can take an example of security of armed forces deployed in dealing with naxalite problems. The security vehicles should be camera enabled to record the events around it. This will be based on the learning from videos of the known terrains in (a) good weather condition and (b) artificially-created poor

visibility weather conditions. The information available based on this learning can be useful when a vehicle moves in actual poor visibility weather conditions.

Research Goals/Outcomes: The proposed research will rely on the tools from multiple disciplines including: dynamical systems and nonlinear control theory, graph theory, signal processing, and network optimization theory etc. Broadly, the TIH aims to offer solutions to the following challenges pertaining to autonomous navigation using single and multi-vehicle systems:

a) Path Planning, Obstacle Avoidance, Vision-Based Navigation in GPS Denied Environments: Navigation in an unknown environment requires creating a map and identifying the obstacles. The appropriate path for navigation and motion can be planned accordingly. Vision can play a crucial role in this respect. Visual information together with information from other available sensors is utilized to update/add the explored map. The information from the vision-based sensor is complementary to other sensors. This is useful in case of the unavailability of other sensors e.g. GPS due to buildings and indoor environment. The combined information can also be utilized to correct the errors in data obtained from individual sensors.

b) Collective Localization: For improving positional information of individual agents, each member of the team will exchange their local information with their neighbouring agents. Then, the individual and neighbours positional information will be fused to obtain more precise positional information. Now to execute this conceptual method, it is necessary to exchange the information among the neighbouring agents. Therefore, it is also within our focus to formulate the problem in such a way that it will allow for distributed processing with minimal communication requirements. Initially, to understand the concept, the proposed method will be validated in the simulation environment. Then to realize the proposed localization technique in real-life, we will use both ground and aerial robotic platforms. The analysis will rely on theoretical tools like Kalman filter, distributed computing, an optimization technique, task scheduling, event-triggered sensing, etc.

c) Source Seeking: Detecting the source of a signal is relevant to many defence applications, such as search-and-rescue operations, odour source detection, sound source localization etc. In these applications, the vehicle collective makes a movement by estimating the intensity of the signal source signal. The problem is challenging if the source is moving and its intensity varies in spatiotemporal fashion. In these applications, the operation of vehicles in a formation will help in accurately estimating the strength of the source signal and hence reaching to it precisely. Secondly, when the vehicles are in the vicinity of the source, vision information can also be infused for early and precise convergence. Of course, in such a dull, dirty and dangerous mission, there are many obstacles of static or dynamic nature that should be avoided towards the completion of the mission.

d) Vision- based Road Condition Monitoring and Safety Alert System: Having a reliable input on the road condition can be enormously beneficial in reducing the number of accidents in unstructured/cluttered environments. The main challenges are: i) A system for safe driving based on detected damages, impediments and other road conditions including cases of unexpected objects encroaching the safety range of the vehicle ii) Incorporation of the Road Quality score in path recommendation system such as Google Map for driving. To address the aforementioned challenges, a system incorporating the use of cameras, accelerometers, and proximity sensors is proposed. The acquired video, the accelerometer signal, and the proximity

sensors will be automatically analysed to address the above issues. We aim to develop Algorithms to analyse the recorded data. In this direction, the data from independent sensors (camera, and sensors for tilts/jerks) will be employed to provide information about the unknown environment.

To emulate the foggy condition, under clearly visible weather conditions such as summer, films of different levels of transparency can be identified and put on the lense of one of the cameras. Different films will emulate different fogg densities and video recorded by the camera will depict as if it was recorded during actual foggy weather conditions. The second camera will acquire actual scenes that the Driving sees during clear weather conditions. In this way, foggy video is obtained as mentioned above. These two videos can be used for learning enhancement algorithms using appropriate machine learning techniques

e) Personalization in Autonomous Driving and Driving Assistance Systems (Prof. Chetan Arora, IIT Delhi): One of the active areas of research in self driving cars is vehicle to vehicle and vehicle to pedestrian interactions. To achieve the desired safety levels, the ego-vehicle must be able to predict how the other vehicles present on the road at the same time are going to maneuver, or how the pedestrians, who may be interested in crossing the road are going to behave. We note that many of these behavioral patterns are specific to countries, cities, and communities. Similarly, visual appearance of various objects appearing in the scene, such as road, sidewalk, vegetation, buildings, and clothing of the pedestrian change depending on the location and associated weather conditions. Human beings have remarkable abilities to adapt to the context which is missing in modern techniques. We would like to focus on techniques where such location specific appearance structures, visual features, rules and behavioral patterns could be learnt from the data itself.

f) Personalization in Human-Machine Collaborative Systems (Prof. Chetan Arora, IIT Delhi): One of the biggest application areas of computer vision is Autonomous Systems. With the advances in technology, the autonomous systems are no-more operating in a stand-alone fashion far from the human operators. But in collaboration with the humans, and sharing the same workspace. Similar to the context of self-driving cars, it is imperative for the autonomous systems to anticipate a human action and then act accordingly. While many of the rules for human-machine collaboration can be codified, it is important for an autonomous system to learn person specific traits to effectively collaborate with its human operator. We would like to explore whether such operator personality traits can be learnt from the visual data alone.

4.3 Visual SLAM

Introduction: Simultaneous Localization and Mapping (SLAM) is the backbone of most of the autonomous systems for efficiently navigating. In the defense applications, the navigation of the UAVs or UGVs is mainly done through the SLAM techniques which use the visual data of the surrounding environment. The localization of the autonomous systems and the map of the surrounding environment is used for tracking and targeting the enemy targets. Accurate localization of the autonomous systems and efficient representation of the surrounding environment determines the success of the SLAM based navigation systems. Furthermore, omnidirectional SLAM has become a very important technique for defense applications.

Research Challenges: The open challenges in the SLAM from the perspective of defense applications is the accurate localization of the autonomous system in the dynamic, unstructured, and unknown environments. A small error in the localization of an autonomous

system can lead to missed targets or fatal accidents. Another research challenge is to efficiently represent the reconstructed map of the surrounding environment for performing more efficient actions. Instead of simply representing the surrounding map by a 3D point cloud or a topological mesh, a semantically meaningful topological representation will play a critical role in efficient navigation and performing cognitive action. Furthermore, most of the existing algorithms solve the SLAM problem for indoor applications or city navigation. However, defense application areas are dynamically changing and are unstructured.

Research Problem: Under this TIH, we will develop fast and accurate algorithms for solving the SLAM problem in dynamic and unstructured environments. Furthermore, we will propose deep learning-based algorithms for learning semantically meaningful maps while simultaneously and accurately localizing the autonomous systems for performing efficient actions. For an accurate localization of the autonomous systems, we will use the multimodal sensor data such as visual or 3D scan data and data acquired through GPS and IMUs. In order to make the proposed algorithms fast, we will use the Direct and Sparse reconstruction approach. We will further propose better techniques for solving the SLAM problem for the omnidirectional vision sensors.

4.4. Future Prediction From Video (Prof. S. Das, IIT Madras)

The problem of predicting multiple future frames given a video clip has gained enormous interest in ComputerVisionand Machine Learning Communities over the past few years [Mathieu et al, 2015; Lotter et al., 2016, Bhttacharjee et al., 2017, Xu et al., 2018, Kwon, et al, 2019]. This problem is important as human behavior is immensely guided by anticipation or prediction of the future based on information from the past and present. Thus, the ability to generate future frames from an input video demonstrates the ability of a machine to understand the visual content and predict the future conditioned on the given input. Potential applications of this problem are improved decision making for self-driving cars/vehicles, self-guided planning for action of robots with visual sensors, the anticipation of illegal events in surveillance scenarios, unsupervised feature learning (by modeling the non-trivial spatio-temporal relationships among the scene elements), video compression, etc.

The problem of video prediction is intrinsically ill-posed, compounded by the lack of availability of proper ground truth. Even though most works assume the future to be deterministic and thus predict a single future, the possible future is generally not unique. Also,for most prior published work, thegeneratedfutureframesaretypically blurred and predicted only for a short duration. Therefore, the focus of this plan of work would be to design methods that will reduce the amount of blur in the predicted videos, generate prediction with semantic content and provide multiple tracks of potential probabilistic futureframes.

4.5 Bathymetry Mapping with Obstacle Avoidance using an Autonomous Surface Vehicle (ASV) [Prof. P. B. Sujit, IISER Bhopal]

Motivation: With increased marine activities along the coast, there is a need for additional marine monitoring resources to perform persistent surveillance. Manually performing these long range operations along thousands of kilometers is difficult. Autonomous vehicles with their ability to perform remote operations while enabling human-in-the-loop operations enables the operators to monitor long range missions with minimal operator overhead. A key

application is detection of mines in the shallow water of the ocean. Performing manual boat-based missions is costly and time consuming. This proposal focuses on developing an autonomous surface boat that is capable of navigating in the open waters autonomously and perform bathymetry operations to detect mines in the sea bed. In order to perform autonomous operations, several onboard modules need to be built.

The primary aim of this proposal is to develop an autonomous surface vehicle (ASV) that is capable of performing autonomous navigation in the presence of obstacles for bathymetry mapping in shallow waters. We will be constructing a catamaran ASV and automate its operation using off-the shelf open source autopilot. We will develop novel navigation algorithms for the boat to navigate in different environmental conditions while avoiding incoming obstacles that may include rocks, boats and other floating objects. The obstacle detection will be performed using an onboard camera and novel algorithms to detect the objects robustly under different lighting conditions and sea states will be carried out. The boat will be equipped with a side-scan sonar for bathymetry mapping that continuously acquires the seabed information. Such an ASV will be highly useful for Mine-Counter-Measure operations, habitat mapping, etc.

Objectives: The objectives of the proposed project are

- Construct an autonomous surface boat (ASV) using several off the shelf components
- Develop autonomous navigation capabilities
- Develop robust vision based obstacle detection and avoidance techniques for different environments
- Efficiently map building for a given region while avoiding moving obstacles.
- Develop informative path planners to acquire efficient target information

Relevance and importance of the project: India has a long coastline that needs to be protected. This project plans to provide a capability for Indian maritime authorities to operate an ASV autonomously in the ocean under different environmental conditions. Especially in unstructured environments. The sensing systems consisting of cameras can acquire reasonable sensor information about the environment to steer away from any incoming potential threats. The threats could be both incoming boats and stationary/floating objects. The outcomes of this project will add new capability to the existing surveillance systems of the maritime forces.

4.5 Haptics-based Supervised Navigation - Teleoperation System (Jointly by IIT Jodhpur and Prof. Subhasis Chaudhuri, IIT Bombay)

Touching an object causes rich haptic cues that enable us to understand the object's physical properties and adeptly control the interaction. Although human experience centers on physical contact with tangible items, few computer systems provide the user with high-fidelity touch feedback, limiting their intuitiveness. Haptic interfaces are mechatronic systems that modulate the physical interaction between a human and his or her tangible surroundings. Such interfaces will typically involve mechanical, electrical, and computational layers that work together to sense user motions or forces, quickly process these inputs with other information, and

physically respond by actuating elements of the user's surroundings. Well-designed tactile feedback can greatly increase the realism of virtual and remote worlds. Haptics has a lot of interesting applications in the defence sector, in particular, for teleoperation kind of applications.

In recent times, both AR and VR have received significant research and industry attention because of bringing a certain amount of realistic and immersiveness into end-user's experience while interacting and using digital technologies. These emerging technologies have demonstrated applications in healthcare, tourism industry, education and entertainment (e.g., digital cinema). In addition, augmented/virtual reality represents key technologies to design, simulate and optimize cyber-physical production systems such as TPTA and interact with it remotely or in a collaborative way. In this TIH, we will also focus on using AR/VR technologies along with the haptics to create a dynamic and immersive telepresence and teleaction (TPTA) systems for defense applications.

Research Goals/Outcomes

There are various interesting and challenging research problems to realize a realistic and immersive telepresence and teleactions systems such as the good design of manipulators, kinesthetic and tactile sensing, data communication, delay compensating controllers, and immersion into the virtual workspace. The data communication module requires meeting quite a restrictive quality of service guarantee. In line with this vision, we, at IIT Jodhpur would plan the following research activities under the proposed TIH. We also emphasize desirable overlaps between the proposed research activities.

a) Haptic data compression: As mentioned above, haptic information transmitted during a teleoperation task needs to be reduced while maintaining the quality of perception and ensuring the stability of the system. In the literature, perceptually adaptive sampling mechanisms have been proposed to reduce the kinesthetic information. It is still not studied fully how the perceptual deadband is affected by the direction of haptic stimuli, tactile information (vibration information), rate of change of the force and in the presence of the other stimuli like audio and video. In this TIH, we would like to address these issues.

b) Multimodal perception: In a typical teleoperation task, we need to deal with different kinds of stimuli (audio, video, haptic- kinesthetic and tactile). For better execution of the task, it should be known how these stimuli interact with each other. Inter-perceptual relationships among them will help in various tasks like controlling, transmission, navigation, etc. The TIH will focus on these inter-perceptual relationships using psychophysical experiments.

c) QoE-centric communication: One of the key issues in interactive and dynamic teleoperation is that of optimized and latency-free communication. To that end, we will undertake research to develop models for perceptually-optimized communication of audiovisual data. This will be particularly meaningful in scenarios where resource constraints including limited bandwidth can seriously hinder teleoperations due to poor quality of audiovisual streams received. It will also cater to the requirement of larger field of view (FOV) and multi-view synthesis (both will obviously increase required data rates) which are critical for efficient teleoperation.

d) Haptics, AR/VR Multimodal Interface: To develop an immersive AR/VR interface with haptic modality, the following parts will be undertaken:

Scene Understanding: Scene understanding refers to the identification of relevant information so that appropriate action can be taken. For instance, planning how the robot needs to be manoeuvred. In the context of this TIH, the term "scene" is not restricted to visual but can also involve audio, smell and haptic (touch) information. Such holistic information processing will obviously be advantageous over single modality analysis. To achieve our goals, we will undertake fundamental studies to understand the five basic human senses and their interactions with each other. This will involve targeted psychophysical studies using appropriate stimuli and equipment. It will also involve extensive use of novel computer vision techniques (for instance, occlusions can be a major challenge in unstructured and unknown territory). Finally, we will develop computational models that can help in the decision making process about the navigation task/sub-task.

Haptics Modelling and Rendering for Immersive and realistic Remote/Virtual Environment: The results from psychophysical studies (previous point no. 1) will also be exploited for creating high definition haptic modeling and rendering algorithms for interacting with various real/remote and virtual objects. This research activity will help in creating haptic-augmented reality.

Retrieval algorithms for objects mapping: To realize the above-planned steps in a cost-effective and practical manner, we will develop fast retrieval algorithms for correctly mapping objects to their corresponding senses. For example, determination of the hardness/roughness, texture perception of a remote object through visual sensing data.

Based on the above-mentioned research activities, a novel multi-model AR and VR interface will be built for interaction with virtual or remote objects that will be used for teleoperation kind of applications in the defence area.

4.6 Technology Fallouts

1. Development vision-based guidance and navigation algorithms
2. Personalization in Autonomous Driving and Driving Assistance Systems
3. Framework of future prediction from video
4. Robust framework for simultaneous localization and mapping
5. Telepresence and Tele-action (TPTA) system for supervised Autonomy

5. Doing and Acting

5.1. Introduction

Action through Reasoning, Control and Manipulation of Autonomous systems is an important stage while its physical interaction with the environment. Typical defences applications in this regard include collaborations between soldiers and autonomous agents, dexterous manipulation of objects during bomb disposal, vision-based target identification and acquisition, deception operation, assisting and enabling soldiers, mine detection and diffusion, identifying enemy positions using autonomous agents etc. The vision information is playing a crucial role to execute all these tasks. Hence, action is the essence of all such interaction with the environment applying appropriate control law. Thus, Action or reasoning, manipulation and control can be referred to as a mechanism of translating the desired action to obtain a certain output. This acts as a bridge between the initial state to the final desired outcome for a given set of hardware/software system. This involves parameterization of the initial state and final state. The higher-level actions include motion planning, development of optimal control strategy to perform the task based on available hardware constraints and task performing/manipulation. The parameterization of the state is basically performed for a design of the control in terms of Observability. However, these states can be decomposed into various hidden states based on observations or captured data e.g. intensity, feature points, touch etc. The captured data/information acts as a feedback for the underlying system to evolve the state steadily to perform the task by optimally motion planning for the manipulation. The acquired data plays an integral role to make the underlying system robust under disturbance. In case of a completely new task, such an interaction with the environment becomes very crucial. Thus, the optimal acquisition of Data and real-time processing is important in all such cases.

Task-oriented procedure/strategy such as Manipulation and Control, Imitations, Procedural Learning, Visual Servoing etc. can be utilized to perform various tasks in different applications such as Task Training, Maneuvering in a complex unknown environment in variable conditions, Servoing etc. Only different components of the basic task execution change with the variation of the task and the system. Figure below shows a typical functional diagram for a task execution system. Thus, in this part of the iHub Drishti, we will focus on developing such general strategies, algorithms which can be applied/adapted for various defence applications using the knowledge of previous developed strategies for similar kinds of tasks. The following section summarizes some of the focus areas in which the iHub Drishti plans to contribute towards.

We have divided the focus areas in 'Action' into following sub-areas based on possible strategies to be followed for the execution of the various desired tasks for the given anonymous systems. Figure below summarizes the various strategies categorically in the Action part of iHub Drishti.
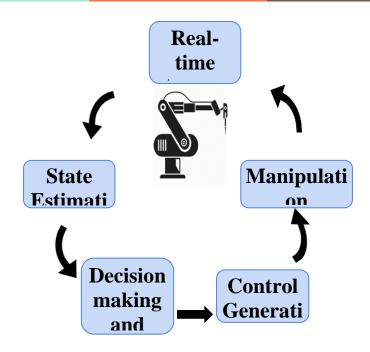
Figure: Basic Task Execution Strategy
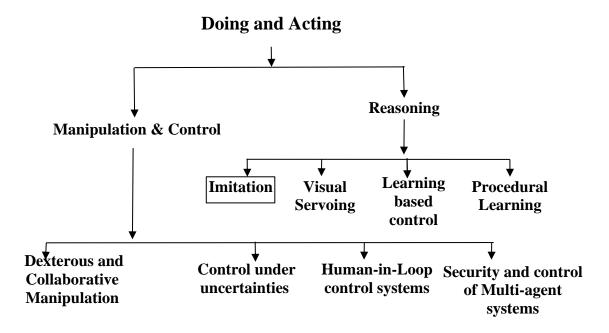
Figure: Various Strategies in Action.

The 'Manipulation & Control' basically refers to a predefined coded task category. Whereas, `Reasoning' refers to the tasks where a visual demonstration of the task is available to follow and learn. These two are major focus areas for the action part in iHub Drishti. We describe each of the sub-areas in the details in the subsequent sections. In the following sections, we describe the strategies in detail for focus areas shown in the figure above with the introduction and discussion of challenges and limitations of the state of the art methods.

5.2. Manipulation and Control through Vision Feedback

There is now a major international push to design autonomous systems that should interact with each other and the physical world with cooperative objectives. This is achieved through advances in wireless communication, the internet, and the introduction of new and cheap vision-based sensing devices. Such systems are usually referred to as cyber-physical systems (CPS) and are the cornerstone of the present social and technical infrastructure. Autonomous systems, for instance, unmanned aerial vehicles (UAVs) and remotely operated vehicles (ROVs) and manipulators, is one paradigm that has recently emerged as an enabling technology to facilitate a diverse range of applications. Representative examples include, but are not limited to, aerial photography, inspection and monitoring, border patrol missions, remote driving of unmanned cars, collaborative manipulations, pick and place operations, forest fire detection and monitoring, surveillance, data collection, agriculture, and crowd monitoring.

5.2.1 Vision-based Control of Autonomous Systems

It is typically desired that the autonomous vehicles move in a geometrical configuration so as to complete the overall group objective effectively, by using their onboard vision-based sensing. For instance, [Leonard et al., 2007] discusses that the data collection is optimized by mobile sensors operating in a circular formation. There are many other works where various formation patterns have been explored depending upon the nature of an application please refer to the survey papers [Oh et al., 2015, Cao et al., 2012].

Because of the practical reasons as mentioned above, the study of the vision-based distributed formation control of autonomous unmanned vehicles, in recent years, has been of great interest among researchers from interdisciplinary areas. Providing theoretical as well as experimentally tested solutions to the problems pertaining to formation control of autonomous multi-vehicle systems, comprising various types of ground and aerial robots, is one of the prime technical activities planned at the TIH-Drishti at IIT-Jodhpur.

Research Goals

Broadly, the TIH aims to offer solutions to the following challenges pertaining to autonomous multivehicle systems:

- Vision-based Distributed Control for Autonomous systems
- Distributed multi-sensor data fusion for vision based sensors (DYSL - AT)
- Resource-Constrained Controller Design for Autonomous Systems with Event-triggered Visual Sensors

5.2.3 Human-in-loop Control for Vision-based Autonomous Systems :

Though advances in autonomous systems, still in some scenarios require human involvement or help to overcome uncertainties in the environment. For this reason, it is necessary to develop robot controllers which can use the benefits of both autonomous agents, adept at handling lower-level control tasks, and humans, superior at handling higher-level cognitive tasks. The key question of the Human-robot interaction domain is how to combine human operator control with autonomous control so that each entity can provide input to carry out

high-level cognitive tasks (that humans excel at) as well as low-level tasks (that automatic controllers excel at).

Research Challenges: We will address the following research challenges:

1. Developing autonomous robots which can navigate uneven complex terrain is not an easy task or often impractical to program into robots.
2. Driving autonomous/semi-autonomous passenger vehicles remotely in both known and unknown roads through vision feedback is a challenging task.

5.2.4 Vision-assisted Collaborative and Dexterous Manipulation

Collaborative Manipulation is designed for safe operation around and in collaboration with humans. It has increasingly become popular in industrial setup due to these high-level interactions between humans and robots. Human robot collaboration has shown promise of increased productivity due to flexibility and repeatability of robots combined with planning and decision making capability of humans. Collaborative robots have evolved as one of the key elements in Industry 4.0. In order to minimize impact due to collision cobots are typically designed to work at lower speed and forces and are made of lighter material. Akella et al. (1999) presented broad design principles for human-machine interaction, and described industrial cobots designed for testing and validation. A detailed review of human-robot interaction and analysis of future trends was presented by Matheson et al.(2019). Collaborative robots are expected to perform difficult and tedious tasks in the presence of humans and participate in higher level collaborative activities. Present day cobots are mainly equipped with force sensors and rely on contact to happen for active compliance in human collaborative environments. However, it is always desired if the robot can understand the actions of human workers through visual cues and learn to adapt to the human collaborative environment around it. This will be one of the critical requirements of collaborative robots, and computer vision combined with learning will play a crucial role in achieving this autonomy. Some of the challenges in this regard involve dynamic communication, optimization, and reprogramming and planning.

Research Challenges:

1. Adapting in collaborative environment through visual cues and learning:

2. Vision assisted human robot interaction for collaborative tasks

Dextrous manipulation is aimed at cooperated movement of multiple arms or fingers for grasping and manipulation of objects. This is a crucial task for a mobile agent when it acts on the environment for doing various tasks. Even though classically dexterous manipulation is designed as precise control of force and motion, computer vision can play an important role in achieving autonomy while doing multi-fingered or multi-arm grasping and manipulation. For example, humans use their eyes predominantly in achieving dexterous manipulation. Furthermore dextrous manipulation through teleoperation aided with AR and VR can help in achieving semiautonomous grasp and manipulation and lends application in robot-assisted remote operation in search and rescue missions. Such teleoperation can also achieve simultaneous control of multiple robots for dexterous manipulation in an industrial setup. Teleoperation imparts robot reasoning ability, creativeness in task operation and intuitiveness in accomplishing a task. Combining multimodal feedback with AR and VR for teleoperated dexterous manipulation with adaptability to learn from uncertainties is the main challenge.

Research Challenges

1. Vision assisted real-time adaptive learning and replanning for dexterous manipulation
2. Learning dexterous manipulation through video demonstration
3. Dextrous manipulation through teleoperation with integrated AR VR

5.3. Reasoning

Reasoning is an important step for selecting appropriate control actions. Reasoning enables autonomous systems to learn from the existing knowledge, demonstration or from previous actions. For autonomous systems operating in unstructured environments, visual feedback plays an important role in achieving the above. The visual servoing, procedural learning and imitation are the key tools to obtain this aim and discussed next.

5.3.1 Visual Servoing

Visual Servoing refers to a technique utilizing information from vision sensors as feedback to control the pose/motion of a robotic system. Figure shows an example of a typical visual servoing system
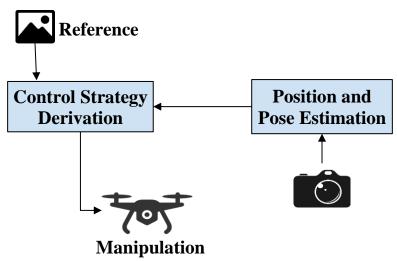
Figure A Typical Visual Servoing System.

The configuration of the visual servoing system can be broadly defined into two modes: eye-in-hand and eye-to-hand based on relative position of the camera. Further, operation of the Visual Servoing System can be subdivided into two major tasks: Sensing and Control. The Sensing in the Visual Servoing Systems can be classified majorly into three classes of techniques viz. Position-based System, Image-based System, and Hybrid System as described below:

The Position-based visual servoing involves estimating the 3D pose of the navigation system to derive the control law for maneuvering. This involves visual scene metrology to obtain the 3D model of the scene. Thus, 3D shape of the system, configuration and calibration matrix is essential in this case. The performance of such a system becomes highly sensitive to the calibration parameters.

The Image-based Visual Servoing system utilizes images directly to derive the control law for the navigation. The predefined pose of the destination or target is utilized as a reference to generate a control law using relative position of destination in the current view of the scene. This type of navigation scheme is useful in applications such as parking, landing and thus, it has a limited scope in application of real-life tasks. Also, the Image-based proportional control has asymptotic convergence and thus requires high time.

The Vision-based Visual Servoing is a hybrid system of Position-based and Image-based servoing to reap the benefits of the both systems by overcoming the challenges. Classical Position-based visual servoing is known as 3D servoing strategy and Image-based visual servoing is known as 2D servoing strategy. Thus, the hybrid approach often referred to as 2 1/2D visual servoing strategy, a term coined by [Malis et al. 1999]. The hybrid approach does not require a 3D model of the object contrary to the Position-based visual servoing scheme. Also, the convergence of the control law is much more robust as contrary to Image-based methods without any inherent assumptions. The hybrid visual servoing system utilizes vision based various techniques to estimate the 3D pose for subsequently deriving the control for the navigation.

Research challenges:

- Reliable depth information estimation under high speeds for creation of depth map in an unknown environment for motion planning by adaptively avoiding the obstacles of various types in GPS-denied areas. Utilization of frameworks such as Action and policy iteration using Deep Reinforcement learning can be useful in such aspects.

- Accurate calibration and transformation of information from camera plane to Servoing System plane to derive a stable control strategy for vision based navigation.

- Semantic scene understanding by Mapping and Labeling for robust maneuvaring.

- Robustness to Dynamic scene conditions such as illumination, indoor-outdoor environment for navigation

- Optimized resource utilization framework for a given task specific accuracy.

Outcomes: In the Action part in iHub Drishti, we would focus on these Sensing and Control challenges to develop a more robust and generalized framework for Visual Servoing systems e.g. Saez et al. 2020 that can be adapted to dynamical scene conditions with intelligent navigation capabilities. The effective vision based depth estimation techniques or effective sensors such as event based cameras can be utilized for this purpose. Such a servoing system can be utilized for mobility of the various indoor systems such as robot/drone based inspection, autonomous carriage, cleaning robots etc. including various outdoor systems.

5.3.2 Procedural Learning

Procedural learning refers to utilization of cognitive framework to perform new tasks based on previous learning. Procedural learning usually involves repetitive performance of an activity to improve the associated learning which is reflected through improved task performance. Such a learning framework is a backbone for conscious machines and futuristic autonomous robotics with broader scopes. The procedural learning framework can be simply defined by using few strategic steps as given in figure.
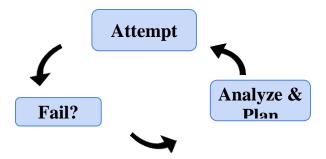
Figure: Procedural Learning Framework.

Research Challenges:

- Failure can be fatal and can result in breakdown of the system. Thus, Optimal selection of the strategy for performing a complex task as well as in presence on limited or Higher DOF is essential. The scene understanding framework along with hardware limitations has to be incorporated to address these aspects.

- Interpretation of the task and decomposition of task with respect to interaction with the external environment. The cognition framework has to be simplified in terms of cause-effect space which is mapped to an action-control space with a scene understanding framework.

Outcome: In the Action part in iHub Drishti, we would focus on addressing these fundamental challenges to develop an autonomous system which can navigate in an unknown environment and perform various tasks such as gripping by interacting with the environment. Such a generalization framework can be further employed in developing new generation games, industry 4.0 etc.

5.3.3 Imitation

The task of imitation is the backbone for future robotics applications. This can be utilized to train the robots the novel tasks very easily by simply demonstration. This will circumvent the requirement of cumbersome task-specific training framework. Such an Imitation framework will require far less computation and storage resources. This will result in optimal resources utilization with real-time processing time. Thus, the resulting robotic system would be a more practical, task robust system with applicability to a wide range of tasks.
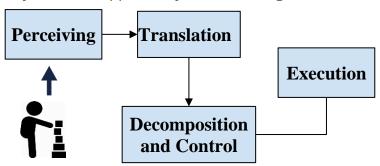
Figure: An Imitation System Example.

The task of Imitation can be simplified in terms of few strategic tasks as given in the symbolic diagram in the Figure. The process can be simplified into perceiving, translation, decomposition, and control strategy and performing. The perceiving involves identifying the states of the task and capturing the relevant information from the scene. The translation involves the translating the real-life task to machine interpretable tasks based on available states or degrees of freedom. The decomposition involves dividing the complex task into simpler tasks. Finally, control strategy is developed based on state information and decomposed tasks are performed based on a feedback based system.

Research Challenges:

- A general task-independent Imitation system which can be scaled up for a complex motion imitation. This can be implemented as a generalized framework for decomposition and policy robustness. Further, scaling up a task for a high DOF imitation system and optimized decomposition of the task is another major challenge in this regard. An action and control space can be utilized for such a task.

- A real-time robust and stable control for imitation system with low vibrations for higher stability and accuracy of the system with optimized resource utilization.

Outcome: In the Doing part in iHub Drishti, we would focus on addressing these challenges to develop a highly accurate imitation framework which can be utilized for general purpose practical applications such as Teleoperation, Bomb Diffusion, Disaster Management, Industry 4.0, Gaming, Aerial Robotics etc.

5.3.4 Cognitive Imitation Learning (Prof. Laxmidhar Behera, IIT Kanpur)

HUMANS are gifted with the ability to adapt in uncertain and dynamic situations while performing a complex task using the a priori basic skills with little or no knowledge of the situations. They generalize and perform novel motions based on their previously learned motor skills. In contrast, current robots can only work efficiently in structured environments separated from humans and are limited to a small number of predefined tasks. The task is hard coded by a human expert after careful analysis of a robot's workspace. Thus, robots are still not able to adapt their skills to novel situations or to learn new motor tasks in real time. Issac Asimov envisioned robots as companions in the domestic environment, workers in automated commerce industries and soldiers in war. The solution to this problem is to learn new motor skills autonomously using prior learned basic skills and improve them over time.

This poses an interesting scientific problem: can the robot observe humans doing such tasks and then imitate? Additionally can we program the robot to exploit the knowledge of its own embodiment and dynamically adapt to moving obstacles, or people working in the same workspace? In this context, learning motor skills through human demonstrations needs to leverage prior information about the objects being picked, robot's movement capabilities and exploit abstract representation of the task that allows generalizing the learnt skill to novel situations. The robotic arm system with a mobile base considered in this proposal has several redundant degrees of freedom and is equipped with a depth-sensing camera to operate in an unstructured environment as well as distinctly identify, reach and manipulate different objects. Using these demonstrator platforms, our objective is to increase robustness and adaptability of autonomous solutions to robotic applications by addressing the following scientific goals:

1. Represent the human demonstrations in a scalable, alternate relational space, that will allow us to transfer the task to heterogeneous robots robustly

2. Exploit offline computation to achieve fast, robust and collision-free motion planning in a dynamically changing environment

3. Enable robots to acquire visual servoing capabilities to reach any target and manipulate different objects

4. Equip the sensing system to visually perceive its unstructured workspace repeatedly and robustly

5. Write software that will allow us to learn, and transfer these skills across various robotic platforms, minimizing the platform-specific setup

6. In the context of robotic automation, our techniques will be validated through experiments and stowing mock-ups. We will use metrics such as success rate, accuracy, amount of demonstration data needed, motion power efficiency and runtime speed.

5.4 Technology and Product Development

Both private and government industries are aggressively working to offer semi-autonomous and autonomous solutions to society, based on large scale interconnected dynamical systems. These classes of systems are widely used in numerous civilian and military applications, ranging from food-monitoring, agriculture, border management, surveillance to cooperative space missions. The TIH at IIT Jodhpur looks forward to providing experimentally tested solutions to various projects in the areas of autonomous systems in unknown and unstructured environments. The research and technological contributions of this segment of work can be broadly summarized as follows:

- Framework for vision-assisted collaborative interaction between humans and autonomous systems

- Algorithms for visual servoing and dexterous manipulation for the purpose of target acquisition or object manipulation.

- The development of distributed controller strategies for heterogeneous multiple vehicles indulges in a cooperative mission.

- Technology demonstration and validation through various experimental testbeds.

- Design and implementation of decentralized algorithms with integrated cyber-security for swarms of autonomous systems.

- Learning based planning and decision making framework for vision based autonomy for task accomplishment.

6. Problems posed by Stakeholders - DRDO Labs

The work proposed under this vertical is aimed at solving relevant problems for the defence sector. DRDO and MHA would be the main user agencies for this work. The end-users have already been approached and feedback on the proposal has been taken. The problems floated by varisould DRDO Labs are given below:

Seeing Through Occlusion (DRDO - CAIR): The main aim would be to develop real-time image enhancement techniques to enable autonomous vision-assisted navigation in adverse weather conditions such as Haze, Fog, Rain, dust etc. The task would also include creation of synthetic imagery to address dynamic occlusion. The objective would be to produce imagery more suitable for subsequent tasks like object detection/recognition.

Asynchronous Sensing (DRDO - CAIR): The main aim would be to exploit asynchronous spatially distributed imagery captured from multiple platforms (swarm of drones) and generate wide-area visual intelligence from the fusion of imagery. Additionally, it should exploit non-uniformly temporal sampled data streams from event based recording sources.

Real time 3D reconstruction (DRDO - CAIR): In this topic, the main aim would be to create a 3D model of the environment in real-time. The outcome would enable real-time situational awareness using unmanned ground and aerial vehicles. The topic can be further augmented with real-time object detection to visualize the objects in the 3D environment. Use of active vision, where the surveillance plan is altered in real time by computed need of the 3D reconstruction in progress, may also be explored.

Tustable Object Recognition (CAIR): While there has been significant progress in the area of visual recognition (both objects and activities), the biggest concerns have been the robustness and efficiency of the AI models particularly with small samples, dirty data and high clutter. Hence, the main objective of this task would be to develop AI solutions that are robust and resistant to deceptive and conflicting inputs (adversarial). This should ultimately lead to establishment or calibration of trust both for usage of learnt models in new environments (fixed AI) and usage of in-field learned models (live AI).

Ground / aerial based autonomous systems for surveillance (DRDO - DLJ): Development of ground / aerial based autonomous systems for surveillance with capability of

- Facial Recognition
- Voice Recognition
- People count
- Number Plate Recognition
- Detection of intrusion
- Moving object detection
- Object missing
- Abandoned object
- Gesture control
- Heat Map
- Smoke detection

Semi-autonomous Navigation (DRDO - CAIR): There have been significant research efforts to create autonomy for air, sea and land vehicles. Currently, these are in various stages of development. While a trustable fully autonomous systems are some distance away, there is a need to study the trade-offs between the partitioning of tasks, response times, HMI and communication. This would help in establishment of semi-autonomous UAV/UGV that could be deployed for defence applications.

Navigation of Nano UAV with monocular camera (DRDO DYSL - AT): Vision based navigation of an UAV is a necessity in an indoor environment, which is GPS denied. Navigation can be seen as a process through which an UAV plans on how to safely and quickly reach the

target location, which mostly relies on the current environment and location. In order to successfully complete the scheduled mission, a UAV must be fully aware of its states, including location, navigation speed, heading direction as well as starting point and target location. Vision based sensors are passive sensors having anti-interference ability and in a GPS-denied environment, development of vision based navigation would be highly useful for development of efficient and mission specific nano UAVs. With input from a vision sensor, the system has to carry out internal processing of localization and mapping, obstacle avoidance and path planning and output continuous control to drive the UAV to target location through state estimation. With Nano UAVs possessing very limited payload capability, if all these task can be achieved through a single camera without resorting to stereo camera set up , it would be very useful.

SLAM and 3D mapping for navigation of UAV SWARM (DYSL-AT): Vision based SLAM algorithms provide simultaneous recovery of poses of cameras and structure of the environment. 3D mapping of the environment through data collected from various sensors provides useful information for navigation of swarms of UAVs. Monocular visual SLAM for single camera localization with real-time performance via mapping a sparse set of natural features has been considerably developed. Modern feature based SLAM systems have been divided into two parallel independent threads: tracking and mapping. These steps are necessary for visual loop-closure detections that are used to correct the drifts of trajectories which are common in long-term and large-scale environments. The goal of this development should be development of visual SLAM algorithms along with creation of a 3D map of an indoor environment where the layout is unknown and GPS is unavailable. After the maps are developed, these can be utilized for an obstacle-free path planning for the swarm of UAVs to achieve the mission objectives.
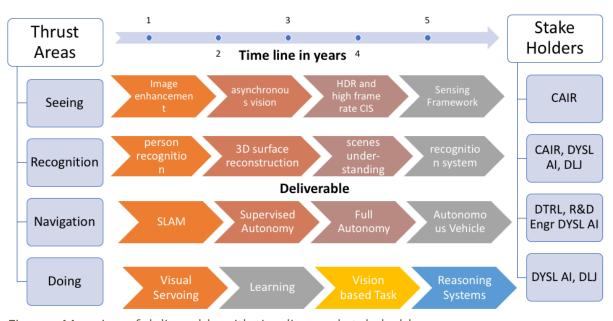

Figure: Mapping of deliverable with timeline and stakeholders

SWARM of UAVs: Distributed multi-sensor data fusion for vision based sensors (DYSL - AT): In a swarm of UAV scenarios, multiple vision based sensors can be mounted on different UAVs of the swarm. This will provide value addition in terms of sensing through various vision

based sensors such as CCD camera, IR/ thermal camera and LiDAR etc, while processing these data on a ground station for creation of larger overview of the surrounding as compared to what could be captured through a single sensor. One of the futuristic technologies in the domain of vision based SWARM is to enhance the target detection capabilities using disimilarities of the object image in the FOV among various distributed vision sensors. Vision based sensors capability enhancement will be enabler for SWARMs to reduce the size of UAV from nano to pico and increase in nos available from tens to lakhs.

Autonomous System for Terrain Navigation – Route Planning (DRDO - DTRL): In this project, it is proposed to use cognition driven intelligence for designing autonomous systems for terrain navigation.

7. Application Areas

This vertical mainly focuses on development of Vision-based Autonomous Systems with applications to defence and security. The following applications area in defence and security have surfaced based on join discussions with DRDO Labs:

- Reconnaissance surveillance and Target acquisition: Reconnaissance operations generally cover large areas with broader intelligence requirements and therefore require mobility and include multiple points of observance. Target acquisition is another requirement of such missions. This calls for a network of ground and air based autonomous systems with vision capabilities for successful completion of mission.

- Border patrol: India shares more than 15000 km long borders with neighbouring countries, and also has border disputes with a few countries. This necessitates the requirement of continuous patrol of the border day and night. Further border areas commonly have diverse geography and non favourable weather conditions for humans. Therefore autonomous systems with vision capability can be a perfect substitute for border patrol requiring large involvement of armed forces.

- Battle damage prediction: Battle damage prediction is an important part of combat damage repair which requires prediction of manpower and material for quick evacuation and repair. Combat in future war would be more intensive and fluid and will produce a large number of reparable combat vehicles. Rapid combat repair can increase combat effectiveness significantly, and hence, timely completion this task is crucial. Set of vision-based autonomous vehicles can play a crucial role in effective battle damage prediction.

- Electronic warfare and Deception operation: Electronic warfare would be a critical element of future's battles and autonomous systems will play a pivotal role in it. For example an aerial autonomous system can emit electromagnetic energy to overwhelm, confuse, deceive or otherwise render ineffective the radar systems of an enmily unit.

- Meteorology missions: Meteorology missions are critical requirements for predicting adverse weather conditions while movements of troops in challenging terrain. Aerial autonomous systems with sensors for meteorological measurement can play a crucial role in quick prediction of adverse conditions enabling preventive measures.

- Radio and data relay: Secured Radio and data relay is a critical requirement of armed forces during combat. In many instances, line of sight is not guaranteed due to limitation of ground tactical networks. This limitation can be overcomed using airborne relay communications using aerial autonomous systems as a possible solution.

- Maritime operation: Maritime operations are the actions of naval forces to combat sea–based terrorism and other illegal activities, such as hijacking, piracy, and human trafficking. Maritime operations also involve the marine environmental protection, creating a safer and clean environment and assisting seafaring vessels in distress. Vision based autonomous systems play an important role in completing such operations in a short time with minimum risk.

- Landmine detection and diffusal: Landmine detection and diffusal is a critical requirement during combat and can cause significant loss of human life in not dictated priori. Here, vision based autonomous or teleoperated systems with manipulation capability and mine detection sensor can be risk free alternatives for landmine diffusal.

- Disaster and crisis management: Defence forces often participate in search and rescue operations during disaster and crisis management. A set of vision based autonomous ground and air based systems can help in by damage prediction, structural inspection, identifying appropriate path for evacuation, and identifying hazardous conditions, etc.

In addition the the above the technology developed through this vertical will also lend its usage in civil application area such as Autonomous Driving, Precision Framing, Green cover monitoring, Yield prediction, Delivery of medicine and organs, Monitoring of animal movements, Traffic spotting, Environmental monitoring, and Fisheries protection.
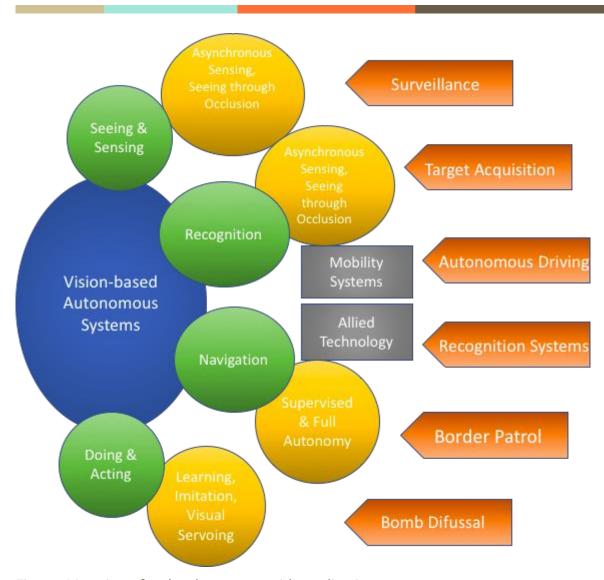
Figure: Mapping of technology areas with applications

8. Work plan

Year 1				Year 2				Year 3				Year 4				Year 5			
Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
Staff and PhD Students Hiring																			
	Equipment Purchase and Setup of lab infrastructure																		
	Research in core areas of Vision based autonomous systems																		
		Development of technologies for Autonomous Systems																	
			Specification and Development of technologies for targeted applications																
						Integration and Testing of all application framework													
	Education and training for developing manpower in Vision-based Autonomous Systems																		

9. Outcomes and Deliverables

9.1 Research and Technological

Outcome and deliverable in four areas are listed below:

Seeing and Sensing

1. An adaptive real-time method for image enhancement in scattering mediums such as water, haze, fog, smoke etc. with respect to variation in density which is robust against non-uniform atmospheric scattering, varying illumination conditions including saturations and high-speed operations.

2. Design of high dynamic range and high frame rate CIS with inbuilt FPN cancellation for defense and autonomous systems.

3. Vision based autonomous systems with asynchronous vision sensors for defence application with the capabilities of automatic 3D reconstruction of the territories, obstacle avoidance, object detection.

4. Deep learning based approach for handling the data acquired through the event-based cameras and solving the problems such as 3D reconstruction, obstacle detection, object detection and classification, visual odometry, pose estimation, and tracking.

Recognition and Understanding

1. Person recognition system in constrained and unconstrained environments

2. System for describing unseen objects and scenes

3. Fast and accurate techniques for real-time incremental 3D surface reconstruction of the defence application zones from images, range scans, and asynchronous vision sensors data.

4. Vision-based sensing systems for reconstructing, recognition, and tracking objects present at a very large distance from the autonomous system.

Navigation

1. Development of guidance and navigation algorithms

2. Robust framework for simultaneous localization and mapping

3. Telepresence and Tele-action (TPTA) system consisting of various sensing modalities like vision, audio, haptics and smell, supported by AR technologies

Doing

1. Framework for vision-assisted collaborative interaction

2. Target acquisition or object manipulation through visual servoing and dexterous manipulation.

3. Distributed control strategies for heterogeneous multiple vehicles cooperation

4. Planning and decision making framework for vision-based autonomy

9.2 Infusion in Curriculum

In the current scenario where the industry is introduced with different aspects of Autonomous Systems, it is important to provide an appropriate academic discourse at different levels for the audience. This TIH will provide avenues to discuss and create appropriate courses in Autonomous Systems for the curriculum at different levels from undergraduate to pre-doctoral level. This may help in capacity building in the area of Autonomous Systems for the future.

9.3 Entrepreneurship and start-ups

The TIH will encourage students and scholars to become entrepreneurs by creating start-ups. It is expected that activities in the various projects undertaken in TIH will provide avenues for initiating many start-ups. The parent institute of a student working in the project may provide extra support and assistance from their technology incubation cells. Some of the start-ups may be further formed as Section-8 companies. This will also promote the "make in India" motto of the nation.

9.4 Industrial Collaboration

One of the main emphasis of this hub will be to create a strong link between academia and industry. This is important to make the research more relevant. As there is an enormous interest in industry in developing technologies for vision-based Autonomous Systems, the TIH will offer the right platform for collaborative activities both in the core technology and applications of Autonomous Systems. The TIH will also provide technology consultancy and transfer to the relevant industries.

10. Grand Problems

The TIH will steer a series of multidisciplinary research initiatives to harness transformational power of computer vision, augmented and virtual reality to address the following grand problems

- *Building dependable autonomous systems for surveillance*
- *Building dependable autonomous systems for healthcare*

These two grand projects aim to produce cost-efficient, indigenous, trustworthy solutions to national needs in the area of defense and security and healthcare. The hub will take advantage of advancements in technological developments of small size but efficient vision sensors, servo controls, responsive actuators, communication devices, artificial intelligence, etc. The core objective of this mega project is to achieve semi to full autonomy for mobility systems, which will allow autonomous systems to navigate and interact in unknown and unstructured environments. The current autonomous systems lack human intelligence, robustness and flexibility and may not be suitable for doing many tasks that humans can do. Many countries have initiated activities in this area and countries which can solve these issues will have an edge in future's warfare. For countries like India, research and development of such autonomous systems is crucial due to strategic location in South Asia.

In the Indian context, technologies related to vision guided autonomous systems can also be deployed in handling natural disasters for rescue and reconstruction operations, operations in nuclear power plants, mining, under-sea explorations etc. F

This mega project will further extend to incorporate virtual reality and augmented reality based solutions in Military Simulation Applications such as seamless 3D visualization with VR effects of battlefield scenarios. The TIH is also looking forward to the development and integration of VR and 3D capabilities in context of over the Sea and underwater Scenario Simulation. The core AR/VR technology stack suitably augmented by research in the related disciplines of computer graphics, computer vision, haptics and machine learning will transform and find many other application domains such as Digital Heritage, Remote Collaboration, Education and Training, Sports and Entertainment, Rehabilitation & Therapeutics, Digital Twin Applications, AR/VR for Telemedicine. The TIH will follow through the framework of trustability, fairness, ethics and privacy in designing and developing the solutions under this mega project to make them robust, responsible and dependable.

11. Lab Infrastructure

Autonomous systems augmented with computer vision, and artificial intelligence have the potential to change the world of mobility systems. With the advancement in technology and wider applications, there has been significant interest in developing vision-based autonomous mobility systems. The development of such systems is made possible with advancements in technological developments of small size but efficient vision sensors, servo controls, responsive actuators, communication devices, artificial intelligence, etc. Vision-based Autonomous mobility systems laboratory aims at knowledge generation towards vision-based control of Autonomous mobility systems. National level lab infrastructure will be set up at the Hub that has access to cutting edge hardware.The lab will house aerial and ground-based all-terrain vehicles with manipulation and other capabilities. The lab will also develop the capability of vision assisted teleoperation and communication system for robust visual feed transmission. The lab will also focus on developing technologies in the area of navigation, manipulation, control, communication and computing and will have related equipment and tools to support activities in these areas. The development of technologies for Vision-based Autonomous Systems will be the primary focus of the proposed laboratory.

12. Training and Outreach

There has been significant interest in developing Vision-based Autonomous systems in industry and research organization. This has led to an infusion of Autonomous systems in education and training space as well. We will develop short and long course modules on different topics related to Vision based Autonomous systems. Courses on Autonomous Systems would be designed and incorporated in all the undergraduate, postgraduate and research scholarly curriculum inorder to engage in the future of Autonomous Systems to solve the community problems from these labs.

The idea is to create opportunities for integrating science and technology, enhancing the ability to capture young people who are educated and motivated to learn and grow. The agenda would be to facilitate training to faculty, staff, undergraduate, postgraduate students and research scholars. Tap on the gaps in learning/industrial training and partner with industries to provide knowhow in the area of Autonomous Systems Technologies and focus on increasing

the accessibility, affordability and availability. On the outreach front, the aim is to operate more programs, and initiatives meeting business and community needs across the nation.

Vision based Autonomous systems laboratory developed as a part of this initiative would provide a space to students undertaking their projects and to use the lab for designing and testing their various algorithms for Vision based Autonomous Systems. The Autonomous systems Lab can also be a resource for training operators, service personnel and production workers, who do not have access to expensive machines as stated earlier. Training young engineers and students in Autonomous Systems and their applications shall open new employment opportunities for them.

13. Conclusion

Vision-based Autonomous Systems for navigation and manipulation in unknown and unstructured environments would be the Main cyber-physical system that will be developed under this vertical of TIH.

14. References

[Daugman, 2004] J. Daugman, (2004), How iris recognition works, IEEE Transactions on Circuits and Systems for Video Technology, vol. 14, no. 1, pp. 21–30.

[Frahm et al. 2010] Frahm, J.M., Fite-Georgel, P., Gallup, D., Johnson, T., Raguram, R., Wu, C., Jen, Y.H., Dunn, E., Clipp, B., Lazebnik, S. and Pollefeys, M., (2010) September. Building Rome on a cloudless day. In European Conference on Computer Vision (pp. 368-381). Springer, Berlin, Heidelberg.

[Hinami et al. 2017] R. Hinami, T. Mei, S. Satoh, (2017), "Joint Detection and recounting of abnormal events by learning deep generic knowledge", IEEE Proc. on ICCV.

[Jayaraman and Grauman 2018] Dinesh Jayaraman, Kristen Grauman (2018): Learning to Look Around: Intelligently Exploring Unseen Environments for Unknown Tasks. CVPR.

[Jianwei Yang et al. 2019] Jianwei Yang, Zhile Ren, Mingze Xu, Xinlei Chen, David J. Crandall, Devi Parikh, Dhruv Batra (2019): Embodied Visual Recognition. CoRR abs/1904.04404

[Lee et al. 2017] W. Lee, S. Kim, Y. Lee, H. Lee, and M. Choi, (2017), "Deep neural networks for wildfire detection with an unmanned aerial vehicle," in IEEE International Conference on Consumer Electronics (ICCE).

[Liang et al. 2019] J. Liang, L. Jiang, J. C. Niebles, A. G. Hauptmann, L. Fei-Fei, (2019), "Peeking into the future: Predicting future person activities and locations in videos", IEEE Proc. on CVPR.

[Maltoni et al., 2009] D. Maltoni, D. Maio, A. Jain, and S. Prabhakar, (2009), Handbook of fingerprint recognition, Springer Science & Business Media.

[Milani et al. 2018] G. Milani, M. Volpi, D. Tonolla, M. Doering, C. Robinson, M. Kneub¨uhler, and M. Schaepman, (2018), "Robust quantification of riverine land cover dynamics by high-resolution remote sensing," Remote Sensing of Environment, vol. 217, pp. 491–505.

[Mittal et al, 2017] P. Mittal, A. Jain, G. Goswami, M. Vatsa, and R. Singh, (2017) Composite Sketch Recognition using Saliency and Attribute Feedback, Information Fusion, Volume 33, Pages 86–99

[Mou et al. 2020] Mou, Lichao, Yuansheng Hua, Pu Jin and Xiao xiang Zhu, (2020), "ERA: A Dataset and Deep Learning Benchmark for Event Recognition in Aerial Videos." ArXiv abs/2001.11394

[Schonberger and Frahm, 2016] Schonberger, J.L. and Frahm, J.M., 2016. Structure-from-motion revisited. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 4104-4113).

[Simonyan and Zisserman 2015] Karen Simonyan, Andrew Zisserman (2015): Very Deep Convolutional Networks for Large-Scale Image Recognition, ICLR.

[Singh et al., 2019] M. Singh, R. Singh, M. Vatsa, N. Ratha, and R. Chellappa, 2019, Recognizing Disguised Faces in the Wild, IEEE Transactions on Biometrics, Behavior, and Identity Science, Volume 1, No. 2, Pages 97-108.

[Tripathi et al. 2017] Tripathi, R. K., Jalal, A. S., Agrawal, S. C., (2017), "Suspicious human activity recognition: a review", Artif Intell Rev, Springer.

[Vetrivel et al. 2018] A. Vetrivel, M. Gerke, N. Kerle, F. Nex, and G. Vosselman, (2018), "Disaster damage detection through synergistic use of deep learning and 3D point cloud features derived from very high resolution oblique aerial images, and multiple-kernel-learning," ISPRS Journal of Photogrammetry and Remote Sensing, vol. 140, pp. 45–59.

[Wu, 2013] Wu, C., 2013, June. Towards linear-time incremental structure from motion. In 2013 International Conference on 3D Vision-3DV 2013 (pp. 127-134). IEEE.

[Xu et al. 2018] Z. Xu, L. Wu, and Z. Zhang, (2018), "Use of active learning for earthquake damage mapping from UAV photogrammetric point clouds," International Journal of Remote Sensing, vol. 39, no. 15-16, pp. 5568–5595.

[Zhao et al., 2003] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, (2003) Face recognition: A literature survey, ACM computing surveys, vol. 35, no. 4, pp. 399–458.

[Zheng et al. 2010] Zheng, Q., Sharf, A., Wan, G., Li, Y., Mitra, N.J., Cohen-Or, D. and Chen, B., 2010. Non-local scan consolidation for 3D urban scenes. ACM Trans. Graph., 29(4), pp.94-1.

[Zhu et al, 2014] J. Y. Zhu, W. S. Zheng, J. H. Lai, and S. Z. Li, (2014), Matching NIR face to VIS face using transduction, IEEE Transactions on Information Forensics and Security, 9(3):501–514.

A. Cardenas, S. Amin, and S. Sastry, "Secure control: Towards survivable cyber-physical systems," in 2008 The 28th International Conference on Distributed Computing Systems Workshops. IEEE, 2008, pp. 495-500.

Cetinkaya, H. Ishii, and T. Hayakawa, "Networked control under random and malicious packet losses," IEEE Transactions on Automatic Control, vol. 62, no. 5, pp. 2434-2449, 2016.

Jain and D. Ghose, "Stabilization of collective formations with speed and controller gain heterogeneity and saturation," Journal of the Franklin Institute, vol. 354, no. 14, pp. 5964-5995, 2017.

Jain and D. Ghose, "Trajectory-constrained collective circular motion with different phase arrangements," IEEE Transactions on Automatic Control, 2019.

Rahnama and P. J. Antsaklis, "Learning-based event-triggered control for synchronization of passive multi-agent systems under attack," IEEE Transactions on Automatic Control, 2019.

Teixeira, K. C. Sou, H. Sandberg, and K. H. Johansson, "Secure control systems: A quantitative risk management approach," IEEE Control Systems Magazine, vol. 35, no. 1, pp. 24-45, 2015.

Adams R.J., & Hannaford, B. (1999). Stable haptic interaction with virtual environments. In IEEE Transactions on Robotics and Automation, vol. 15, no. 3, pp. 465-474, June 1999.

Akella, Prasad, et al. "Cobots for the automobile assembly line." Proceedings 1999 IEEE International Conference on Robotics and Automation (Cat. No. 99CH36288C). Vol. 1. IEEE, 1999.

Almeida, C., Franco, T., Ferreira, H., Martins, A., Santos, R., Almeida, J. M., ... & Silva, E. (2009, May). Radar based collision detection developments on USV ROAZ II. In Oceans 2009-Europe (pp. 1-6). Ieee.

Alves, J. C., & Cruz, N. A. (2015, August). Ais-enabled collision avoidance strategies for autonomous sailboats. In World Robotic Sailing championship and International Robotic Sailing Conference (pp. 77-87). Springer, Cham.

Arai, T., Yasue, T., Kitamura, K., Shimamoto, H., Kosugi, T., Jun, S.W., Aoyama, S., Hsu, M.C., Yamashita, Y., Sumi, H. and Kawahito, S., 2017. A 1.1 mm, 33-Mpixel 240-fps 3-D-Stacked CMOS Image Sensor with three-stage Cyclic-Cyclic-SAR Analog-to-Digital Converters. IEEE Transactions on Electron Devices, 64(12), pp.4992-5000.

Arkin, R., & Ali, K. (1994, August). Integration of reactive and telerobotic control in multi-agent robotic systems. In Proc. Simulation of Adaptive Behavior.

Babaeizadeh, C. Finn, D. Erhan, R. Campbell and S. Levine, "Stochastic Variational Video Prediction," in Sixth International Conference on Learning Representations, 2018.

Barmish, B. R., Corless, M., & Leitmann, G. (1983). A new class of stabilizing controllers for uncertain dynamical systems. SIAM Journal on Control and Optimization, 21(2), 246-255.

Bermejo, C., & Hui, P. (2017). A survey on haptic technologies for mobile augmented reality. arXiv:1709.00698 [cs]. http://arxiv.org/abs/1709. 00698 arXiv: 1709.00698.

Bhattacharjee et al., "Directional Attention based Video Frame Prediction using Graph Convolutional Networks" in Proceedings of the International Joint Conference on Neural Networks, 2019.

Bibuli, M., Bruzzone, G., Caccia, M., & Lapierre, L. (2009). Path-following algorithms and experiments for an unmanned surface vehicle. Journal of Field Robotics, 26(8), 669-688.

Bibuli, M., Caccia, M., Lapierre, L., & Bruzzone, G. (2012). Guidance of unmanned surface vehicles: Experiments in vehicle following. IEEE Robotics & Automation Magazine, 19(3), 92-102.

Borrelli, F., Bemporad, A., & Morari, M. (2017). Predictive control for linear and hybrid systems. Cambridge University Press.

Brandli, C., Berner, R., Yang, M., Liu, S.C. and Delbruck, T., 2014. A 240× 180 130 db 3 μs latency global shutter spatiotemporal vision sensor. IEEE Journal of Solid-State Circuits, 49(10), pp.2333-2341.

Briñón-Arranz, L., Schenato, L., & Seuret, A. (2015). Distributed source seeking via a circular formation of agents under communication constraints. IEEE Transactions on Control of Network Systems, 3(2), 104-115.

Caccia, M., Bibuli, M., Bono, R., & Bruzzone, G. (2008). Basic navigation, guidance and control of an unmanned surface vehicle. Autonomous Robots, 25(4), 349-365.

Cadena, C., Carlone, L., Carrillo, H., Latif, Y., Scaramuzza, D., Neira, J., Reid, I. and Leonard, J.J., 2016. Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age. IEEE Transactions on robotics, 32(6), pp.1309-1332.

Carter, T., Seah, S.A., Long, B., Drinkwater, B., & Subramanian, S. (2013). UltraHaptics: multi-point mid-air haptic feedback for touch surfaces. In Proceedings of the 26th annual ACM symposium on User interface software and technology (UIST '13). ACM, New York, NY, USA, 505-514.

Caruso, D., Engel, J. and Cremers, D., 2015, September. Large-scale direct SLAM for omnidirectional cameras. In 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (pp. 141-148). IEEE.

Chaplot, D.S., Salakhutdinov, R., Gupta, A. and Gupta, S., 2020. Neural Topological SLAM for Visual Navigation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 12875-12884).

Chen, Jian, et al. "Multiple UAV navigation with finite sensing zone." 2006 American Control Conference. IEEE, 2006.

Chen, Y. H. (1988). Design of robust controllers for uncertain dynamical systems. IEEE Transactions on Automatic control, 33(5), 487-491.

Concannon, D., Flynn, R., & Murray N. (2019). A quality of experience evaluation system and research challenges for networked virtual reality-based teleoperation applications. In Proceedings of the 11th ACM Workshop on Immersive Mixed and Virtual Environment Systems (MMVE '19). ACM, New York, NY, USA, 10-12.

Corke, P. I., & Hutchinson, S. A. (2001). A new partitioned approach to image-based visual servo control. IEEE Transactions on Robotics and Automation, 17(4), 507-515.

Cortes, Jorge, et al. "Coverage control for mobile sensing networks." IEEE Transactions on robotics and Automation 20.2 (2004): 243-255.

Denton and E. L., "Unsupervised learning of disentangled representations from video," in Advances in Neural Information Processing Systems, 2017.

E. Malis, and S. Boudet, "2 1/2 D visual servoing," in IEEE Transactions on Robotics and Automation, vol. 15, no. 2, pp. 238-250, April 1999, doi: 10.1109/70.760345.

Engel, J., Koltun, V. and Cremers, D., 2017. Direct sparse odometry. IEEE transactions on pattern analysis and machine intelligence, 40(3), pp.611-625.

F. L. Lewis, H. Zhang, K. Hengster-Movric, and A. Das, Cooperative control of multi-agent systems: optimal and adaptive design approaches. Springer Science & Business Media, 2013.

F. Pasqualetti, F. Dorer, and F. Bullo, "Control-theoretic methods for cyber physical security: Geometric principles for optimal cross-layer resilient control systems," IEEE Control Systems Magazine, vol. 35, no. 1, pp. 110-127, 2015.

Fan, Xiying, Chuanhe Huang, Bin Fu, Shaojie Wen, and Xi Chen. "UAV-Assisted Data Dissemination in Delay-Constrained VANETs." Mobile Information Systems 2018 (2018).

Fawaz, Wissam, RibalAtallah, ChadiAssi, and Maurice Khabbaz. "Unmanned aerial vehicles as store-carry-forward nodes for vehicular networks." IEEE Access 5 (2017): 23710-23718.

Fragkiadaki, Katerina, and e. al., "Motion prediction under multimodality with conditional stochastic networks," arXiv:1705.02082, 2017.

Freeman, E., Anderson, R., Williamson, J., Wilson, G., & Brewster, S.A. (2017). Textured surfaces for ultrasound haptic displays. In Proceedings of the 19th ACM International Conference on Multimodal Interaction (ICMI '17). ACM, New York, NY, USA, 491-492.

Frey, C., Zarzhitsky, D., Spears, W. M., Spears, D. F., Karlsson, C., Ramos, B., ... & Widder, E. A. (2008, September). A physicomimetics control framework for swarms of autonomous surface vehicles. In OCEANS 2008 (pp. 1-6). IEEE.

Garcia, G. J., Pomares, J., Torres, F., & Gil, P. (2014). Event-based visual servoing with features' prediction. In ROBOT2013: First Iberian Robotics Conference (pp. 679-691). Springer, Cham.

Girard, Anouck R., Adam S. Howell, and J. Karl Hedrick. "Border patrol and surveillance missions using multiple unmanned air vehicles." 2004 43rd IEEE Conference on Decision and Control (CDC)(IEEE Cat. No. 04CH37601). Vol. 1. IEEE, 2004.

Goodfellow et al., "Generative Adversarial Nets", in Advances in Neural Information Processing Systems, 2014.

H. Cai and C. Bai, "Deep Video Generation, Prediction and Completion of Human Action Sequences," in Proceedings of the European Conference on Computer Vision, 2018.

H. Yang, Q.-L. Han, X. Ge, L. Ding, Y. Xu, B. Jiang, and D. Zhou, "Fault tolerant cooperative control of multi-agent systems: A survey of trends and methodologies," IEEE Transactions on Industrial Informatics, 2019.

Healey, A. J., Horner, D. P., Kragelund, S. P., Wring, B., & Monarrez, A. (2007). Collaborative unmanned systems for maritime and port security operations. NPS Montrey Bay CA, Center for autonomous underwater vehicle research.

Heidarsson, H. K., & Sukhatme, G. S. (2011). Obstacle detection and avoidance for an autonomous surface vehicle using a profiling sonar. In 2011 IEEE International Conference on Robotics and Automation (pp. 731-736). IEEE.

Isleyen, A., Vardar, Y. & Basdogan, C. (2019). Tactile Roughness Perception of Virtual Gratings by Electrovibration. In IEEE Transactions on Haptics.

J. Xu, B. Ni and X. Yang, "Video Prediction via Selective Sampling," in Advances in Neural Information Processing Systems, 2018.

K. Wu and D. Dunning, "Hypocognition: Making Sense of the Landscape Beyond One's Conceptual Reach," Review of General Psychology Special Issue: Meaning of Meanings, vol. 1, no. 2, pp. 1-11, 2017, http://dx.doi.org/10.1037/gpr0000126.

K.-K. Oh, M.-C. Park, and H.-S. Ahn, "A survey of multi-agent formation control," Automatica, vol. 53, pp. 424-440, 2015.

Kaur, A., Mishra, D. and Sarkar, M., 2018. A 12-bit, 2.5-bit/Phase Column-Parallel Cyclic ADC. IEEE Transactions on Very Large Scale Integration (VLSI) Systems, 27(1), pp.248-252.

Kim, H., Leutenegger, S. and Davison, A.J., 2016, October. Real-time 3D reconstruction and 6-DoF tracking with an event camera. In European Conference on Computer Vision (pp. 349-364). Springer, Cham.

Kitts, C., Mahacek, P., Adamek, T., & Mas, I. (2011, September). Experiments in the control and application of Automated Surface Vessel fleets. In OCEANS'11 MTS/IEEE KONA (pp. 1-7).

Kiumarsi, B., & Başar, T. (2018, December). Human-in-the-loop Control of Distributed Multi-Agent Systems: A Relative Input-Output Approach. In 2018 IEEE Conference on Decision and Control (CDC) (pp. 3343-3348). IEEE.

Kumar, H., Gupta, S., & Venkatesh, K. S., A Novel Method for Inferior Mirage Detection in Video, Digital Image & Signal Processing (DISP2019), Vol. , pp, 2019, IEEE.

Kumar, H., Yadav, A.S., Gupta, S. and Venkatesh, K.S., 2018. Depth map estimation using defocus and motion cues. IEEE Transactions on Circuits and Systems for Video Technology, 29(5), pp.1365-1379.

Kwon YH and Park MG, "Predicting Future Frames using Retrospective Cycle GAN, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019.

Kwon, J.H., Ryu, S., Ko, G., Lee, J., Suh, K., & Nam, S. (2019). Design of a VR Prototyping System Based on a Haptic Controller. In IEEE 6th International Conference on Industrial Engineering and Applications (ICIEA), Tokyo, Japan, 2019, pp. 41-44.

Lancelot, F., Causse, M., Schneider, N., & Mongeau, M. (2015). Human-in-the-loop multi-agent approach for airport taxiing operations. In Trends in Practical Applications of Agents, Multi-Agent Systems and Sustainability (pp. 235-236). Springer, Cham.

Lewis, F. L., & Vrabie, D. (2009). Reinforcement learning and adaptive dynamic programming for feedback control. IEEE circuits and systems magazine, 9(3), 32-50.

Lichtsteiner, P., Posch, C. and Delbruck, T., 2008. A 128$\times$128 120 dB 15μ s Latency Asynchronous Temporal Contrast Vision Sensor. IEEE journal of solid-state circuits, 43(2), pp.566-576.

Lin, F., & Olbrot, A. W. (1996, December). An LQR approach to robust control of linear systems with uncertain parameters. In Proceedings of 35th IEEE Conference on Decision and Control (Vol. 4, pp. 4158-4163). IEEE.

Lin, M. (Ed.), Otaduy, M. (Ed.). (2008). Haptic Rendering. New York: A K Peters/CRC Press,

Litts K.B., & Lewis E.W. (2019). Mobile Augmented Reality: Exploring a new genre of learning. GetMobile: Mobile Comp. and Comm. 22, 3, 5-9.

Long, B., Seah, S.A., Carter, T., & Subramanian, S. (2014). Rendering volumetric haptic shapes in mid-air using ultrasound. ACM Trans. Graph. 33, 6, Article 181 (November 2014), 10 pages.

Lu, J., & Zhou, H. (2017). Numerical reproduction and explanation of road surface mirages under grazing-angle scattering. Applied optics, 56(19), 5550-5558.

Lu, Yuncheng, et al. "A survey on vision-based UAV navigation." Geo-spatial information science 21.1 (2018): 21-32.

Luc et al., "Predicting Deeper into the Future of Semantic Segmentation" in Proceedings of the IEEE International Conference on Computer Vision, 2017.

M. A. Saez, F. P. Maturana, K. Barton and D. M. Tilbury, "Context-Sensitive Modeling and Analysis of Cyber-Physical Manufacturing Systems for Anomaly Detection and Diagnosis," in IEEE Transactions on Automation Science and Engineering, vol. 17, no. 1, pp. 29-40, Jan. 2020, doi: 10.1109/TASE.2019.2918562.

M. De Queiroz, X. Cai, and M. Feemster, Formation Control of Multi-agent Systems: A Graph Rigidity Approach. John Wiley & Sons, 2019.

M. Mathieu, C. Couprie and Y. LeCun, "Deep multi-scale video prediction beyond mean square error," in arXiv preprint arXiv:1511.05440, 2015.

M. Mesbahi and M. Egerstedt, Graph theoretic methods in multiagent networks. Princeton University Press, 2010, vol. 33.

M. S. Rahman, M. A. Mahmud, A. M. T. Oo, and H. R. Pota, "Multi-agent approach for enhancing security of protection schemes in cyber-physical energy systems," IEEE transactions on industrial informatics, vol. 13, no. 2, pp. 436-447, 2016.

Matheson, Eloise, et al. "Human–Robot Collaboration in Manufacturing Applications: A Review." Robotics 8.4 (2019): 100.

Misra, Sohum, et al. "Cooperative Search Area optimization using Multiple Unmanned Aerial Vehicles in a GPS-denied Environment." 2019 International Conference on Unmanned Aircraft Systems (ICUAS). IEEE, 2019.

Morris, A., Donamukkala, R., Kapuria, A., Steinfeld, A., Matthews, J. T., Dunbar-Jacob, J., & Thrun, S. (2003, September). A robotic walker that provides guidance. In 2003 IEEE International Conference on Robotics and Automation (Cat. No. 03CH37422) (Vol. 1, pp. 25-30). IEEE.

Mueggler, Elias, Guillermo Gallego, Henri Rebecq, and Davide Scaramuzza. "Continuous-time visual-inertial odometry for event cameras." IEEE Transactions on Robotics 34, no. 6 (2018): 1425-1440.

N. Crombez, E. M. Mouaddib, G. Caron and F. Chaumette, "Visual Servoing With Photometric Gaussian Mixtures as Dense Features," in IEEE Transactions on Robotics, vol. 35, no. 1, pp. 49-63, Feb. 2019, doi: 10.1109/TRO.2018.2876765.

N. E. Leonard, D. A. Paley, F. Lekien, R. Sepulchre, D. M. Fratantoni, and R. E. Davis, "Collective motion, sensor networks, and ocean sampling," Proceedings of the IEEE, vol. 95, no. 1, pp. 48-74, 2007.

Nayar, S. K., & Narasimhan, S. G. (1999, September). Vision in bad weather. In Proceedings of the Seventh IEEE International Conference on Computer Vision (Vol. 2, pp. 820-827). IEEE.

Neves, R., & Matos, A. C. (2013, September). Raspberry PI based stereo vision for small size ASVs. In 2013 OCEANS-San Diego (pp. 1-6). IEEE.

Nomura, R., Unuma, Y., Komuro, T., Yamamoto, S., & Tsumura, N. (2017). Mobile augmented reality for providing perception of materials. In Proceedings of the 16th International Conference on Mobile and Ubiquitous Multimedia (MUM '17). ACM, New York, NY, USA, 501-506.

Orsag, M., Haus, T., Tolić, D., Ivanovic, A., Car, M., Palunko, I., & Bogdan, S. (2016, June). Human-in-the-loop control of multi-agent aerial systems. In 2016 European Control Conference (ECC) (pp. 2139-2145). IEEE.

Osgouei, R.H., Kim J.R., & Choi, S. (2019). Data-driven Texture Modeling and Rendering on Electrovibration Display. In IEEE Transactions on Haptics.

P. Bhattacharjee and S. Das, "Temporal Coherency based Criteria for Predicting Video Frames using Deep Multi-stage Generative," in Advances in Neural Information Processing Systems, 2017.

P. Bhattacharjee and S. Das, "Temporal Coherency based Criteria for Predicting Video Frames using Deep Multi-stage Generative," in Advances in Neural Information Processing Systems, 2017.

P. G. Trepagnier, J. Nagel, P. M. Kinney, C. Koutsourgeras, and M. Dooner, "KAT-5: Robust systems for autonomous vehicle navigation in challenging and unknown terrain," DARPA 2005, STAR 36, M. Buehler, K. Iagnemma, S. Singh (Eds.), Springer, pp. 103-128, 2007.

Plonski, P. A., Vander Hook, J., Peng, C., Noori, N., & Isler, V. (2015). Navigation around an unknown obstacle for autonomous surface vehicles using a forward-facing sonar. In IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR).

Ploplys, N. J., Kawka, P. A., & Alleyne, A. G. (2004). Closed-loop control over wireless networks. IEEE Control Systems Magazine, 24(3), 58-71.

Qu, Yaohong, et al. "A fault-tolerant cooperative positioning approach for multiple UAVs." IEEE Access 5 (2017): 15630-15640.

R. T. Fomena and F. Chaumette, "Visual Servoing from Spheres using a Spherical Projection Model," Proceedings 2007 IEEE International Conference on Robotics and Automation, Roma, 2007, pp. 2080-2085, doi: 10.1109/ROBOT.2007.363628.

Rebecq, H., Horstschäfer, T., Gallego, G. and Scaramuzza, D., 2016. EVO: A geometric approach to event-based 6-DOF parallel tracking and mapping in real time. IEEE Robotics and Automation Letters, 2(2), pp.593-600.

Redo-Sanchez, A., Heshmat, B., Aghasi, A., Naqvi, S., Zhang, M., Romberg, J., & Raskar, R. (2016). Terahertz time-gated spectral imaging for content extraction through layered structures. Nature communications, 7(1), 1-7.

Reed, C. M. (2008). Experiments in real-time path planning for riverine environments (Doctoral dissertation, Virginia Tech).

Saif, Osamah, Isabelle Fantoni, and Arturo Zavala-Río. "Distributed integral control of multiple UAVs: precise flocking and navigation." IET Control Theory & Applications 13.13 (2019): 2008-2017.

Salisbury J.K., & Srinivasan M.A. (1997). Phantom-based haptic interaction with virtual objects. In IEEE Computer Graphics and Applications, vol. 17, no. 5, pp. 6-10, Sept.-Oct. 1997.

Salisbury, J.K., Conti, F., & Barbagli, F. (2004).Haptic rendering: introductory concepts. In IEEE Computer Graphics and Applications, vol. 24, no. 2, pp. 24-32, March-April 2004.

Shen, S., Mulgaonkar, Y., Michael, N. and Kumar, V., 2014, May. Multi-sensor fusion for robust autonomous flight in indoor and outdoor environments with a rotorcraft MAV. In 2014 IEEE International Conference on Robotics and Automation (ICRA) (pp. 4974-4981). IEEE.

Sun, Y., Yoshida, S., Narumi, T., & Hirose, M. (2018). Handheld Haptic Interface for Rendering Size, Shape, and Stiffness of Virtual Objects. In Proceedings of the 2018 ACM International Conference on Interactive Surfaces and Spaces (ISS '18). ACM, New York, NY, USA, 411-414.

Sun, Y., Yoshida, S., Narumi, T., & Hirose, M. (2019). PaCaPa: A Handheld VR Device for Rendering Size, Shape, and Stiffness of Virtual Objects in Tool-based Interactions. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19). ACM, New York, NY, USA, Paper 452, 12 pages.

T. Qin, P. Li and S. Shen, "VINS-Mono: A Robust and Versatile Monocular Visual-Inertial State Estimator," in IEEE Transactions on Robotics, vol. 34, no. 4, pp. 1004-1020, Aug. 2018.

Tripathy, N. S., Kar, I. N., Paul, K., An event-triggered based robust control of robot manipulator. In International Conference on Control Automation Robotics & Vision, 2014.

Tzeng, C. Y., & Chen, J. F. (1999). Fundamental properties of linear ship steering dynamic models. Journal of Marine Science and Technology, 7(2), 79-88.

Vardar, Y., Güçlü, B., & and Basdogan, C. (2017). Effect of Waveform on Tactile Perception by Electrovibration Displayed on Touch Screens. In IEEE Transactions on Haptics, vol. 10, no. 4, pp. 488-499, 1 Oct.-Dec. 2017.

Vardar, Y., Güçlü, B., & and Basdogan, C. (2018). Tactile Masking by Electrovibration. In IEEE Transactions on Haptics, vol. 11, no. 4, pp. 623-635, 1 Oct.-Dec. 2018.

Vondrick, Carl, H. Pirsiavash and A. Torralba, "Generating videos with scene dynamics," in Advances In Neural Information Processing Systems, 2016.

W. He, X. Gao, W. Zhong, and F. Qian, "Secure impulsive synchronization control of multi-agent systems under deception attacks," Information Sciences, vol. 459, pp. 354-368, 2018.

W. Lotter, G. Kreiman, and D. Cox, "Deep Predictive Coding Networks for Video Prediction and Unsupervised Learning," in arXiv preprint arXiv:1605.08104, 2016.

Walker, Jacob, and e. al., "The pose knows: Video forecasting by generating pose futures," in Proceedings of the IEEE International Conference on Computer Vision, 2017.

Wang, J., Huang, P., Chen, C., Gu, W., & Chu, J. (2011, January). Stereovision aided navigation of an autonomous surface vehicle. In 2011 3rd International Conference on Advanced Computer Control (pp. 130-133). IEEE.

Wegner, R., & Anderson, J. (2006). Agent-based support for balancing teleoperation and autonomy in urban search and rescue. International Journal of Robotics and Automation, 21(2), 120-128.

Xiao, T., Liu, Y., Zhou, B., Jiang, Y. and Sun, J., 2018. Unified perceptual parsing for scene understanding. In Proceedings of the European Conference on Computer Vision (ECCV) (pp. 418-434).

Xingjian et al., "Convolutional LSTM network: A machine learning approach for precipitation nowcasting", in Advances in Neural Information Processing Systems, 2015.

Xu, B. Ni and X. Yang, "Video Prediction via Selective Sampling," in Advances in Neural Information Processing Systems, 2018.

Xu, B., Stilwell, D. J., & Kurdila, A. (2009, October). Efficient computation of level sets for path planning. In 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems (pp. 4414-4419). IEEE.

Y. Cao, W. Yu, W. Ren, and G. Chen, "An overview of recent progress in the study of distributed multi-agent coordination," IEEE Transactions on Industrial Informatics, vol. 9, no. 1, pp. 427-438, 2012.

Y. Wan, J. Cao, G. Chen, and W. Huang, "Distributed observer-based cyber-security control of complex dynamical networks," IEEE Transactions on Circuits and Systems I: Regular Papers, vol. 64, no. 11, pp. 2966-2975, 2017.

Yakimenko, O. A., & Kragelund, S. P. (2011). Real-time optimal guidance and obstacle avoidance for umvs. Autonomous Underwater Vehicles, 67-98.

Yang, N., Stumberg, L.V., Wang, R. and Cremers, D., 2020. D3VO: Deep Depth, Deep Pose and Deep Uncertainty for Monocular Visual Odometry. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 1281-1292).

Yen, S.C., Fan, C.L., & Hsu, C.H. (2019). Streaming 360° videos to head-mounted virtual reality using DASH over QUIC transport protocol. In Proceedings of the 24th ACM Workshop on Packet Video (PV '19). ACM, New York, NY, USA, 7-12.

Z. Yang and S. Shen, "Monocular Visual–Inertial State Estimation With Online Initialization and Camera–IMU Extrinsic Calibration," in IEEE Transactions on Automation Science and Engineering, vol. 14, no. 1, pp. 39-51, Jan. 2017, doi: 10.1109/TASE.2016.2550621.

Zhuang, J., Su, Y., & Liao, Y. (2011). Unmanned surface vehicle target tracking based on marine radar. In 2011 International Conference on Computer Science and Service System (CSSS) (pp. 1872-1875). IEEE.

Computer Vision for Better Living: Healthcare & Biosphere

From Imaging to Image Analysis

Collaborating **Organizations:**

IIT Jodhpur

National Collaborators (Healthcare): CSIR-IGIB, IIIT Hyderabad, IIT Kharagpur, IISc Bangalore, AIIMS Jodhpur, PGI Chandigarh, Tata Medical Centre, Kolkata, Medanta Institute of Education and Research, TeleRadiology Solutions, Bangalore, CARING, Bangalore, Max Super Speciality Hospital New Delhi

National Collaborators (Biosphere): IIIT Delhi, ICAR-CMFRI, ICAR-IARI, Department of Forest Himachal Pradesh

Foreign Collaborators (Healthcare): John Hopkins University

Foreign Collaborators (Biosphere): University of Lincoln, UK

Introduction

Healthy humans depend on healthy environments. Biodiversity, ecosystems, and the essential services that they deliver, are central pillars for all life forms on the planet. They are sources of food and essential nutrients, medicines and medicinal compounds, fuel, energy, livelihoods and cultural and spiritual enrichment. They also contribute to the provision of clean water and air, and perform critical functions that range from the regulation of pests and disease to that of climate change and natural disasters. Each of these functions has direct and indirect consequences on our health and well-being. Studies have shown that biodiversity is important for both mental and physical well-being of humans. Despite our reliance on nature, we are losing species and ecosystems, that support healthy environments, at alarming rates, resulting in environmental degradation and adverse effects on human health. The novel coronavirus is a prominent example that sheds light on the close links between the loss of biodiversity and habitats on one hand and healthy functioning human societies on the other.

For decades, researchers have been exploring the use of computer vision technologies for human healthcare. However, the application of computer vision for monitoring and conserving biodiversity, and educating about flora and fauna is relatively recent. But the possibilities are immense. Therefore, the second vertical of the iHub Drishti will focus on different aspects of

good living - human healthcare, plant health, and marine life, and is accordingly divided into two sub-verticals: human health and biosphere. Unlike the autonomous systems vertical which will have an end-to-end cyber-physical system with several technology fallouts, this will involve designing a series of cyber-physical systems for specific tasks, for instance, individual CPS will be build for detecting a specific disease, there will be different systems for animal and plant phenotyping.

Figure: Showcasing the components of biodiversity and human health[8].

SUMMARY

A consortium mode project has been framed based upon careful requirement analysis with stakeholders and end users. Based upon the need, we have focused on delivering dependable Computer Vision based technology for diagnostics, prognostics and therapeutics using multi-modality imaging. We have addressed problems of chronic, infectious diseases as well traumatic injuries. Responding to the future trend for precision medicine, we have designed a major initiative in the area of radio-genomics in collaboration with our research partners. As we are facing the unprecedented crisis of COVID-19, IIT Jodhpur is working on a multi-institutional collaborative project RAKSHAK for developing AI based technologies for diagnostics, disease surveillance and mitigating effect of COVID-19 on different aspects of our life.

A.1 Prologue: Imaging for Human Health

Image based tools are widely used in healthcare applications. Different modalities of imaging solutions are being developed for better non-invasive examination of patients. Computer vision techniques provide means for automating the interpretation and report generation process for image-based investigations. The role of computer vision solutions is not just restricted to detection, diagnosis and treatment in hospitals, rather, the technology can also be used to facilitate public healthcare and universal healthcare schemes.

Today, computer vision technologies are being used for a variety of applications in the healthcare industry. These systems are achieving remarkable performances at the cost of increased complexity for medical applications of image analysis. As these systems are being ubiquitously introduced into critical domains, such as medical image computing and computer-assisted intervention, it becomes imperative to develop methodologies that explain their

---

[8]Images is taken from https://www.cbd.int/gbo/

predictions as well. Such methodologies would help physicians decide whether they should follow/trust a prediction. Detection, diagnosis and treatment are the three primary pillars of healthcare tasks. Depending on the specialization and the disease, the information modalities and source of detection and diagnosis are different. For instance, for tuberculosis diagnosis, X-ray and sputum are the primary sources of confirmation, whereas for anaemia, blood test is used. According to the market study done by Frost and Sullivan (refer to the figure below), the global medical imaging market is predicted to increase significantly in the next 10 years, out of which the computer vision based imaging is predicted to increase to 13% of the market share by 2028. This validates the potential of technology development in the area of computer vision for medical imaging.

Starting from acquisition, to processing, diagnostics, robotic surgeries, and analyzing results based on the specificity and sensitivity for a given population, the intersection of computer vision and healthcare automation covers areas such as sensing, detection, navigation, mixed reality, and trustability. iHub Drishti, given the market opportunity and national need has chosen Computer Vision for Healthcare as a major thrust area.
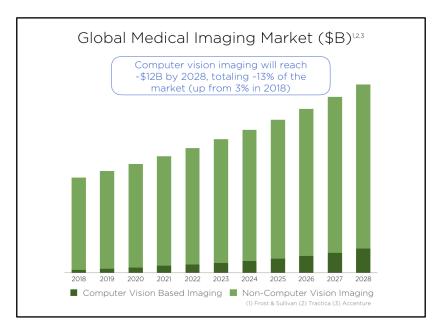
Figure: Prediction of the global imaging market by Frost & Sullivan, Tractica and Accenture.

A.2 National Need & Background

The current Indian healthcare system in India is under immense pressure. The Central Bureau of Health Intelligence (CBHI) data from the National Health Profile 2017 states that there is only one government allopathic doctor for every 10,189 people in India. Moreover, the World Health Organization (WHO) study 'The Health Workforce in India', published in 2016 shows that more than 50% of those calling themselves allopathic doctors do not have a medical qualification. The rural/remote areas are the most affected where self-styled doctors provide up to 75% of primary care. In such scenarios, automated healthcare solutions using imaging, a key investigation tool, would be highly convenient. These could aid in remote monitoring, fast and accurate diagnosis, identification of critical cases, and recommendation of correct treatment.

Primary healthcare centers are required to provide essential and quality health services, from health promotion to prevention, treatment, rehabilitation, and palliative care. However, medical experts are not available in every city, and are far more accessible in metro cities in secondary and tertiary care centers or specialized hospitals. This increases the burden of care on the secondary and tertiary care centers in India. Again, there are countless instances of patients, unaware of their medical conditions, are rendered victims of tragic circumstances, like loss of vision in cataract patients losing their eyesight if not diagnosed at the right time. The use of computer vision technology can not only help identifying patients at the right time, but also help them receive quality healthcare. Development of awareness and encouragement in research in the field of uncertainty modeling in medical image interpretation is required to enable dependable implementation of machine learning tools in the clinical world.

The world during the COVID pandemic and post-COVID has increased acceptance to telemedicine. The combination of Computer Vision with Natural Language Processing has immense possibilities in this area. Applications and chatbots are being developed to answer medical questions and perform examinations just as a health professional would. The use of speech and language processing helps patients provide symptoms of medical concern with ease; and these terms can be used for identifying and collating medical history documents through simple search operations by a machine. In such applications, the introduction of computer vision can raise the level of functionality to far greater heights than a human doctor.

IIT Jodhpur has been identified as the coordinator for the RAKSHAK initiative, a flagship program of the Department of Science and Technology (DST), which focuses on Remedial Action, Knowledge Skimming and Holistic Analysis of COVID-19. Faculty members at IIT Jodhpur are working towards this enterprise and iHub Drishti will be supporting these activities. With this motivation, the TIH will be focusing on the following three broad applications of computer vision and mixed reality in healthcare for better living.

- Diagnosis, Prognosis and Therapeutics
- Public Health Care Services
- RAKSHAK for COVID19

It is of paramount importance that all predictions and decisions made by automated healthcare solutions are explainable to its users. Therefore, the technologies to be designed and built in iHub Drishti will have explainability and dependability as integral components of the solutions. Our faculty members have considerable experience in working on different aspects of healthcare and some of the major areas in which faculty members have made significant contributions are: Neural connectomics, Ultrasound image analysis, MR image analysis, and Histopathology image analysis. These initiatives were completed in collaboration with leading medical institutions and hospitals of the country such as AIIMS Jodhpur, PGI Chandigarh, AIIMS Rishikesh, National Brain Research Centre and Medanta hospital.

A.3 Objectives

TIH at IIT Jodhpur has organized technology development work in this vertical keeping two essential requirements of Indian healthcare systems in mind

(i)   Reducing cost of image-based diagnostics

(ii)   Making improved healthcare accessible to the bottom of the pyramid

At the same time iHub DRISHTI will put in consistent and collective effort towards contributing technologies for advancing the state-of-the art of image analysis for patient care. Objectives of the vertical on computer vision for healthcare will be the following:

1. ***Development of computational systems for capability enhancement of targeted imaging modalities:*** Key motivation will be to achieve better detectability of abnormalities despite low cost imaging modalities.

2. ***Design of advanced image based dependable diagnostic and prognostic tools for dealing with trauma, chronic and infectious diseases:*** We shall address the problem of dependability of the automated decision of the deep learning-based computer vision tools so that the systems developed can provide meaningful measure of confidence in the final decision.

3. ***Developing machine learning techniques for radiomics and radio-genomics to support individualized treatment plan based upon imaging and multi-omic data:*** We intend to develop dependable techniques for detecting quantitative radiological features and joint mining of image features, genomics, and clinical data to support the discovery of synergistic patterns for precision medicine

4. ***Development of computational techniques for low cost digital pathology:*** We want to develop low cost techniques for point-of-care pathology assessment minimizing requirement of labs for initial diagnosis.

5. ***Developing image-based analysis tools for ophthalmologic diagnostics***: Machine learning can help in developing techniques with better usability for detection and grading of diabetic retinopathy and cataract ensuring better reach and coverage of public health systems.

6. ***Developing image driven AI for offering new patient services at PHC's:*** We intend to develop techniques for deployment of image based AI systems for better management of pregnant women, malnourishment of children and cardiac care.

7. Creating an AI based platform for responding to medical and other national emergencies – RAKSHAK platform

A.4 Research Challenges & Application Development

In this section, we shall highlight the specific problems and our proposed approach for solving them. These approaches have been designed in consultation with collaborators and research partners.

A.4.1 Enhanced Capability for Imaging Modalities

Medical imaging techniques, with the ability to look beyond visible surfaces, have become critical diagnostic tools. Apart from diagnosis, medical images are useful for intraoperative and postoperative monitoring, surgical planning, cohort studies etc. There exist several different modalities, such as X-Ray, Computed Tomography (CT), Magnetic Resonance Imaging (MRI), Ultrasound Imaging and Polarization Imaging, which use different characteristics of electromagnetic and mechanical waves and serve different purposes. A common challenge with all medical images is their interpretation, which requires both expertise as well as experience. In addition, their capability of generating 3D volume data comes handy; further

increasing the interpretation challenge. For example, the appearance of cardiac anatomy on ultrasound can sometimes be confusing. The projects that we plan to undertake in this vertical range range from improving the quality of images captured using existing imaging techniques such as ultrasound waves, to designing computer vision algorithms to synthesize one modality from the other.

Improving the Quality of Diagnostic Imaging: Medical ultrasound has a wide range of applications. It is a convenient and low-cost imaging technique. Unlike other medical image modalities, it uses safe ultrasound (US) waves for imaging (Ortiz et al., 2012). However, despite its many advantages, the poor image quality of US images is always a concern. US images experience quality degradation due to the stochastic patterns being generated from the constructive and destructive interferences of the backscattered US pulse echoes known as speckle. Computer vision techniques, both classical as well as recent deep learning techniques, can help in US image despeckling and interpretation [Mishra et al. 2018].

Proposal: iHub Drishti will develop a software suite for context and application driven feature detection and feature enhancement of USG images using deep learning techniques. In particular, we plan to deploy the software suite on a combination of hand-held USG devices and linking it to cloud based services in a framework of IOT based architecture.

Synthesizing One Imaging Modality from Another: In Indian hospitals, the most easily available imaging modality is the X-ray machines. Plain radiography and sonography have wider penetration and affordability compared to CT and MRI. This presents multiple challenges including availability of imaging technologies for detection and diagnosis and affordability. One of the many technologies that we plan to develop is to use plain radiography and sonography to synthesize the corresponding CT. Synthesis of the MRI head from a CT head is another project that we aim to work on. These would be of significant value at district hospital level, in particular, in trauma situations where time is critical. It is also helpful in the current pandemic situation where one wants to strive towards providing low-cost diagnostics minimizing exposure and contamination. Researchers today are working on these problems and the figures below show the state-of-the-art results, obtained using deep learning architectures.

(a)   (b)
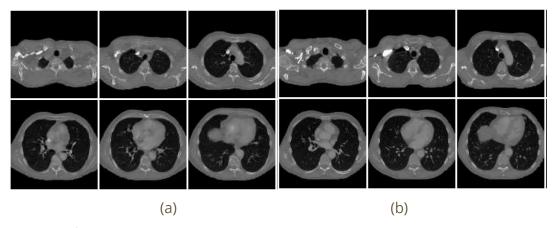
Figure: Examples[1] of (a) reconstructed CT slices and (b) corresponding groundtruth.
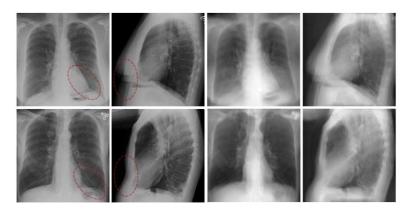
Figure: CT reconstruction from real-world X-rays. The first and second columns are real X-rays in two views. The following two columns are transformed X-rays by CycleGAN[9].

Proposal: We not only aim to generate accurate reconstructions, but also incorporate measures to explain the reconstructions. Often, nodes of interest like tumours are far too small in size in X-rays, and it is undesirable for these to be treated/neglected as noise in the reconstruction. We thus plan to incorporate semantic segmentation and explainability in the reconstruction process, such that the reconstructed images are useful for detection and diagnosis. Research efforts related to shortening acquisition times of different imaging modalities is another project that we intend to undertake; this would be useful from the perspective of patient compliance and limited resources.

A.4.2 Design of Advanced Image Based Dependable Diagnostic and Prognostic Tools for Dealing with Trauma, Chronic and Infectious Diseases

Diagnostic & Prognostic Systems: Disease diagnosis is one of the most important tasks of healthcare services and it requires knowledge and experience. As a part of this TIH, we plan to build efficient and accurate computer assisted diagnosis (CAD) systems using computer vision technologies. Computer vision algorithms can detect complex patterns within images that may go unnoticed even by trained clinicians and can significantly assist non-experts. In the current scenario, the expert doctors in secondary and tertiary care centers are overwhelmed and unable to provide time to patients who need it. On the contrary, non-experts require assistance in proper diagnosis. Computer vision assisted diagnosis systems will not only help non-experts with diagnosis, but it will also be of help to experts in optimizing their time and resources. Apart from diagnosis, medical images are useful for intraoperative and postoperative monitoring, surgical planning, cohort studies etc.

---

[9] Ying et al.; X2CT-GAN: reconstructing CT from biplanar X-rays with generative adversarial networks, IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019.

Diagnosis and Prognosis of Lungs Diseases- Silicosis: X-ray and CT scans are two preferred medical imaging techniques when it comes to lung examination in diseases such as silicosis. Chest X-ray, which is very common to diagnose shortness of breath, persistent cough, fever, pain and other such medical conditions, is the first step recommended by the physicians to evaluate silicosis, primarily due to the fast, convenient and cost-effective nature of X-ray. However, a major challenge with chest radiographs is their interpretation, which requires both expertise as well as experience. Silicosis is marked by appearances of nodular lesions in the upper lobes of the lungs. The figure below shows some examples of lung nodules and scarring originated due to silicosis.

(a)   (b)   (c)
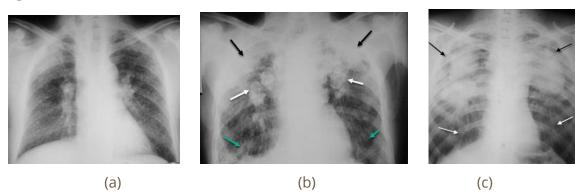
Figure: Chest radiographs showing lung conditions at various stages of silicosis. (a) Diffuse micronodular lung disease with an upper lobe predominance, (b) Progressive massive fibrosis, large conglomerate upper lobe (black arrows), enlarged and calcified nodes (white arrows), scarring in lower lobes (green arrows), (c) Conglomerate soft-tissue densities in upper lobes (black arrows) and scarring in lower lobes (white arrows). [Image source: http://learningradiology.com/notes/chestnotes/silicosis.htm]

The analysis of chest radiographs for silicosis mainly requires classification and grading of cases using X-ray / CT images and segmentation of regions of abnormalities (nodules, scarring) in the images. For example, an AI model can be built to discriminate between the X-ray images of Normal, TB and Silicosis patients. The classification of images is a challenge given the fact that silicosis is often misdiagnosed as pulmonary edema (fluid in the lungs), pneumonia, or tuberculosis. Additionally, the model can also delineate regions in the images containing damaged tissues. It can predict the possibility of future progression of a disease and provide recommendations like change in lifestyle etc. In general, such segmentation is manually performed by the clinicians, which reduces objectiveness of the diagnosis and also makes the process labour intensive. Therefore, an automatic framework for segmentation is desirable. CNNs will be used for both automatic classification as well as segmentation of these images. Apart from X-ray, CT and pathology images can also be used for similar analysis or together for multimodal analysis. Further, time series data recorded using various sensors over multiple visits of the patients will be used for cohort studies and development of prognosis models.

Diagnosis and Prognosis of COVID-19 using Clinical Data and Imaging Data: With the increasing numbers of COVID-19 cases in the country, it is crucial to increase the number of individuals getting tested so that appropriate measures can be taken. The traditional way of COVID testing is a time taking process with a result of confirmation taking more than a day. Due to the mismatch in the availability of testing kits and the number of individuals to be tested,

it is essential to find additional accurate ways of testing and screening. Further, an accurate estimate of coronavirus infection progression and the level of severity can be of significant help for optimized treatment and morbidity reduction.

AI-driven Diagnosis of COVID-19 using Chest X-ray images: This project aims to address the problem of COVID-19 detection using cues present in frontal chest X-ray images. Chest X-ray contains vital information such as the presence of opacities, mismatch in lung ratios, and cues from bronchi. While these details are prominent, we currently have limited samples to train the deep learning model. This project aims to build efficient computer vision and machine learning models that not only predict whether the chest X-Ray has symptoms of COVID-19 or not, but it also presents the regions and characteristics based on which the model has given a particular prediction to ensure explainability.
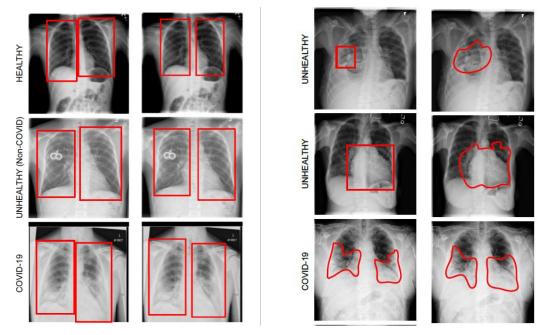
Figure: Lung segmentation and semantic segmentation in chest x-ray images for healthy, unhealthy, and COVID19 individuals. The first and the third columns represent the ground truth annotation, while the second and the fourth columns present the prediction outputs.

AI driven COVID-19 prognosis estimation using multimodal data: We aim at the prognosis of COVID-19 infection based on the longitudinal studies on multimodal data including X-Ray images, X-Ray reports, CT scans, clinical data such as body temperature, and treatment given to the patients. Our main goal is to predict the severity of infection in lungs due to COVID-19 for subsequent days given the data for current days, primarily to augment the decisions of physicians. We will also consider the effect of the presence of comorbidities on COVID-19 prognosis and the impact of COVID-19 on comorbidities and co-existing diseases. For this purpose, we propose to build a computer vision and artificial intelligence driven hybrid model using a combination of advanced deep learning methods and belief networks. Deep learning has shown success in accurate diagnosis of many diseases such as pneumonia and also in developing prognostic biomarkers for diseases such as colorectal cancer. Very recently, studies have been reported to demonstrate effectiveness of AI tools in COVID-19 diagnosis, however, to the best of our knowledge no work has been reported to build an AI driven prognosis model of COVID-19. The figure below summarizes the overall structure of the proposed tool.
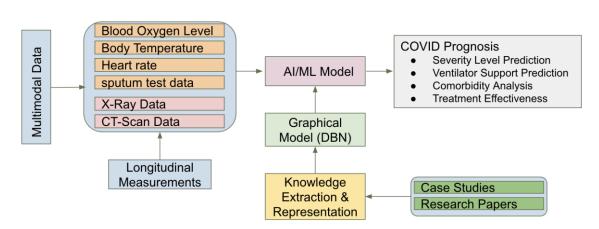
Figure: COVID19 prognosis using imaging and clinical data with knowledge representation

Tools to predict survival and post-traumatic long term morbidity: Traumatic brain injury (TBI) is a significant healthcare and economic problem in India (LMICs), accounting for a major burden regarding morbidity, mortality, disability, socioeconomic losses, reduced life expectancy, and the quality of life. Road traffic accidents account for over 60% of all traumatic brain injuries, followed by falls (30%) and violence (10%), with the most affected age group being between 21 and 30 years old. Although being able to predict a patient's prognosis at presentation accurately is a significant step in clinical decision making and the assessment of the quality of care, there are few prediction models that have used machine learning methods based on data directly captured from hospital registries. Our objective is to develop a machine learning predictive model based on a possible prospective registry with consecutively enrolled TBI patients. It is our goal to design highly predictive models that are intrinsically interpretable.

Dependable Diagnostic and Prognostic Systems: Explainable diagnostic systems are expected to explain the basis for diagnostic decisions. In addition, robust explainable diagnostic systems are required to answer counterfactual questions, and be robust to significant data distribution shifts and uncertain and erroneous measurements.

Computer aided diagnosis systems or Clinical decision support (CDS) systems today operate in complex socio-technical environments - comprising data, software, hardware, networks and people. Computer vision, artificial intelligence and machine learning driven CDS systems are being envisioned as mechanisms to augment the capabilities of doctors, clinicians, staff, patients or other individuals with knowledge and person-specific information, intelligently filtered or presented at appropriate times, to enhance health and health care. It is thus essential that these environments run in sync and that an aura of trust or dependability - across man and machines - exists in the ecosystem. Errors or lack of transparency in CDS-based decisions / inferences with no aspect of explanability or traceability embedded therein, can have catastrophic effects.

WIth the above in mind, inculcating the practice of designing dependable systems is one of the primary objectives of iHub Drishti. We shall be working towards addressing the following critical aspects with regards to designing healthcare systems:

- Bias and fairness,
- Explainability and interpretability,
- Governability and robustness,

- Trustworthiness, and
- Privacy and security

The envisioned ML-based system (refer figure below) will incorporate causal deep learning-based models, such as Deep Bayesian Networks, and temporal neural networks, like Dynamic Deep Bayesian Networks, as example, for prognostics of oral cancer using longitudinal multimodal medical data - like CT-Scan, MRI scan, genomic data, Hyperspectral imaging data, and clinical measurements - of the patient to generate an explainable prognosis of oral cancer.
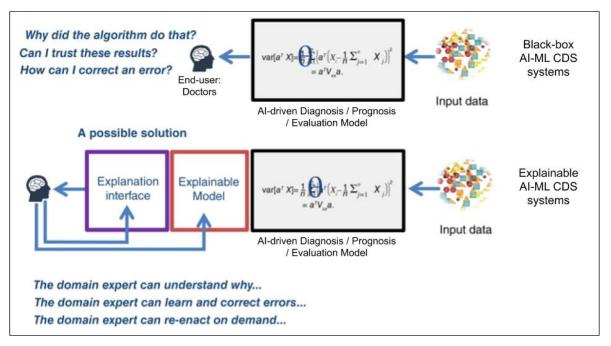
Fig. Figurative representation[10] of the dependable healthcare systems to be designed under the TIH.

A.4.3 Radiomics and Radiogenomics: Technology for Precision Medicine

Precision medicine aspires to leverage new knowledge emanating from heterogeneous imaging, genomic, environmental, and clinical data analysis, facilitating increased understanding of disease progression, treatment efficacy, and prevention, towards developing new, personalized therapies and interventions.

Radiomics research uses quantitative imaging methods to associate multi-modal image features to clinical outcomes. Radiogenomics research extends radiomics scope by investigating the association of these features to –omics analysis, therefore creating imaging phenotypes as surrogate, non-invasive biomarkers. Moreover, by creating radiogenomics signatures. That is, the development of joint models leveraging imaging and genomic features that better describe and correlate to clinical outcomes (features), and as a result can be used for enhancing prognosis, diagnosis, and treatment response. Naturally, radiomics and readiogenomics studies share the same methodological concepts.

---

[10] Adapted from Holzinger, A., Langs, G., Denk, H., Zatloukal, K., & Müller, H. (2019). Causality and explainability of artificial intelligence in medicine. In *WIREs Data Mining Knowl Discov.* (vol. 9, e1312).

Large-scale datasets for radiogenomics research are currently available through open research data initiatives, such as the Cancer Imaging Project (CIP) and the Alzheimer's Disease Neuroimaging Initiative (ADNI). These datasets provide multi-modal imaging, genomic, and clinical data. In this project we consider CIP and ADNI as case studies with the expectation that methods developed for these datasets will find wider applications in other diseases as well.

Refer to the figure on the right for an abstract system diagram that depicts the benefits of adopting radiogenomics methods for precision medicine. Based on the disease, we use the corresponding (multi)-modal imaging approach. Following manual and/or automatic segmentation of the clinically important regions, quantitative and/or qualitative features are then extracted. Radiogenomics investigates the relationships between imaging and genomic features and how radiomics and genomics signatures are associated with clinical outcomes. Radiomics is focused on characterizing the relationship between quantitative (multi-modality) imaging and clinical features (Panayides et al. 2020).

We propose to develop deep learning based dependable algorithms for quantitative image feature extraction. Joint analysis of image features and multi-omic data will be done in collaboration with CSIR-IGIB with participation of CARING & Teleradiology Solutions.

- One of the key tasks is to determine the Diagnostic Region of Interest (ROI), for example tumour region for brain tumour case.
- Next challenge is to build models for mapping quantitative imaging phenotypes to clinical receptors which can be used for exploring construction of prognostic indicators.
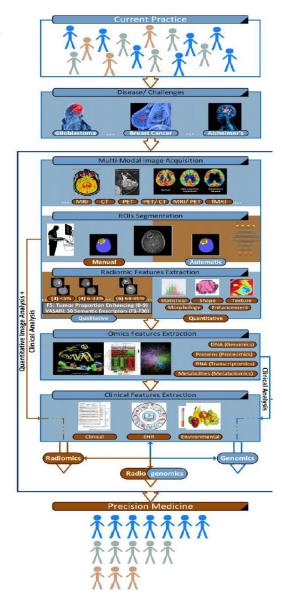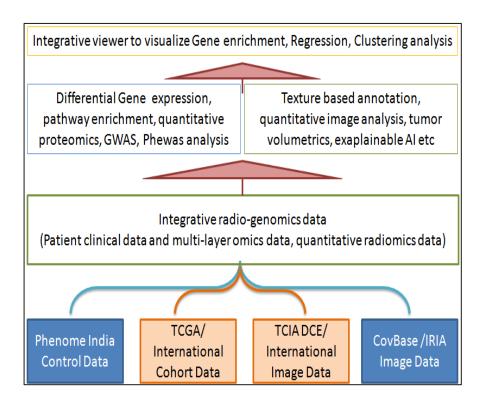
Fig. Radiogenomics System Diagram

- A challenge in radiogenomics lies in the lack of uniformity of imaging data. Images acquired using different scanners, scanning parameters, protocols, and by different technologists can vary significantly in appearance which can cause imaging variables extracted from images acquired in one setup to be incompatible with images acquired using another setup. Uniform acquisition of images is likely to be an unreachable goal in the near future. We shall develop techniques to normalize the imaging data in order to bring it to a common framework. We believe that image normalization is one of the most important future directions related to radiogenomics and significant work is needed before it reaches maturity.

Ultimately, the goal is to develop imaging biomarkers for outcomes that incorporate both phenotypic and genotypic metrics and provide useful applications of radiogenomics in brain, liver, prostate, gynecological, and breast tumors etc. We propose to build an integrative data and computational architecture for AI driven personalized diagnosis, prognosis and therapy design as shown in the following figure.

Connectomics

Connectomics is another source of quantitative information not much explored in radio-genomics which has a potential for yielding interesting insight.Connectomics is the study of whole brain association maps, i.e.,the connectome, with a focus on understanding, quantifying, and visualizing brain network organizations. Connectomics research is of interest to the neuroscientific community largely because of its potential to understand human cognition, its variation over development and aging, and its alteration in disease or injury. As such, big data in connectomics are rapidly growing with emerging international research initiatives collecting large, high-quality brain images with structural, diffusion, and functional imaging modalities. Large, open source datasets, such as the Human Connectome Project (HCP) and the Autism Brain Imaging Data Exchange (ABIDE), have spurred the development of new and increasingly powerful machine learning strategies in connectomics. This TIH, in collaboration with AIIMS and other institutes including NBRC (National Brain Research Centre), will work on connectomics to explore and understand cognitive networks for devising techniques and methods for better diagnosis and treatment of diseases linked with cognitive disabilities and cognitive degenerations. This study can yield significant input for creation of radio-genomics for diseases like Alzheimer's.

Human Phenotyping

An organism's phenotype is a function of its genotype, and the influence of environmental factors. Advances in human genetics (particularly genome-wide association studies) have led to the identification of large numbers of chromosomal loci that are responsible for specific human traits and disease. Radio-genomics exploits phenotyping of human beings. The Human Phenotype Ontology[11] (HPO) is a standardized vocabulary of phenotypic abnormalities encountered in human disease, that currently lists over 13,000 terms and over 156,000 annotations to hereditary diseases. Such data find application in clinical diagnostics – the understanding of disease mechanisms and susceptibility, mapping between phenotypes of model organisms, and as a standard vocabulary for clinical databases. The challenge that continually befronts us is the identification of new biomarkers of disease risk towards enhanced processes of disease diagnosis, prevention, and treatment. Primary obstacles [Ausiello & Shaw, 2014] in this direction are: a) the functional and medical significance of most susceptibility alleles is not well understood, and b) any disease trait is not only influenced by many genetic loci, but also by environmental and behavioural parameters.

Under the TIH, we would like to explore and contribute to the understanding of this correlation between biomarkers, environments, behaviours and disease, through the use of AI-driven computer-vision based techniques, as part of our integrative AI based Medicine. A possible direction of work would be to harness the convergence of mobile technology to facilitate collection of longitudinal crowdsourced inputs on actions and environment information, data from wearable sensors that track phenotypic markers (like gait, tremors, fatigue etc.), and sensitive imaging technologies for the identification and analysis of genomic, epigenetic, proteomic and metabolic markers in blood and tissue. We believe that an assemblage of these technologies – aided by explainable deep-learned based models to derive associations, reasons and disease progressions, would not only contribute to the collection of markers of health and disease data for entire human populations, but also lead to new avenues in biomedical research. Some possible fallouts could be contributing to the comprehension of the meaning of 'being healthy', monitor transitions from 'healthy' to 'unwell', as well as facilitation of early diagnosis and intervention. The TIH would ensure coverage of all sections of population and enforcement of strict privacy and ethical practices on the collection and usage of such information. Building dataset for Indian population for specific disease types is the most critical challenge for implementing radio-genomic models.

Using Radiomics for predicting toxicity and response to radiotherapy (AI-RAD)
(Proposed by IIT Kharagpur & TMC, Kolkata)

Approximately 70% of all cancer patients require radiotherapy. However, response to radiotherapy and toxicities of radiotherapy can be difficult to predict amongst patients with cancer. Multiple factors, including cancer biology and radiotherapy dose accumulation may be responsible for response. Toxicity to adjacent normal structures may depend on radiotherapy dose to these structures, but knowledge is currently limited to empirical evidence based on relatively small cohorts of patients. The science of using data from imaging to determine cancer biology (radiomics) and radiotherapy dose plans for prediction (dosiomics) of response and toxicity is nascent but there is strong initial evidence that these have the potential to improve

---

[11] Human Phenotype Ontology: https://hpo.jax.org/

our ability for prediction. However, so far, there have not been efforts to develop a broad mechanism of analysing this information across cancer sites and types. Our goal is to build a framework for radiomic and dosiomic data accumulation and processing using artificial intelligence (AI) tools that will be available at a national level for multi-institutional access for academic researchers and prospective testing in clinical trials in the future.
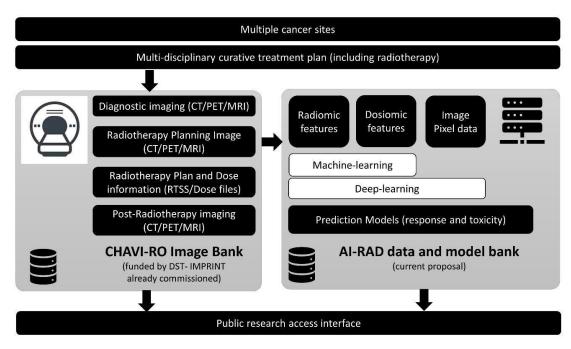
Fig.: A schematic diagram of the proposed framework

Outcome: Outcomes of this proposal is

1. The setting up of the data extraction, database creation and computational processes outlined above.
2. Model generation in selected cohorts (Table 1)
3. Development of a software for prediction of toxicity and response to radiotherapy under the proposed context (Table 1).

A.4.4 Digital Pathology: Sensors and Techniques

Digital pathology is a sub-field of pathology that focuses on data management based on information generated from digitized specimen slides or acquired through digital techniques. The field of digital pathology is growing and has applications in diagnostic medicine, with the goal of achieving efficient and cheaper diagnoses, prognosis, and prediction of diseases. In this section, we shall focus on developing sensors and algorithmic techniques which can facilitate point-of-care digital pathology solutions.

a. Multispectral and Hyperspectral Sensor Design and Sensing

Multispectral imaging acquires more wavelengths as compared to the conventional RGB (red, green, and blue) filter based optical imaging. It captures image data in a range of wavelengths, in some cases even beyond the optical spectrum. Hyperspectral imaging contains an even larger range of electromagnetic spectrum as compared to multispectral imaging.

Hyperspectral imaging (HSI) is a hybrid (combining imaging and spectroscopy) modality for medical applications, particularly for non-invasive disease diagnosis and image-guided surgery. HSI systems capture the reflectance, fluorescence, and transmitted light from tissues, which is representative of the quantitative diagnostic information of tissue physiology, morphology and composition. HSI systems generate three-dimensional hypercubes (Refer the figure below) (of hundreds of narrowband channels) – comprising two spatial dimensions and one spectral dimension. The spatial information helps locate the source of each spectrum, and the spectral signature of each pixel in the images enables identification of various pathological conditions. Application of computer vision techniques like spectral unmixing and classification [Lu et al., 2014] on HSI hypercubes enable precise segmentation, visualization and identification of tissue features, and cell morphology, structure and function.

RGB cameras are limited in their scope of representation; and interpretation of RGB microscopy images is often inadequate because of limitations in human vision to distinguish subtle colour differences arising from spatially overlapping emissions. HSI covers UV, VIS, NIR, and mid-IR spectral ranges which apply to specific medical problems. The MIR absorbance spectrum captures rich information about genomics, proteomics, and metabolomics of cells [Gendrin et al., 2008]. Visible light can penetrate to 1-2 mm below the skin and is thus effective in capturing information from subpapillary regions [Nouvong et al., 2009]. Light in the NIR region penetrates deeper into the tissue than VIS or mid-IR radiation, and is therefore preferred for surgical guidance [Zuzak et al., 2007]; seeing through connective tissues to observe critical anatomical structures. HSI finds application [Lu & Fei, 2014] in:

- Optical biopsy [Tuan, 2004] – in vivo diagnosis of tissue without the need for sample excision and processing.

- Mapping the spatial and temporal relationship between blood oxygen delivery and hypoxia at microvascular levels during tumor growth and angiogenesis [Sorg et al., 2005].

- Visualization of chemical contents of vessels and organs, and monitoring of tissue blood volume and oxygenation during surgery [Fawzi et al., 2011].
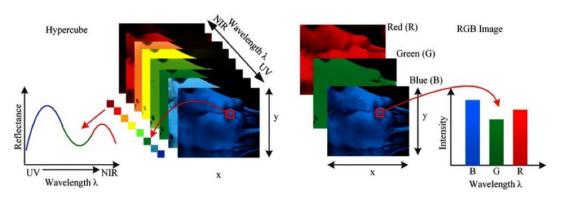
Figure: Difference between an HSI hypercube and an RGB image. [Lu & Fei, 2014]

The HSI technology describes an indirect mechanism of arriving at a spatial map of optical properties of tissues, since it attempts deduction of the interaction coefficients from measurements of reflectance and transmittance [Patterson et al., 2009]. This is essentially an ill-posed inverse problem with no unique solution [Kim et al., 2011]. It is possible for two media of substantially different optical properties to yield similar optical measurements [Patterson et

al., 2009]. In practice, it is rather difficult to eliminate ambiguities arising due to matching spectral profiles across biological samples.

Advanced classification algorithms [Grigoroiu et al., 2020] could facilitate precise delineation between healthy, premalignant, and malignant tissue for image-guided biopsy and surgery. Advanced spectral unmixing algorithms could offer insights into correlations between intrinsic biomarkers and disease states, as well as support identification of biomarkers [Lu et al., 2019] for early cancer detection by recovering subpixel compositional information.

Proposal: HSI based diagnostics

Prominent challenges of HSI based pathology – and some consequent focus-areas for the TIH are:

- Techniques for real-time acquisition of high-resolution HSI video datasets that capture and enhance subtle spectral and spatial dynamics of different tissue types; to aid in intraoperative imaging of organs, tissues, cells, and molecular biomarkers of interest.

- Design of explainable AI-driven algorithms for real-time processing of HSI data – extraction of high-quality diagnostic information, and generation of a quantitative map of different tissue types as well as disease-specific endogenous parameters. Such algorithms could be deployed in diagnostic systems for rural or remote health centers, where facilities are hard to come by. Generation of an ever-growing spectra database for significant molecular biomarkers and all tissue-categories including skin and subcutaneous tissue, ocular tissue, head/brain tissue, epithelial/mucous tissue, breast tissue, cartilage, liver, muscle, aorta, lung, myocardium, etc. Such a database would not only contribute to the design of traceable dependable algorithms, but also lead to the enhanced training of the same to accurately differentiate between oxygenated and deoxygenated blood, distinguish between all tissue categories, as well as identify molecular biomarkers for early cancer detection

Proposal: Multi-spectral Imaging from RGB

From the perspective of a screening or home monitoring device, multispectral imaging is an expensive option due to the high cost of multispectral cameras. An alternative is to reconstruct multispectral images from the RGB images using state-of-the-art processing/machine learning techniques of image reconstruction.

In the past, studies have been conducted to understand the phenomenon of light-muscle interactions [Ozgulbas, 2013; Shuaib et al., 2011].

Figure: Multispectral vs hyperspectral imaging [Issa et. al., 2015].

The blood composition - as a ratio of water, haemoglobin etc. - affects the absorption and scattering of light passing through it. The amount of absorption [Ozgulbas, 2013] and scattering measured using polarization properties of light [Shuaib et al, 2011] can be used to differentiate

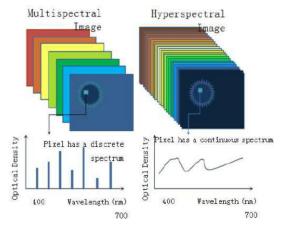among different elements of the blood and evaluate blood-oxygen saturation levels (SpO_2). Most of these studies have been restricted to RGB wavelengths. Construction of multispectral images from RGBcan provide a cheap option for constructing multi-spectral output from RGB imaging. Our objective here is to extend these studies to multispectral wavelengths and develop devices like non-invasive non-contact devices to measure real-time SpO_2 which would be useful for neurological and stroke patients, and athletes.

b.   Fluorescent Light Based Imaging for Oral Cancer Screening

Oral cancer is amongst the top three types of cancers in India. The incidence of oral cancer is highest [Dikshit et al., 2012] in India, south and Southeast Asian countries. In India, 20 per 100,000 people are affected by oral cancer; thereby accounting for approximately 30% of all types of cancer. Over 15 people in India die every hour per day because of oral cancer, cancer in the oropharynx and hypopharynx. Oral cancer primarily affects people in rural areas due to higher exposure to risk factors such as the use of tobacco and lack of health care services. Patients with a history of oral cancers are at risk of developing second primary tumours at a rate of 3.7% per year because of 'field cancerization', and a quarter of all oral cancer-related deaths are caused by second primary tumors [Day et al., 1992]. Hence, patients who are successfully treated for oral cancer should be closely monitored, preferably using a non-invasive diagnostic test. No significant advancement in the treatment of oral cancer has been found in recent years, and the overall survival rate of 5 years has not improved in the past decades. Fortunately, oral cancers are mostly preceded by oral precancerous lesions before malignant transformation which takes around 5-7 years. Early detection of oral cancer offers the best chance for long term survival and has the potential to improve treatment outcomes and make healthcare affordable.

Fluorescent-light based imaging tools are effective for oral cancer screening, especially in rural areas with relative lack of oral cancer screening mechanisms. Direct visual inspection of tissue autofluorescence has shown encouraging results in high-prevalence populations. However, the technique relies on visual recognition skills of the examiner.

The TIH thereby aims to develop a portable, low-cost ([Skandarajah et al., 2017; Cybulski et al., 2014]), robust, explainable CV-driven fluorescent imaging and diagnostic system on edge.

A.4.5 Image-based tools for ophthalmologic diagnostics

Age-related macular degeneration, diabetic retinopathy, and glaucoma are the main causes of blindness in both developed and developing countries. The cost of blindness to society and individuals is huge, and many cases can be avoided by early intervention. Early and reliable diagnosis strategies and effective treatments are therefore a world priority. At the same time, there is increasing research on the retinal vasculature and neuro-retinal architecture as a source of biomarkers for several high-prevalence conditions like dementia, cardiovascular disease, and of course complications of diabetes. Automatic and semi-automatic software tools for retinal image analysis are being used widely in retinal biomarkers research, and increasingly percolating into clinical practice. Significant challenges remain in terms of reliability and validation, number and type of conditions considered, multi-modal analysis (e.g., fundus, optical coherence tomography, scanning laser ophthalmoscopy), novel imaging technologies,

and the effective application of advanced computer vision and machine learning technologies, to mention a few. This TIH will work on these aspects in collaboration with medical institutes to advance state of the art in image diagnostics in this field.

Diabetic Retinopathy (DR)

DR is one of the most common causes of vision loss in the world and it can be prevented by early detection and treatment. Currently, India is the diabetes epicenter of the world and is projected to have 109 million diabetics in the next 20 years. Consequently, early detection and treatment of DR is critical in minimizing the rate of the blindness epidemic in high-burden areas. DR screening and early detection programs, utilizing non mydriatic portable fundus imaging for screening and grading by non-ophthalmologists, have been widely adopted in many countries. However, with the increasing diabetes prevalence and limited numbers of trained graders, the implementation of DR screening programs employing human graders is challenging. Thus computer vision based predictive algorithms for the automated diagnosis and grading of DR has immense potential in the context of the public health settings; providing critical services, enhanced clinical decision-making and diagnostic accuracy, while reducing health system burden and augmenting human resources. The TIH will focus on designing such algorithms for screening, detection and grading of DR.

Cataract Detection

Cataract is a primary cause of visual impairment worldwide and cataract surgery is the commonest elective surgical intervention. There have been coordinated national efforts. As a result, from around 1.2 million cataract surgeries per year in the 1980s, the cataract surgical output increased to 3.9 million per year by 2003. Though the prevalence of cataract blindness is projected to decrease, the absolute number of the cataract blind would increase from 7.75 million in 2001 to 8.25 million in 2020 due to a substantial increase in the population above 50 years in India over this period. These numbers reflect the tremendous increased demand on health care services. Currently, there are no specific thresholds or criteria that define appropriate patients for referral and/or cataract surgery. Referral methods and pathways to surgery vary greatly and there are no definitive guidelines. Due to the lack of systematic screening, information transmission and robust referral within frontline healthcare delivery settings, this project aims at combining expertise of computer vision researchers, physicians, public health workers, to design and implement an efficient and accurate 'Cataract Detection on Mobile Platform', to help screen and quantify wait time and individual risk for patients being considered for cataract surgery. Secondly, it is recommended that surgery should not be offered to patients who are unlikely to gain sufficient benefit to outweigh the risk involved. Therefore, adaptation of appropriateness criteria to clinical practice guidelines and the production of electronic decision-making aid could assist in measuring the appropriateness of cataract surgery recommendation.

Several other ophthalmology related problems which include vision-based assessment can be taken up in this TIH. For instance, automated glaucoma detection and intelligent laparoscopy are some other problems that can be addressed using computer vision technologies.
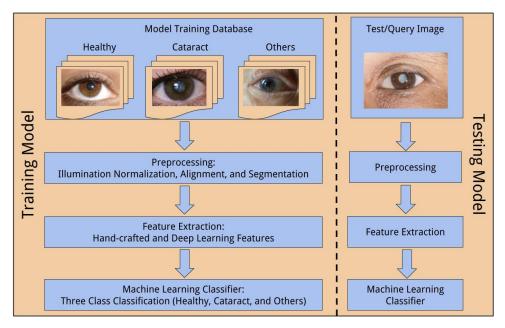
Figure: Block diagram illustrating the proposed model for cataract detection.

A.4.6 Computer Vision for Improving Public Health Related Services

According to the World Health Organization (WHO), in 2015 around 303,000 women died around the globe due to pregnancy complications and in 2017 there were 5.4 million under-5 deaths. Most of these deaths could have been prevented with proper health tracking and care. Therefore, low-cost primary health care systems are essential and play a critical role in achieving the sustainable development goal of "Good-health and well being". Problems that need to be solved to avoid such deaths include monitoring the health condition of pregnant women (blood pressure, anemia), monitoring the growth of infants and newborns, and their nutrition levels, ensuring clean water and sanitation, etc. Other major causes of death are cancer, malaria, tuberculosis, cardiovascular diseases, respiratory diseases, etc; most of which can be avoided through timely intervention. This calls for early and accurate disease-diagnosis and progression-mapping.

The Indian government as well as global organizations like the WHO are running several different schemes to promote women and child welfare in rural areas. The outcomes of these programs are also visible through reduced mortality rates and increasing awareness in the community. However, shortage of health workers adequately trained in the detection and triage of suspected cases, especially in the rural areas, continues to remain a major contributor towards maternal mortality. A critical clinical challenge in this regard is to ensure availability of a detection tool that is affordable and that can be easily applied to effectively improve decision making skills among health workers; enhancing women's linkages to care without heightening healthcare burden. Therefore, under this vertical of the TIH, we would like to build systems that could be installed at primary health centers to assist doctors in diagnosis and accurate determination of the stage of severity or progression of common diseases. Considering the low doctor-patient ratio in rural areas of India, such automatic processes would enable doctors at primary health centers to treat more patients per day.

We will use computer vision and machine learning techniques across different modalities of medical data obtained from patients, such as clinical measurements (heart rate, respiratory rate, lung and heart sounds, blood oxygen saturation levels, body temperature, blood test

measurements, ECG data, EEG data, etc.) and medical imaging data (X-Rays and other available modalities), to accurately diagnose and stage disease progression. Furthermore, we will enrich the developed systems to analyze a patient's follow-up data to monitor the progression of the severity of diseases and automatically suggest possible treatments. We will also build systems for automatic screening and diagnosis of diseases caused by viral outbreaks, such as COVID-19, for better handling of such pandemics in the future. Some of the specific projects that we plan to undertake are discussed below.
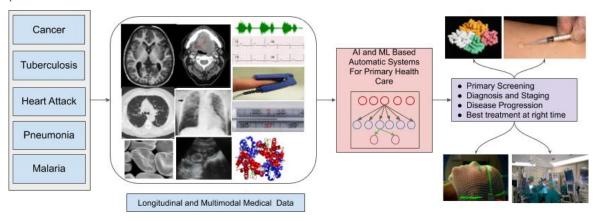
Figure: Flowchart of automatic system for primary health care. [Images credit: Wikipedia]

Detection of Preeclampsia in Pregnant Women

Preeclampsia, occurring in 5–8% of pregnancies worldwide, is among the most important causes of perinatal and maternal mortality and is also a leading cause of fetal complications, which includes low birth weight, premature birth, and stillbirth. It can affect pregnant women before or during pregnancy and is related to increased blood pressure (BP≥140/90 mmHg) and protein in mother's urine (urinary albumin protein ≥300 mg/24 h). The clinical spectrum of preeclampsia ranges from mild to severe. Life-threatening complications due to preeclampsia are difficult to predict and prevent. The detection of the classic form of the disease is difficult since its natural progression is still not well understood, and the relative importance of degree and timing of the symptoms (hypertension, proteinuria, edema or other biochemical abnormalities) is unclear during pregnancy. Preeclampsia often remains unrecognized because of fragmented healthcare services, shortage of trained staff for screening, and situational barriers like poor access and uptake of health specialty services. Given the public health significance of preeclampsia and the limited resources available for providing evidence-based treatment, there is a need to develop effective screening tools that aid health workers in making timely clinical decisions that are often missed or delayed.

Malnourishment Detection in Children

In 2016, around 101 million children below the age of 5 years were underweight[12]. According to The State of Food Insecurity in the World, in 2017, there were over 800 million undernourished people in the world, which is about 11% of the total world population[13]. Apart from malnourishment, there are about 155 million children under five who are suffering from stunting. Poor living conditions are responsible for approximately 23% of all deaths in children

---

[12] UNICEF-WHO-The World Bank Joint Child Malnutrition Estimates. Levels & Trends in Child Malnutrition.

[13] Food and Agricultural Organization of the United Nations. The State of Food Insecurity in the World 2018.

under five, which is about 323,000 deaths in 2015[14]. Furthermore, several countries have very high Infant Mortality Rate (IMR). As per 2017 data, almost 70 countries in the world have over 25 IMR, implying we lost a large number of infants undernourishment and weaknesses in the healthcare system.

Currently, there are standard techniques approved by the World Health Organization (WHO) to identify malnourished children. These are generally performed by medical practitioners and involve use of visual signs and weight to height ratio. Clinical approaches involve analyzing visible signs on the body of individuals such as prominent bones, skinny limbs, or swollen belly. There are different levels and characteristics of malnourishment: prominent bones, skinny limbs, and loose skin are primary visual signs of marasmus, while bilateral pitting edema, hair color, skin color, are major signs of kwashiorkor[15]. Another method which is standardized by the World Health Organization is based on the wrist width calculated using Mid-Upper Arm Circumference (MUAC) tapes. This method is mainly used for children of age 12 months up to the age of 59 months. However, these methods have three key shortcomings:

- Cumbersome measurement process - these methods are difficult to scalable as they require trained medical practitioners to manually perform these tasks.

- Readings are recorded in paper-based formats - despite some pilot initiatives in digitizing this, paper-based recordings remain the mainstay in most low and middle-income economies.

- Interpretation of charts and follow up interventions require significant training of the frontline staff.

Two major Sustainable Development Goals (SDGs) by the United Nations are (i) Zero hunger (Goal 2) and (ii) Good health and well-being (Goal 3). These two goals highlight the importance of good health and well-nourishment in children. One of the major aspects to achieve these goals is an accurate and fast measurement approach which is scalable and inexpensive. The central government, state government and several other NGOs are exploring the use of mobile phones to increase awareness and accessibility of information and facilities in the villages. Towards this objective, this proposal aims to build a computer vision based system to measure malnourishment along with classifying its type using photographs captured by mobile phones. This tool can be integrated into other such mobile systems to provide an overall framework of services related to children.
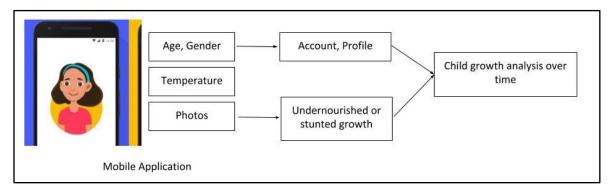
Figure: Illustrating the flow of the information in the proposed tool.

[14] GBD 2015 Mortality and Causes of Death Collaborators. A systematic analysis for the Global Burden of Disease Study 2015.
[15] Mother and Child Nutrition. Early Detection and Referral of Children with Malnutrition.

Prediction of Coronary Artery Disease in Primary Health Care Settings

Cardiovascular disease is still the main cause of death in the world, and coronary artery disease (CAD) is the largest contributor. Screening asymptomatic individuals for CAD in view of preventive treatment is therefore of crucial interest. The Global Burden of Diseases Study reported that the disability-adjusted life years lost by CAD in India during 1990 was 5.6 million in men and 4.5 million in women; the projected figures for 2020 were 14.4 million and 7.7 million in men and women, respectively. The prevalence of CAD in the US is 2.5%. The prevalence of CAD risks in Indians living in India is 11% for non-diabetic patients and 21.4% for diabetic patients A study showed that ischemic heart disease, a manifestation of CAD, has a relative risk of 1.0 for Chinese population, 1.8 for Japanese population and 6.6 for South Asian population (all living in US) compared to the USA national average. Atherosclerosis is a slowly progressing disease, characterized by a long asymptomatic period of several decades. Risk scores like Framingham risk score are relevant to the population in which these have been studied. In Indians the disease occurs earlier, and is more malignant. Apart from the traditional risk scores new biomarkers and imaging methods have emerged which will help in early detection of the disease as well as risk stratification of patients at "high risk". Under this project, we aim to develop a risk score for patients at high risk of developing CAD, and a strategy for risk stratification and treatment in patients with stable CAD.

A.4.7 Creating an AI based Platform for Responding to Medical and other National Emergencies – RAKSHAK Platform

RAKSHAK Project for COVID19

The project Remedial Action, Knowledge Skimming and Holistic Analysis of COVID-19 (RAKSHAK) attempts to find quick and long-lasting solutions of problems faced by different segments of the society from the Covid-19 pandemic using Artificial Intelligence (AI) as the enabling core technology. AI has, no doubt, the potential to control the COVID-19 pandemic and help to curb its worst effects.

The RAKSHAK project attempts to create a unique platform that allows academia, scientists, clinicians, industries, and start-ups to work together to develop AI based innovative solutions to mitigate the effects of pandemic's. It is coordinated by a team of academic experts, startup founders, and startup investors. The main main objectives are:

1. Setting up of a distributed and accelerated cloud-based IT infrastructure for collaborative development

2. Create a data-bank of Covid-related data from sources including clinical agencies, research institutions, public health agencies, etc. The databank and associated tools will facilitate open innovation and serve as the foundation for AI-based solutions.

3. Develop, translate and deploy AI-based solutions for dealing with COVID for the benefit of all segments of the society through a consortium of researchers, developers, startups, and industry.

4. Creating the nodal AI Technology group for enabling the ecosystem for application-based technology development, deployment and transfer to respond to the challenge of medical or other national challenges calling for AI-based solutions.

Expected Services

Apart from creating the cloud-based computation platform and a comprehensive databank, RAKSHAK concentrates on the following three major themes that are necessary to combat Covid and to restore the society rapidly after the pandemic concludes:

1. Diagnostics: Technologies for rapid and reliable diagnostics. One of the key focuses of the project is to develop diagnostic technologies. Powerful early diagnostic techniques have the potential for identifying and isolating cases fast to restrict spread of the disease. This effort is divided into projects that handle different aspects of the problem: Early Diagnostics and Monitoring, Medical Imaging based Diagnostics & Prognosis, Tracking and Surveillance, AI for Test management, Remote Diagnosis using AI. Multiple projects are proposed for each of the above aspects.

2. Advanced tools: Tools and techniques to safely manage the post-lockdown scenario and. It is important to handle the post-lockdown period with healthy social interactions and ample help to enable people to live through the new normal in life. Several projects have been proposed in the following sub-categories under RAKSHAK: Marketplaces and Supply Chain, Medical & Home Condition Management, Public Social Distancing, Sanitation, and Screening & Authentication.

3. Enabling Technology: Technology tools to help restore economic activity faster and equitably. Restoring the economic activity rapidly is of critical concern as several lives have been thrown into disarray by the pandemic and the accompanying lockdown. AI-based tools and techniques can help with this. It is also important to not violate the privacy and dignity of individuals in an effort to achieve these goals. Several projects are proposed to measure the impact, control the situation, and improve the activities. The success of these aspects determine the speed of recovery of the economy for the medium and long terms.

Consortium: The project RAKSHAK will be executed through a consortium of academia, researchers, and industry. The unique execution model is designed with the ease of transferring the solutions developed to the society through startups and the industry.

1. Researchers: A total of 50 Academic researchers from the institutes like IITs, IIITs, IISc, etc., are the primary investigators of the projects.

2. Domain knowledge partners: The project will also have active collaborations with agencies like ICMR and PHFI; medical sciences research institutions like IGIB-CSIR and SCTIMST; hospitals' like AIIMS;

3. Industry partners: Companies like TCS, Persistent, Nvidia; and startups like Algo8, AI Foundry and others.

Expected Outcomes

1. Short-term Outcomes

Diagnostics: Deployment of AI-based technological solutions to augment diagnostics and prognosis for COVID-19 patients, Machine interpretation of mass X-Ray data, Rapid analysis, Test Kits data and other medical sensor data.

Real-time surveillance: Implement intelligent real-time surveillance, containment and tracking of health of quarantined individuals.

Analytical insights for effective Governance: Manage issues after the lockdown and provide useful insights to facilitate quick decision making.

2. Long-Term Outcomes

Strategic Group & First responders: Beyond the immediate technologies and tools to alleviate the effects of Covid, project RAKSHAK envisages the establishment of a nodal group for AI-based solutions in emergency situations that the nation may encounter in the future. Technology and Artificial Intelligence will become important tools in responding to all sorts of crises in the coming years. The country needs to be ready with the required computation platforms, software tools, academic experts, industry teams, etc., to rapidly respond to future emergencies. We expect RAKSHAK to provide a start towards establishing such an entity.

Idea to Impact employment to wealth creation: The long-term outcomes will be also an opportunity to create an "impact innovation ecosystem", where universities, research translation parks and innovation studios work closely with the venture ecosystem to convert problem opportunities into unicorns of tomorrow. With close collaboration via the RAKSHAK "Idea to Impact" innovation engine, we can not only quickly create technology that can be used on the ground quickly but also set up an engine that can work for future advancement of Make-In-India objectives with cutting edge companies and technologies. If we can create such technology companies with potential high gross margins and retain them in India, we can create a large tax base to drive the social schemes that are needed in a diverse country like India. Otherwise if we only have businesses that have margins in the 8-10% range, our tax bases are always going to be small and we will struggle to invest towards the $5tn to $10tn mark. This close collaboration engine will spur the next generation of companies that are going to be sorely needed for India's post Covid19 resurgence as a strategic global leader in the science and technology domain.

IIT Jodhpur has been designated as the coordinator for the RAKSHAK project, which is the COVID19 initiative by the DST. Under this project, during the pandemic and post-COVID19 situations, we aim to pursue a multi-dimensional approach for monitoring, diagnosis and prognosis. Some of the projects that we have undertaken under RAKSHAK are as follows:

Non-invasive Estimation of Vital Parameters at Real-time Using Hyperspectral Reconstruction of RGB Videos:

Hyperspectral reconstruction from RGB images, based solely on computational techniques, is a technique that attempts to infuse and augment common imaging systems with capabilities particular to the hyperspectral domain. This technique is envisioned to facilitate interpolation of RGB images/information into hyperspectral images with higher resolution for possible extraction of finer (higher band-specific) information from RGB data for critical analyses, and to serve as a potential basis for resolution of trade-offs between spatial-structure and spectral

fidelity. The question that confronts us is, given the scope of hyperspectral reconstruction, can we utilize the method for the real-time non-invasive non-contact abstraction of vital health parameters (like core body temperature - despite ambient temperatures, blood-oxygen levels, heart-rate and respiration rate) from RGB videos? Besides offering a low-cost reproducible mechanism for the non-invasive detection of COVID19 patients across sectors, such systems could also be of use for the unbiased holistic screening and monitoring of individuals living in critical habitation zones, in-patient wards, homes for the elderly and other such areas. With this vision in mind, our activities can be broadly enumerated as:

- Approximation of core body temperature from tear-duct regions of face

- Approximation of blood-oxygen saturation levels (as derived from heart-rate and respiration-rate) from facial regions (like, cheek, forehead, nostrils, and mouth)

- Integration of solutions for the above problems, into an all-encompassing real-time screening/monitoring system.

Automated Detection of COVID-19 Suspects from Crowd through Thermal Sensed Video Surveillance

Silent hypoxia (low blood-oxygen saturation levels with no external indicators like shortness of breath) has been identified as a primary symptom of COVID 19 [Ottestad et al., 2020]. Thus, a non-invasive mechanism to detect individuals suffering from silent hypoxia, particularly in crowd-settings, is one of the fundamental aims under the RAKSHAK initiative. Given the current scenario, where most individuals in a crowd can be assumed to be in masks, the envisioned computer-vision based system would be capable of tracking of individuals on the move, and subsequent estimation of the rate of change in temperatures [Negeshi et al., 2020; Kim et al., 2018] - using thermal cameras - around the cheek-forehead-mask regions as a consequence of the rate of blood-flow to approximate heart rates, respiration-rates and blood-oxygen saturation levels, per person. People with low blood-oxygen saturation levels would be immediately identified as potential suspects of COVID19 and subject to further screening.

Detection of Individuals Wearing Masks in Crowds

In a bid towards control of the spread of COVID19 and practice of good respiratory hygiene, all organizations have mandated the use of masks. Under such circumstances, an AI-driven mechanism for the real-time detection of individuals in masks, and that would also raise an alarm for those without, would indeed be of use across organizations. Under the RAKSHAK initiative, arriving at a comprehensive, unbiased AI-driven solution for this problem is one of our major objectives. Modules for detection of human-beings on the move, crowd-counting, segmentation of individual faces and facial regions, followed by detection of occlusion of the nose-mouth region - as indicative of wearing a 'mask' or 'cover', would be some of the system components.

Person Identification with Masks

In this time of COVID-19, recognizing individuals for providing access is a major challenge. A lot of organizations have attendance mechanisms based on biometrics. Until now, fingerprint has been one of the most used modality for access control, however, it requires touch or close contact with the sensor, which is not recommended. Since face recognition technology operates at a distance, it is one of the most viable options for authentication during the pandemic and post-pandemic. While there are very accurate face recognition algorithms that

provide high accuracy for controlled applications like time and attendance, they generally rely on the availability of more than 60-70%% of the face being visible for processing. The use of masks covers more than 50% of the face and only the upper one-third part of the face is left for recognition. Therefore, it is important to design solutions for person recognition using the remaining area. We aim to design an authentication system that first incorporates face and mask detection so that the facial area visible for recognition can be extracted followed by designing an algorithm that utilizes the periocular and the forehead region for recognition. Our team members have significant experience in biometrics, and are at the forefront of promoting and designing technologies for face recognition with disguise accessories.

Detection of Practice of Social Distancing in Crowds

As a possible precautionary step towards preventing the rapid spread of the COVID19 contagion, the WHO has issued instructions requesting all individuals to maintain a distance of at least a metre (3 feet) between themselves and the mandatory use of masks. Now, as the country gears up towards life post-lockdown, it is imperative that such social distancing rules are adhered to. Thus, the design of a computer-vision driven technology to help law-makers track social distancing between citizens on the move is one of our major objectives under the RAKSHAK initiative.

Detection of Smell Loss as Diagnostic feature

It is necessary to develop technology solutions for quantitative measurement of the loss of sensation of smell in conjunction with vision based sensing. A measurement system and algorithm development is underway to build such a system. Combination of VOC's generated by heating aroma oils in measured quantities can lead to structured response from a patient. Using anomaly detection techniques we need to identify the degree of deviant response from the patient under consideration.

A.5 Technology Fallouts: Human Health Vertical

[All designed solutions to follow protocols of trustability, data collection to be all-inclusive and adhere to standard ethical and privacy practices.]

1. Techniques for image enhancement and analysis towards enhancement of diagnostic images. [2 years]
2. Synthesis of diagnostic images in one modality from another, using state-of-the-art processing/machine learning techniques of image reconstruction, towards sensor enhancement: [2 years]

   - CT from plain radiography and sonography

   - Multispectral and hyperspectral images (HSI) from RGB images – extension of these studies to develop devices like non-invasive non-contact devices to measure core body temperature, heart rates, respiration rates and SpO_2 at real-time

   - Conversion of the MRI head from a CT head.
3. Machine Learning based dependable diagnostic and prognostic system using X-ray and CT, MRI images for trauma, chronic and infectious diseases [3 years]
4. A portable, low-cost robust, explainable CV-driven fluorescent imaging and diagnostic system on edge, that would be able to answer counterfactual questions, and be robust

to significant data distribution shifts and uncertain and erroneous measurements. The system would also analyze longitudinal multimodal medical data - like CT-Scan, MRI scan, genomic data, Hyperspectral imaging data, and clinical measurements - of the patient for prognosis of oral cancer. [3 years]

5. An integrative Radio-genomic platform for personalized diagnosis, prognosis and treat design. [5 years]

6. Techniques for real-time acquisition of high-resolution HSI video datasets that capture and enhance subtle spectral and spatial dynamics of different tissue types; to aid in intraoperative imaging of organs, tissues, cells, and molecular biomarkers of interest. [3 years]

7. Explainable AI-driven algorithms for real-time processing of HSI data – extraction of high-quality diagnostic information, and generation of quantitative maps of different tissue types as well as disease-specific endogenous parameters. These could be deployed in diagnostic systems for rural or remote health centers. [2 years]

8. Generation of an ever-growing spectra database for significant molecular biomarkers and all tissue-categories including skin and subcutaneous tissue, ocular tissue, head/brain tissue, epithelial/mucous tissue, breast tissue, cartilage, liver, muscle, aorta, lung, myocardium, etc. This would contribute to the design of traceable dependable algorithms that can accurately differentiate between oxygenated and deoxygenated blood, distinguish between all tissue categories, and identification of molecular biomarkers for early cancer detection. [5 years]

9. Collaboration with medical institutes to design automatic and semi-automatic software tools for retinal image analysis to identify retinal biomarkers; design considerations to include reliability and validation of decisions, number and type of conditions, multi-modal analysis (e.g., fundus, optical coherence tomography, scanning laser ophthalmoscopy), and novel imaging technologies.[3 years]

10. Predictive algorithms for automated diagnosis and grading of diabetic retinopathy. [2years].

11. Design and deployment of an efficient and accurate 'Cataract Detection on Mobile Platform', to help screen and quantify wait time and individual risk for patients being considered for cataract surgery [2 years].

12. Development of a risk score for patients at high risk of developing CAD and a strategy for risk stratification and treatment in patients with stable CAD [2 years].

13. Build systems that could be installed at primary health centers to assist doctors in diagnosis and accurate determination of the stage of severity or progression of common diseases. Use of computer vision and machine learning techniques across different modalities of medical data obtained from the patients, like clinical measurements (heart rate, respiratory rate, etc.) and medical imaging data (X-Rays and other available modalities), to accurately diagnose and stage disease progression. [3 years]

14. Development of effective screening tools that aid health workers in the detection of preeclampsia in pregnant women towards timely clinical decisions. [2 years]

15. A computer vision-based system to measure and classify malnourishment using photographs captured through mobile phones. This tool can be integrated into other such mobile systems to provide an overall framework of services related to children. [2 years]

16. Explainable algorithms that detect COVID-19 from frontal chest X-ray images. [1 year]

17. A computer vision and artificial intelligence driven hybrid model, that uses a combination of advanced deep learning methods and belief networks, for prognosis of COVID-19 infection based on the longitudinal studies on multimodal data including X-Ray images, X-Ray reports, CT scans, clinical data such as body temperature, and treatment given to the patients. [18 months]

18. A non-invasive solution for the detection of individuals suffering from silent hypoxia – as COVID-19 suspects, particularly in crowd-settings, from thermal videos. [1 year]

19. An AI-driven mechanism for the real-time detection and identification of individuals in masks. [1 year]

20. Methods to help law-makers track social distancing between citizens on the move. [1 year]

21. Development of a software for prediction of toxicity and response to radiotherapy under the proposed context [3 year]

B. Computer Vision for Biosphere Management

B.1 SUMMARY

Computer vision-based techniques can provide very significant tools for conservation of the environment. In this vertical, in collaboration with academic institutions, research labs and stakeholders we have proposed generation of deployable technologies for wildlife conservation, appropriate exploitation of marine fisheries and plant phenomics for food security.

B.2 Introduction

Biological diversity, most commonly used in its contracted form, biodiversity, is the term used to describe the variety of life on earth, including animals, plants and microbial species. It has been estimated that there are roughly 8.7 million eukaryotic species on earth, of which approximately 25% (2.2. million) are marine, and most of them are yet to be discovered [Mora et al. 2011]. Biodiversity not only refers to the multitude of species on earth, it also consists of the specific genetic variations and traits within species (such as different crop varieties), and the assemblage of these species within ecosystems that characterize agricultural and other landscapes such as forests, wetlands, grasslands, deserts, lakes and rivers. One of the major area of application of Computer Vision is management of ecological resources

B.3 National Need & Impact

The most critical need of the time is to ensure safety of the biosphere and development of sustainable strategies for exploitation of resources of the biosphere for long term well-being of humanity.

The population of many species is fast reducing due to loss of habitat caused by deforestation. Conservation of wildlife is essential for ensuring biodiversity. Key issues we need to address for conserving wildlife are: (i) proper enumeration of species in its habitat (ii) ensuring protection of the species from poachers (iii) reducing wildlife-human conflict so that the local population can champion the conservation effort. Computer Vision can provide the enabling technology for achieving these goals.

Many species of chimpanzees and gorillas have recently been added to the list of endangered and critically endangered animals in the International Union for Conservation of Nature (IUCN) Red List. As deforestation sees no reduction in the near future, there is an enormous risk of more primate species being added to the list of inevitable extinction of its current members. To counter these catastrophic consequences, conservation of primates is the need of the hour - not only for conservation and effective mitigation in the wild but also in zoological parks and wildlife sanctuaries. Poaching is another challenge that is faced by wildlife conversation. In the last few years, tiger poaching in India has increased by 60%. Similar is the case for other animals that are hunted for trade. Wildlife institutes of India and related organizations are exploring the use of cameras and sensors to monitor the movement of animals and poachers in the forests. Some countries use thermal imagery to monitor wildlife.

Along with protection of biodiversity we need to address the problem of sustainable exploitation of natural resources ensuring food security, economic prosperity and conservation

of the ecosystem. India's marine fishery has an annual turnover of about Rs. 52,326 crores in 2018. If the marine biodiversity and mariculture potential along our marine waters are valued, it is expected to be worth several billion. With the ever-increasing demand and trade, there is tremendous pressure on these precious resources and often their sustainability is in question. With difficulty accessing a long coastline of 7516.6 km, 2.17 million km^2 of EEZ, monitoring of fish catch, biodiversity, and mariculture systems in Indian coastal and marine waters are hugely challenging. Multispecies nature of fisheries, different fishing methods, and systems, with difficult to access fish landing centers coupled with long-chain marketing and trading systems, adds up constraints to their monitoring and management. Identification of fishes to the species level is a challenging task, particularly when the resources are multispecies and captured in large quantities. In fisheries management, real-time identification of fish, taking their counts, and estimating their biomass are important management inputs for providing fishing guidelines, regulations, and conservation protocols. There is a need for quick and accurate data collection, monitoring, and informed management of these resources. The current monitoring systems in place, though effective, are resource-intensive, time-consuming, and depend largely on human judgment; they might cause delays and errors in decision making. Use of automated means or enumerator free collection of data on species caught, size at which they are caught and the quantity caught and using the data in an appropriate estimation scheme is a critical issue in marine fisheries science.

The major challenge of the world is to double the food production by 2050 for sustaining food security. This increase in production must be achieved mainly through increase in crop productivity per unit area while sustaining the environment and enhanced input use efficiency. This challenge is further exacerbated by the competition for the land and water resources, global warming and associated increase in the frequency and severity of drought and heat stresses in future. Hence, both genetic improvement approaches involving phenomics and genomics-assisted breeding for improving the innate resource use efficiency, stress tolerance and yield, and precision crop management strategies for improving resource use efficiency and reducing inputs and post-harvest losses are key issues to be taken care of for a sustained enhanced production system.

Both for genomics-aided crop improvement and precision crop management, high throughput precision phenotyping (characterization) of plants is necessary. Conventional phenotyping methods remain a bottleneck that limits utilization of the power of cutting edge science in crop improvement and management. The multi-disciplinary science of phenomics, the sensor aided non-destructive high throughput automated acquisition and analysis of high-dimensional phenotypic data on an organism-wide scale, emerged recently to bridge the phenotype-genotype gap in breeding and high throughput plant health monitoring in precision agriculture. Next generation phenomics, the study of plant growth, performance, and composition, utilizes new technologies to better characterize plant responses to the environment and also better describes the growth environment itself (Furbank and Tester, 2011). Computer vision can play a key role in this space.

B.4 Objectives

On the basis of the national need and requirement, we have identified following objectives for this vertical

1. Development and deployment of Computer Vision techniques for facilitating wildlife conservation

2. Design and deployment of Computer vision based multi-sensory system for marine fish catch estimation, biodiversity, and mari-culture monitoring

3. Design and development of Computer Vision applications for plant phenomics to ensure increased food production.

B.5 Research Challenges & Application Development

The research efforts under this head can be segregated into three major groups:

- Wildlife monitoring and conservation
- Underwater fisheries/Marine biology
- Phenomics and Plant health management

B.5.1 Wildlife Conservation

Due to the illegal trafficking of animals, poaching, loss of natural habitats, conservation and the management of wildlife has become a major challenge throughout the globe. There have been efforts on this aspect by the government. Significant goals of wildlife conservation and management are monitoring the population, monitoring the habitats, monitoring the poaching activities, and monitoring human-wildlife conflicts, etc. Recently, computer vision-based techniques, equipped with machine learning tools, have become the major approach for better conservation and management of wildlife.

A. Wildlife Monitoring

In this direction, the project "Wildbook[16]" is a major step that is trying to achieve the above goals using computer vision techniques. There exist techniques to detect and re-identify wild animals for population monitoring [Korschens et al. 2019, Liu et al. 2019, Kupyn et al. 2019]. There also exist approaches for estimation of 3D models with textures from images that help biologists to monitor the behavior and the health of the animals [Zuffi et al. 2019].

Computer vision-based techniques are helping in the conservation and management of wildlife. However, these approaches are in their initial stages and require improvement. Furthermore, there exist several other problems where computer vision can play a critical role. Wildfires lead to the extinction of many endangered and critically endangered species. The 2019 wildfire in Australia led to the death of more than half of the population of Koalas. In such wildfires, fire rescue teams struggle to locate and save the animals. Another challenge lies in monitoring wildlife.

Current systems of animal tracking are mostly invasive in nature. For example, researchers have used GPS collars [Clark et al. 2006], which have to be strapped around the neck of animals and can be monitored from a remote location. Another approach of tracking, proposed by Kim et al. (2013), employs both RFID and GPS tags embedded in the body of animals to detect if

---

[16] Wildbook: https://www.wildbook.org/doku.php

they have escaped their cages in a zoo. These kinds of arrangements are costly, unreliable and require a significant human intervention into the lifestyles of wild animals.

With these motivations, we would like to collaborate with wildlife institutes in the country and undertake the following projects:

- Detecting wildlife in regions of natural disaster: We will devise a system where our goal will be to automatically detect (using learning based computer vision techniques) animals in the line of disaster and alert rescue teams. This is a challenging problem due to the possible presence of thick smoke, severe heat in large-scale wildfires. Besides wildfires, flooding is another natural disaster that claims the lives of many animals. Like in the flood of 2018 in Assam, many rhinoceroses drowned in the Kaziranga National Park. We would like to propose a system that can be of use in floods as well.

- Animal identification and tracking: To minimize the usage of invasive techniques for animal identification, we aim to design identification and tracking algorithms for different kinds of animals like tigers and chimpanzees.

- Surveillance of wildlife conservation areas: We will design human detection and tracking algorithms in wildlife environments that operate with low cost sturdy cameras and can be installed in forest areas. Besides animals, these tracking algorithms - implemented across camera sensor networks - can also be used to detect poachers.

Animal Phenotyping

Animal phenotyping allows researchers to gain a better understanding of complex interactions between the cardiovascular, respiratory, metabolism and central nervous systems, in sync with the environment. This is particularly critical for the monitoring of the well-being of livestock and animals (especially endangered species) in zoological parks and conservatories. Longitudinal observation and analysis of animal characteristics and behaviour, and environmental parameters – gathered from remote sensing and on-ground cameras, and wearable sensors – is of paramount importance.

Under the biodiversity vertical, iHub Drishti aims to develop explainable deep-learning based tools to derive detailed connections between animal health, behaviour and the environment. Educating animal care-takers and farmers on variations in these parameters would not only help them track the fitness of the animals, but also design better mechanisms of sustenance for them.

Animal Identification

Ecological imbalance owing to rapid urbanization and deforestation has adversely affected the population of several wild animals; rendering a number of species extinct, some endangered and forcing others to co-exist in close proximity to human settlements. Natural calamities and uncontrolled poaching are other factors affecting numbers of the wild. Effective wildlife conservation calls for identification of animals as representative of a species, maintaining accounts of their numbers locally and globally, and monitoring their population across locations. However, existing approaches like field visits for data collection and manual analysis by experts are resource intensive, tedious, prone to errors arising from subjective estimations, and time consuming. Remote sensors and on-ground cameras in natural habitats offer unobtrusive ways of gathering vast amounts of data on animals in the wild, and the deep-learning based image analysis techniques (refer figures below) of today present immense capacities for automated relevant information extraction. However, this also calls for arriving

at solutions to certain challenges arising from pose, lighting variations, occlusions due to uncontrolled environments, and scarce all-inclusive training datasets - particularly in the Indian context.

B. Computer Vision for Reducing Human Wildlife Conflicts

Expansion of urban areas and infrastructural developments has led to depletion of forest cover, which is a natural habitat for many animal species. As this habitat shrinks, humans and animals share spaces, and this coexistence leads to various forms of conflicts between humans and wildlife. On one hand, animals may venture into human settlement areas in search of food and other survival needs, which may result in economic losses and sometimes human fatality. On the other hand, humans are responsible for poaching and killing animals for sale of animal products and occasionally in retaliation to the incurred economic loss.

The importance of biodiversity for a sustainable ecosystem is well understood by conservation scientists and government organizations and is accordingly reflected in the wildlife protection policies, e.g., India has increased the number (area) of conservation reserves from 4 (~42km^2) in 2006 to 88 (~4300km^2) in 2019. While policy-level efforts are made to conserve biodiversity, enforcing them is a demanding task and wildlife crimes are still rampant. Similarly, human-wildlife conflicts adversely impact human life and property as well. Various instances of retaliatory killing of tigers, leopards and elephants are reported due to cattle killings, crop depredation and even fatal attacks on humans.

We propose to use technological interventions for mitigating human-wildlife conflict by leveraging recent advances in cyber-physical systems that are becoming pervasive in our everyday life. We plan to achieve this in a two-staged manner: firstly, develop a camera-phone based crowdsourced conflict reporting system, which is intended to be deployed in urban and semi-urban areas for macaques, stray cattle, dogs, and similar commensal (dependent on humans) species. Secondly, we will develop a camera trap and UAV-based system for anti-depredation surveillance of agricultural fields to mitigate crop-raiding by macaques, stray cattle, Nilgai, elephants and other large herbivores.

Stakeholders

We believe this project will be relevant to multiple union ministries, state forest departments as well as international organizations. An initial list is given below:

- Ministry of Environment, Forest and Climate Change
- Ministry of Agriculture & Farmers' Welfare
- Department of Fisheries
- State forest departments, e.g., the Himachal Pradesh Forest department's wildlife wing has shown strong interest
- International organizations like WWF, IUCN, WCS, etc.

B.5.2 Computer Vision for Underwater Fisheries Applications (Collaborator: ICAR-CMFRI)

Image based analysis of underwater imagery requires implementation of image enhancement and restoration for improving the quality of the image data. There have been many techniques

to restore and enhance the underwater images. Schechner and Averbuch [6] exploited the polarization de-hazing method to compensate for visibility degradation. Ancuti et al. [7] used image fusion method in turbid medium to reconstruct a clear image, and Hou et al. combined point spread function and a modulation transfer function to reduce the blurring effect [8]. Ouyang et al. [9] proposed a bilateral filtering based image deconvolution method. Although the aforementioned approaches can enhance the image contrast, these methods have demonstrated several drawbacks that reduce their practical applicability. First, the equipment of imaging is difficult in practice (e.g., range-gated laser imaging system, which is hardly applied in practice). Second, multiple input images are required, which is difficult to capture by hardware. Third, they cannot solve the color distortion very well. Also, these problems with reference to inland water bodies have not been much investigated.

Computer vision and image processing techniques have been successfully applied to different problems for abundance calculation. An interesting as well as important aquaculture problem is measuring the weight of fish. A comprehensive survey on the subject was carried out by Zion [10]. In India, hardly any work has been done involving computer vision, acoustic imaging and robotic interventions for developing robust technology for fish abundance data collection. Automatic counting or characterization of fish species remains a complex and difficult task because of numerous factors: many species move about during observations, individuals often look very similar, some species blend in with their backgrounds, lighting can be variable, and the correlation between measurable features and desired attributes may be weak. Fish recognition algorithms and school and shoal identification against complex background with noisy video data have received little attention globally. Fish bio-mass and volume estimation methods need enhancement to deal with uncertainty inherent in observations obtained through sensor based sampling. Further, in the Indian context, tank/pond, riverine region/ reservoir/ estuary offers further challenges because of the degree of turbidity of water, pollution, immersed impurities and occlusion. In this context, we need to design improved imaging technology; possibly deriving from fusion of acoustic data and visual data towards enhanced classification techniques.

Application of computer vision-based multi-sensory systems for marine fish catch estimation, biodiversity, and mariculture monitoring

Computer vision and machine learning based techniques have great potential to automate the annotation of marine images to reduce the time consumed in manual processing. Currently, preliminary efforts are on to make a deep learning based framework for automatic annotation of coral images using novel classifying algorithms at ICAR-CMFRI. However, as a next step, the development of a computer vision and deep learning based framework for the automatic annotation and detection of the benthic and fish videos using novel classifying algorithms and development of underwater autonomous drones based tools to determine the diversity and abundance in the coral ecosystems would greatly benefit marine ecological studies in terms of cost, speed, accuracy, and thus in better quantifying the status.

Development of novel and robust technologies for the automation of the identification, abundance, and biomass as well as quality data involving multi signal processing, data fusion, pattern recognition with machine learning, and AI platforms are indispensable components of such ventures. Problems needed to be addressed are:

- Development of image-based fish species identification, size, and biomass estimation from single and multi-species catches and data transmission.

- Develop real-time target species identification and advisory systems integrated with Potential Fishing Zone advisories, weather, and territorial water indicators.
- Identification, counting, and assessment of fish species assemblage underwater using imaging techniques.
- Development of mariculture site selection, structural and biological stock monitoring, water quality sensing using onsite sensors and aerial drone-based systems.
- Development of smart underwater autonomous systems integrated with image classifiers to estimate coral reef diversity and abundance from videos.

ICAR-CMFRI will provide the following support:

- 11 research centres, 17 field centres and 2 KVK's in all maritime states and Lakshadweep.
- Survey mechanism covering complete Indian coast to collect landing centre based fishery information based on a stratified multistage random sampling method and a massive manpower strength of survey staff and enumerators for each maritime district.
- Expertise in traditional and molecular taxonomic identification of all marine species.
- GIS based mapping of trawl fishing grounds across the coastline.
- Huge collection of digital images pertaining to all exploited resources in Indian EEZ.
- Professional Association of Diving Instructors (PADI) certified scientific and technical manpower.
- Experience in handling deep learning based frameworks and preliminary efforts on automatic annotation of coral images using novel classifying algorithms.
- With necessary facilities provided, we have capacity to generate both large number of images or videos from fish landing and underwater
- Conventional identification of fish and other aquatic biodiversity
- Real Mariculture and offshore and near shore cage culture installations
- Conventional habitat, fishery, mariculture monitoring
- PFZ advisories available
- Vessel based near shore and offshore activities

B.5.3 Computer Vision for Plant Phenomics (*Collaborator:* ICAR-IARI)

ICAR-IARI, New Delhi has established an automated plant phenomics facility. The facility was inaugurated by the Hon'ble Prime Minister, Shri. Narendra Modi on 11-10-2017, and dedicated to the Nation in the name of "Nanaji Deshmukh Plant Phenomics Centre (NDPPC)". The state-of the art Phenomics facility consists of Climate controlled greenhouses, moving field conveyor systems, 1200 plant carriers with RFID tag, weighing and watering stations and imaging sensors viz., RGB, IR Thermal, NIR-shoot, NIR-root, Chlorophyll Fluorescence, VNIR-Hyperspectral and SWIR-Hyperspectral. During the past two years, the facility has recorded over 36 million images from rice and wheat subjected to drought and nitrogen deficit stresses. The centre has also

identified superior germplasm and advanced breeding lines with high drought tolerance and water use efficiency, and genomics regions associated with drought tolerance and water use efficiency in rice and wheat. Drone phenotyping for high throughput field phenomics has been standardised. Computer vision based phenotyping for nowcasting and forecasting of plant phenotype can significantly enhance the output of both controlled environment phenomics as well as field phenomics (Mochida et al. 2019).

Research Gaps and Scientific Questions

Number of precision agricultural tools and mobile and web based information dissemination and forecasting systems are placed in support of agricultural stakeholders. But still there are many grey areas which need to be addressed. Both public and private agriculture R&D organisations take at least 10 long years for breeding a new variety of crop plants using conventional approach. In traditional approach, phenotyping of crop plants for identification of contrast donors from the diverse population is time consuming and cumbersome exercise. At times, it is rather challenging in plants to detect the cause of the problem, as usually symptoms of different problems are common i.e., yellowing and wilting of plants. The artificial imposition of stress treatments equally to all the biological samples is difficult. Data acquisition of stress levels as well as stress response in each plant of the population at a given target environment is a manually impossible task. One among the best examples is biotic and abiotic stress levels and response of the plants. Hence development of diagnosis tools specific for different stress conditions like nutrient deficiency, biotic and abiotic stress will help both researchers and farmers in arriving at a solution to real time problems. Diagnostic deficiency is one among the grey areas where little is known from the publication. Several researchable issues will find a place in the list of grey areas and need attention. As climate control phenotyping and field phenotyping using drones along with computer vision tools would not only hasten the process of breeding new variety. But also will reduce the product life cycle of the released variety in the market to achieve sustainable agricultural productivity in India.

As precision phenotyping & precision agriculture involves measurement of morpho-physiological traits of crop plants and physical properties of soil. Automatic estimation of traits like leaf area, leaf number, biomass, tiller number, economic tiller number, panicle number, grain number, grain size, grain weight, quality aspects of grain are the few areas of researchable issue to be addressed. Apart from this "viewing the invisible" is one of the prime areas, where prediction algorithms & simulation models are used for timely predicting the flowering dates, yield, occurrence of nutrient deficiency, pest and disease need to be addressed. Acquiring data below ground about soil moisture, mineral nutrient availability, root architecture, microbiome etc., are areas of "black box", need attention to study using different computer vision techniques to address the problem. Some of the other gray areas to work on:

1. Discerning different kinds of stresses and stress levels in field crops.

2. Accurate and efficient measurement of leaf angle distribution (LAD) is important for characterizing canopy structures and understanding solar radiation regimes within the plant canopy. The main challenge for obtaining LAD is measuring the orientations of individual leaves rapidly and accurately in complex field conditions.

3. Under field conditions, counting the economic part of the crops is the key to estimate the yield. kernel /Spikelet counting on maize/wheat ear using RGB images.

4. Counting of tillers in cereals crops using image-based phenotyping

5. Root Phenotyping still required so questions to be answered.

6. Biomass estimation after reproductive phase in food crops

7. Photosynthetic traits estimation using computer vision (Non photochemical quenching)

The major aim of the proposal is to identify and quantify plant traits and donors for tolerance to water and nutrient deficit stresses under controlled environment and field conditions using different imaging sensors and computer Vision. To address this aim, the following problems will be addressed:

Objectives

1. To identify and quantify component traits contributing to yield (Plant architecture related traits - measurement of leaf area, counting of leaves, branches/tillers, automated detection of phenology, detection of flowering and reproductive structures, for all crops; detection of stage of physiological maturity and post-harvest quality for Tomato, etc) by using computer vision

2. To diagnose and quantify drought stress response symptoms under controlled environment conditions using high throughput imaging sensors

3. To diagnose macronutrient (N) and micronutrient (Zn and Fe) deficiency and quantify these nutrients in the tissues of selected crops when they occur individually or in combination using high throughput imaging sensors

4. To diagnose symptoms of combination of drought/salt/water logging stresses with nutrient deficiency and quantify plant responses using computer vision based image analysis.

Deliverables

- High throughput imaging sensor based methods, algorithms and softwares for identification and quantification of plant architectural traits, diagnosis and quantification of abiotic stresses, nutrient deficiency.

- SOPs for high throughput phenotyping for precision breeding and precision agriculture

Outcomes

- The imaging sensor based methods, algorithms and softwares identification and quantification of plant architectural traits, diagnosis and quantification of abiotic stresses will be useful for high throughput germplasm evaluation in crop breeding programs, management of harvest and post-harvest quality in tomato, and precision agriculture.

- The specific wavelengths identified, algorithms and softwares developed will be useful to design handheld high throughput and low-cost sensors useful for farmers to manage their crops.

Benefits to stakeholders

- The imaging sensor based methods, algorithms and softwares developed will be useful for plant breeders and physiologists to phenotype large numbers of germplasm reserves available in our country to develop high yielding and climate resilient crop varieties.

- Industry and start-ups involved in precision agriculture will benefit from the methods, algorithms and softwares developed from this study to develop and market handheld high throughput and low-cost sensors, and use these tools to provide custom hiring services.
- The imaging sensor based methods, algorithms and softwares developed will be useful for precision agriculture of these crops.

B.5.4 Water Management

Use of hyperspectral non-contact online sensors along with traditional water quality sensing devices can provide new application areas of computer vision. Agriculture sector consumes the largest share of India's total water resources and is increasingly subjected to water risks. Around 70%-80% of the water is being used by agriculture. Further, different crops in different regions of the country require different quality of water. Improving agriculture's water management and water treatment are, therefore, essential for a sustainable and productive agro-food sector. Use of field phenomics for assessment of water requirement for crops can lead to water conservation. Exploiting novel use of non-contact multi-spectral sensors, we propose to develop technologies for: i) real time water quality monitoring (ii) water management and drinking water. Further to make the water utilization in agriculture efficient an informed decision on irrigation is required, therefore in addition to water management, following research activities would be conducted for use of hyperspectral imaging data for monitoring: i) soil condition; ii) soil remediation; iii) subsu and iv) geospatial soil imagery.

B.6 Technology Fallouts: Biosphere

[All designed solutions to follow protocols of trustability, data collection to be all-inclusive and adhere to standard ethical and privacy practices.]

- Detection of wildlife along front lines disasters (like wildfires and floods) [2 years]
- Non-invasive, low-cost techniques for fine-grained identification and monitoring of animals in the wild.{2 years]
- Detection of poachers in conservation areas [2 years]
- Explainable deep-learning based tools to derive detailed connections between animal health, behaviour and the environment, to track fitness of livestock and wild animals help design better mechanisms of sustenance for them. [3 years]
- A camera-phone based crowdsourced conflict reporting system, and a camera trap and UAV-based system for anti-depredation surveillance of agricultural fields to mitigate crop-raiding by macaques, stray cattle, Nilgai, elephants and other large herbivores. [2 years]
- An Artificial Intelligence (AI) powered Human-Wildlife Conflict Management System, that will engage regular citizens through a mobile phone app and assist forest authorities to prioritize actions for optimal results. [2 years]
- A web-portal for forest and municipal authorities, to analyze human-wildlife conflict statistics through a GIS-style spatial analysis of conflict-regions. [1 year]
- Image-based fish species identification, size, and biomass estimation from single and multi-species catches and data transmission. [2 years]
- Real-time target species identification and advisory systems integrated with Potential Fishing Zone advisories, weather, and territorial water indicators.

- Identification, counting, and assessment of fish species assemblage underwater using imaging techniques. [4]
- Development of mariculture site selection, structural and biological stock monitoring, water quality sensing using onsite sensors and aerial drone-based systems. [4]
- Development of smart underwater autonomous vehicle systems integrated with image classifiers to estimate coral reef diversity and abundance from videos. [5].
- The imaging sensor based methods, algorithms and software for identification and quantification of plant architectural traits, diagnosis and quantification of abiotic stresses. It will be used for high throughput germplasm evaluation in crop breeding programs, management of harvest and post-harvest quality in crops like tomato, and precision agriculture. [3 years].
- The specific wavelengths identified, algorithms and softwares developed will be used to develop handheld high throughput and low-cost sensors useful for farmers to manage their crops. [3 years]
- Use of non-contact multi-spectral sensors to develop technologies for: i) water management and drinking water, ii) smart water grid, iii) real-time water monitoring, and iv) wastewater treatment and reuse, iv) soil testing and monitoring, v) soil remediation, vi) subsurface irrigation.   [4 years]

References

Healthcare

- Aeschbacher, J., Wu, J., & Timofte, R. (2017). In defense of shallow learned spectral reconstruction from RGB images. In Proc. IEEE International Conference on Computer Vision Workshops (pp. 471-479).
- Akhtar, N., & Mian, A. (2017). Non-parametric coupled Bayesian dictionary and classifier learning for hyperspectral classification. In IEEE Transactions on Neural Networks and Learning Systems (vol. 29, no. 09, pp. 4038 - 4050).
- Akhtar, N., & Mian, A. (2018). Hyperspectral recovery from RGB images using Gaussian processes. In ArXiv: 1801.04654v2.
- Arad, B., & Ben-Shahar, O. (2016). Sparse recovery of hyperspectral signal from natural RGB images. In European Conference on Computer Vision (ECCV) (pp. 19-34).
- Ahn, C. Y., Jung, Y. M., Kwon, O. I., & Seo, J. K. (2012). Fast segmentation of ultrasound images using robust rayleigh distribution decomposition. In Pattern Recognition (vol. 45, no. 9, pp. 3490–3500).
- Ausiello, D., & Shaw, S. (2014). Quantitative human phenotyping: the next frontier in medicine. In Transactions of the American Clinical and Climatological Association (vol. 125, pp. 219–228).
- Brusseau, E., de Korte, C. L., Mastik, F., Schaar, J., & Van der Steen, A. F., (2004). Fully automatic luminal contour segmentation in intracoronary ultrasound imaging-a statistical approach. In IEEE Trans. Biomed. Imag. (vol. 23, no. 5, pp. 554–566).
- Budzan, S., & Wyżgolik, R. (2013). Face and eyes localization algorithm in thermal images for temperature measurement of the inner canthus of the eyes. In Infrared Physics & Technology (vol. 60, pp. 225-234).

- Choi, I., Jeon, D., Nam, G., Gutiérrez, D., & Kim, M. (2017). High-quality hyperspectral reconstruction using a spectral prior. In ACM Transactions on Graphics (vol. 36, pp.1-13).
- Cybulski J. S., Clements, J., & Prakash, M. (2014). Foldscope: origami-based paper microscope. In PLOS ONE (vol. 9, no. 6, E98781).
- Day, G. L., & Blot, W. J. (1992). Second primary tumors in patients with oral cancer. In Cancer (vol. 70, no. 1, pp. 14–19).
- De Veld, D. C., Witjes, M. J, Sterenborg, H. J., & Roodenburg, J. L. (2005). The status of in vivo autofluorescence spectroscopy and imaging for oral oncology. In Oral Oncol. (vol. 41, pp. 117–131).
- Dias, J. B., Plaza, A., Valls, G. C., Scheunders, P., Nasrabadi, N., & Chanussot, J. (2013). Hyperspectral remote sensing data analysis and future challenges. In IEEE Geosci. Remote Sens. Mag (vol. 1, no. 2, pp. 6–36).
- Dikshit, R., Gupta, P. C., Ramasundarahettige, C., Gajalakshmi, V., Aleksandrowicz, L., Badwe, R., Kumar, R., Roy, S., Suraweera, W., Bray, F., Mallath, M., Singh, P. K., Sinha, D. N., Shet, A. S., Gelband, H., & Jha, P. (2012). Cancer mortality in India: A nationally representative survey. In Lancet (vol. 379, no. 9828, pp.1807–1816).
- Dwyer, P. J., DiMarzio, C. A., & Rajadhyaksha, M. (2007). Confocal θ line-scanning microscope for imaging human tissues. In Appl Optics. (vol. 46, no. 10, pp. 1843–1851).
- Esneault, S., Lafon, C., & Dillenseger, J. L. (2010). Liver vessels segmentation using a hybrid geometrical moments/graph cuts method. In IEEE Trans. Biomed. Eng. (vol. 57, no. 2, pp. 276–283).
- Farah, C. S., & McCullough, M. J. (2007). A pilot case control study on the efficacy of acetic acid wash and chemiluminescent illumination (ViziLite™) in the visualisation of oral mucosal white lesions. In Oral Oncol. (vol. 43, pp. 820–824).
- Fawzi, A. A., Lee., N., Acton, J. H. Laine, A. F., & Smith, R. T. (2011). Recovery of macular pigment spectrum in vivo using hyperspectral image analysis. In J. Biomed. Opt. (vol. 16, no.10, p.106008).
- Galliani, S., Lanaras, C., Marmanis, D., Baltsavias, E., & Schindler, K. (2017). Learned spectral super-resolution, ArXiv: 1703.09470v1.
- Gendrin C., Roggo Y., & Collet C. (2008). Pharmaceutical applications of vibrational chemical imaging and chemometrics: A review. In J. Pharm. Biomed. Anal. (vol. 48, no. 3, pp. 533–553).
- Grigoroiu, A., Yoon, J. & Bohndiek, S. E. (2020). Deep learning applied to hyperspectral endoscopy for online spectral classification. In Sci Rep (vol. 10, p. 3947).
- Guazzi, A. R., Villaroel, M., Jorge., J., & Daly, J. (2015). Non-contact measurement of oxygen saturation with an RGB camera. In Biomedical Optics Express (vol. 6, no. 9, pp. 3320-3338).
- Ibraheem, I. (2015). Early detection of melanoma using multispectral imaging and artificial intelligence techniques. In American Journal of Biomedical and Life Sciences (Special Issue: Spectral Imaging for Medical Diagnosis "Modern Tool for Molecular Imaging", vol. 2-3, pp. 29-33).
- Jeyaraj, P. R., & Nadar, E. R. S. (2019). Computer-assisted medical image classification for early diagnosis of oral cancer employing deep learning algorithm. In Journal of cancer research and clinical oncology (vol. 145, no. 4, pp.829-837).
- Kim, D. W., Lee, S., Kwon, S., Nam, W., Cha, I. H., & Kim, H. J., (2019). Deep learning-based survival prediction of oral cancer patients. In Scientific reports (vol. 9, no. 1, pp. 1-10).

- Kim, A., &Wilson, B. (2011). Measurement of ex vivo and in vivo tissue optical properties: methods and theories. In Welch, A. J., Gemert M. J. C. (Eds.). Optical-Thermal Response of Laser-Irradiated Tissue (pp. 267–319). Springer. Netherlands.
- Kim, Y., Park Y., Kim, J., & Lee, E. C. (2018). Remote Heart Rate Monitoring Method Using Infrared Thermal Camera, In International Journal of Engineering Research and Technology (vol. 11, No. 3, pp. 493-500).
- Lane, P. M, Gilhuly, T., Whitehead, P. D., Zeng, H., Poh, C., Ng, S., Willams, M., Zhang, L., Rosin, M., & MacAulay, C. E. (2006). Simple device for the direct visualization of oral-cavity tissue fluorescence. In J. Biomed Opt. (vol. 11, no. 2, p. 024006).
- Li, Y., Xie, W., & Li, H. Hyperspectral image reconstruction by deep convolutional neural network for classification. In Pattern Recognition (371-383).
- Lu, G., Halig, L., Wang, D., Chen, Z. G., & Fei, B. (2014). Hyperspectral imaging for cancer surgical margin delineation: Registration of hyperspectral and histological images. In Proc. SPIE Int. Soc. Opt. Eng.
- Lu, G., & Fei, B. (2014). Medical hyperspectral imaging: A review. In Journal of biomedical optics (vol. 19, no. 1, p. 10901).
- Lu, G., Wang, D., Qin, X., Muller, S., Little, J. V., Wang, X., Chen A. Y., Chen, G., & Fei, B. (2019) Histopathology Feature Mining and Association with Hyperspectral Imaging for the Detection of Squamous Neoplasia. In Sci Rep (vol. 9, p. 17863).
- Maitland, K. C., Gillenwater, A. M., Williams, M. D., El-Naggar, A. K., Descour, M. R., & Richards-Kortum, R. R. (2008) In vivo imaging of oral neoplasia using a miniaturized fiber optic confocal reflectance microscope. In Oral Oncol. (vol. 44, no. 11, pp. 1059–1066).
- Mishra, D., Priyadarshini, N., Chakraborty, S., & Sarkar, M. (2017). Blood oxygen saturation measurement using polarization-dependent optical sectioning. In IEEE Sensors (vol. 17, no. 12, pp. 3900-3908).
- Mishra, D., Chaudhury, S., Sarkar, M. & Soin, A.S. (2018). Ultrasound image enhancement using structure oriented adversarial network. In IEEE Signal Processing Letters (vol. 25, no. 9, pp.1349-1353).
- Mora C, Tittensor DP, Adl S, Simpson AGB, Worm B (2011) How Many Species Are There on Earth and in the Ocean?. PLOS Biology 9(8): e1001127.
- Negishi, T., Abe, S., Matsui, T., Liu, H., Kurosawa, M., Kirimoto, T., & Sun, G. (2020). Contactless vital signs measurement system using RGB-thermal image sensors and its clinical screening test on patients with seasonal influenza. In Sensors (vol. 20, no. 8, p. E2171).
- Norman, S. M., Odibo, A. O., Macones, G. A., Dicke, J. M., Crane, J. P., & Cahill, A. G. (2010). Ultrasound-detected subchorionic hemorrhage and the obstetric implications. In Obstetrics & Gynecology (vol. 116, no. 2, Part 1, pp. 311–315).
- Nouvong, A., Hoogwerf, B., Mohler, E., Davis, B., Tajaddini, A., & Medenilla, E. (2009). Evaluation of diabetic foot ulcer healing with hyperspectral imaging of oxyhemoglobin and deoxyhemoglobin. In Diabetes Care (vol. 32, no. 11, pp. 2056–2061).
- Ottestad, W., Seim, M., & Mæhlen, J. O., (2020). COVID-19 with silent hypoxemia [in Norwegian: Covid-19 med stille hypoksemi]. In Tidsskr Nor Laegeforen. doi:10.4045/tidsskr.20.0299.
- Ozgulbas, D (2013). NIR light transmission through skin and muscle. MS. Thesis. Biomedical Engineering. New Jersey Institute of Technology.
- Ortiz, S. H. Z., Chiu, T., & Fox, M. D., Ultrasound image enhancement: A review. In Biomed. Signal Process. Control (vol. 7, no. 5, pp. 419–428).

- Patterson, M., Wilson, B., & Wyman, D. (1991). The propagation of optical radiation in tissue. II: optical properties of tissues and resulting fluence distributions. In Lasers Med. Sci. (vol. 6, no. 4, pp. 379–390).
- Panayides A. S. et al., Radiogenomics for Precision Medicine With A Big Data Analytics Perspective DOI 10.1109/JBHI.2018.2879381, IEEE Journal of Biomedical and Health Informatics, 2020.
- Roblyer, D., Richards-Kortum, R., Sokolov, K., El-Naggar, A. K., Williams, M. D., Kurachi, C., & Gillenwater, A. M. (2008). Multispectral optical imaging device for in vivo detection of oral neoplasia. In J Biomed Opt. (vol. 13, no. 2, p. 024019).
- Rutman, A. M., & Kuo, M. D., (2009) Radiogenomics: creating a link between molecular diagnostics and diagnostic imaging. In European journal of radiology (vol. 70, no. 2, pp. 232-241).
- Salzar, R., Steiner, G., Mantsch, H. H., Mansfield, J., & Lewis, E. N. (2000). Infrared and Raman imaging of biological and biomimetic samples. In Fresenius J. Anal. Chem. (vol. 366, nos. 6–7, pp. 712–726).
- Shi, Z., Chen, C., Xiong, Z., Liu, D. & Wu, F. (2018). HSCNN+: Advanced CNN-based hyperspectral recovery from RGB Images. In Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) (pp. 1052-10528).
- Shin, D., Vigneswaran, N., Gillenwater, A., & Richards-Kortum, R. (2010). Advances in fluorescence imaging techniques to detect oral cancer and its precursors. In Future oncology (vol. 6, no. 7, pp. 1143–1154).
- Shin, H. J., Pierce, M. C., Lee, D., Ra, H., Solgaard, O., & Richards-Kortum, R. (2007). Fiber-optic confocal microscope using a MEMS scanner and miniature objective lens. In Opt. Express (vol. 15, no. 15, pp. 9113–9122).
- Shuaib, A. S., Li, X., & Yao, G. (2011). Transmission of polarized light in skeletal muscle. In Journal of Biomedical Optics (vol. 16, no. 2, e025001).
- Skandarajah, A., Sunny, S. P., Gurpur, P., Reber, C. D., D'Ambrosio, M. V., Raghavan, N., James, B. L., Ramanjinappa, R. D., Suresh, A., Kandasarma, U., Birur, P., Kumar, V. V., Galmeanu, H. C., Itu, A. M., Modiga-Arsu, M., Rausch, S., Sramek, M., Kollegal, M., Paladini, G., Kuriakose, M., & Fletcher, D. (2017). Mobile microscopy as a screening tool for oral cancer in India: A pilot study. In PloS one (vol. 12, no. 11, e0188440).
- Sorg, B. S., Moeller B. J., Donovan, O., Cao, Y., & Dewhirst, M. W. (2005), Hyperspectral imaging of hemoglobin saturation in tumor microvasculature and tumor hypoxia development. In J. Biomed. Optics (vol. 10, no. 4, p. 44004).
- Tuan, V. -D. (2004). A hyperspectral imaging system for in vivo optical diagnostics. In IEEE Eng. Med. Biol. Mag. (vol. 23, no. 5, pp. 40–49).
- Wang, L., Sun, C., Fu, Y., Kim, M. H., & Huang, H. (2019). Hyperspectral image reconstruction using a deep spatial-spectral prior. In Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (pp.8032-8041).
- Wu, B., Sun, X., Hu, L. & Wang, Y. (2019). Learning with unsure data for medical image diagnosis. In Proceedings of the IEEE International Conference on Computer Vision (pp. 10590-10599).
- Zang, X., Bascom, R., Gilbert, C., Toth, J., & Higgins, W. (2016). Methods for 2-d and 3-d endobronchial ultrasound image segmentation. In IEEE Trans. Biomed. Eng. (vol. 63, no. 7, pp. 1426–1439).
- Zhao, Y., Rada, L., Chen, K., Harding, S. P., & Zheng, Y. (2015). Automated vessel segmentation using infinite perimeter active contour model with hybrid region

information with application to retinal images. In IEEE Trans. Med. Imag. (vol. 34, no. 9, pp. 1797–1807).

- Zhou, M., Leung, A., Echegaray, S., Gentles, A., Shrager, J. B., Jensen, K. C., Berry,, G. J., Plevritis, S. K., Rubin, D. L.,Napel, S., & Gevaert, O. (2018). Non-Small Cell Lung Cancer Radiogenomics Map Identifies Relationships between Molecular and Imaging Phenotypes with Prognostic Implications. In Radiology (vol. 286,no. 1, pp. 307-315).
- Zuzak, K. J., Naik, S. C., Alexandrakis, G., Hawkins, D., Behbani, K., & Livingston, E. (2007). Characterization of a near-infrared laparoscopic hyperspectral imaging system for minimally invasive surgery. In Anal. Chem. (vol. 79, no. 12, pp. 4709–4715).

Biosphere

- Araus, J. L., & Cairns, J. E. (2014). Field high-throughput phenotyping: the new crop breeding frontier. Trends in plant science, 19(1), 52-61.
- Campbell, M. T., Knecht, A. C., Berger, B., Brien, C. J., Wang, D., and Walia, H. (2015). Integrating image-based phenomics and association analysis to dissect the genetic architecture of temporal salinity responses in rice. Plant Physiol. 168, 1476–1489. doi: 10.1104/pp.15.00450
- Chen, D., Neumann, K., Friedel, S., Kilian, B., Chen, M., Altmann, T., et al. (2014). Dissecting the phenotypic components of crop plant growth and drought responses based on high-throughput image analysis. Plant Cell 26, 4636–4655. doi: 10.1105/tpc.114.129601
- Corp, L. A., McMurtrey, J. E., Middleton, E. M., Mulchi, C. L., Chappelle, E. W., & Daughtry, C. S. (2003). Fluorescence sensing systems: In vivo detection of biophysical variations in field corn due to nitrogen supply. Remote Sensing of Environment, 86(4), 470-479.
- Das, Bappa, Sahoo, R.N.Pargal, Sourabh, Krishna Gopal, Gupta, V K, Verma, R and Viswanathan, C. 2016 Measuring leaf area index from colour digital image of wheat crop, Journal of Agrometeorology 18 (1): 22-26.
- Das B., Sahoo R.N., Pargal S., Krishna G., Verma R., Chinnusamy V., Sehgal V.K., Gupta V.K. (2017). Comparison of different uni- and multi-variate techniques for monitoring leaf water status as an indicator of water-deficit stress in wheat through spectroscopy, Biosystems Engineering. 160:69-83.
- Das B, Sahoo, R N, Pargal S., Krishna Gopal, Verma, Rakesh, Chinnusamy Viswanathan., Sehgal, V.K., Gupta, V.K., Dash, S K, Swain, P 2018, Quantitative monitoring of sucrose, reducing sugar and total sugar dynamics for phenotyping of water-deficit stress tolerance in rice through spectroscopy and chemometrics Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy 192: 41-51.
- Das B, Sahoo RN, Pargal S, Krishna G, Verma R, Chinnusamy V, Sehgal VK, Gupta VK.2019 Comparative analysis of index and chemometric techniques-based assessment of leaf area index (LAI) in wheat through field spectroradiometer, Landsat-8, Sentinel-2 and Hyperion bands. Geocarto International doi.org/10.1080/10106049.2019.1581271
- Das B, Sahoo RN, Biswas A, Pargal S, Krishna G, Verma R, Chinnusamy V, Sehgal VK, Gupta VK. 2020 Discrimination of rice genotypes using field spectroradiometry. Geocarto International 35(1):64-77
- Fahlgren, N., Feldman, M., Gehan, M. A., Wilson, M. S., Shyu, C., Bryant, D. W., et al. (2015). A versatile phenotyping system and analytics platform reveals diverse temporal

responses to water availability in Seteria. Mol. Plant 8, 1520–1535. doi: 10.1016/j.molp.2015.06.005

- Fahlgren, N., Gehan, M. A., and Baxter, I. (2015b). Lights, camera, action: high throughput plant phenotyping is ready for a close-up. Curr. Opin. Plant Biol. 24, 93–99. doi: 10.1016/j.pbi.2015.02.006

- Furbank, R. T., & Tester, M. (2011). Phenomics–technologies to relieve the phenotyping bottleneck. Trends in plant science, 16(12), 635-644.

- Ge Y., Atef A, Zhang H, Miao C, Ramamurthy R K, Sigmon B, Yang J and Schnable J C. (2019). High-throughput analysis of leaf physiological and chemical traits with VIS–NIR–SWIR spectroscopy: a case study with a maize diversity panel. Plant Methods 15:66 https://doi.org/10.1186/s13007-019-0450-8

- Ge, Y., Bai, G., Stoerger, V., and Schnable, J. C. (2016). Temporal dynamics of maize plant growth, water use, and leaf water content using automated high throughput RGB and hyperspectral imaging. Comput. Electron. Agric. 127, 625–632. doi: 10.1016/j.compag.2016.07.028Fahmi et al.,2018

- Golzarian, M. R., Frick, R. A., Rajendran, K., Berger, B., Roy, S., Tester, M., et al. (2011). Accurate inference of shoot biomass from high-throughput images of cereal plants. Plant Methods 7:2. doi: 10.1186/1746-4811-7-2

- Hartmann, A., Czauderna, T., Hoffmann, R., Stein, N. and Schreiber, F. (2011). HTPheno: an image analysis pipeline for high-throughput plant phenotyping. BMC bioinformatics, 12(148).

- Knecht, A.C., Campbell, M.T., Caprez, A., Swanson, D.R. and Walia, H. (2016). Image Harvest: an open-source platform for high-throughput plant image processing and analysis. Journal of experimental botany, 67(11), 3587-3599..

- Krajewski P, Chen D, Ćwiek H, van Dijk AD, Fiorani F, Kersey P, Klukas C, Lange M, Markiewicz A, Nap JP, van Oeveren J, Pommier C, Scholz U, van Schriek M, Usadel B, Weise S. 2015. Towards recommendations for metadata and data handling in plant phenotyping. J Exp Bot. 66: 5417-27.

- Krishna G, Sahoo RN, Singh P, Patra H, Bajpai V, Das B, Kumar S, Dhandapani R, Vishwakarma C, Pal M, Chinnusamy V. 2019 Application of thermal imaging and hyperspectral remote sensing for crop water deficit stress monitoring Geocarto International doi.org/10.1080/10106049.2019.1618922.

- Krishna G, Sahoo RN, Singh P, Bajpai V, Patra H, Kumar S, Dandapani R, Gupta VK, Viswanathan C, Ahmad T, Sahoo PM. 2019 Comparison of various modelling approaches for water deficit stress monitoring in rice crop through hyperspectral remote sensing Agricultural Water Management 213:231–244.

- Li, L., Zhang, Q., and Huang, D. (2014). A review of imaging techniques for plant phenotyping. Sensors 14, 20078–20111. doi: 10.3390/s141120078

- Liu W, Gowda M, Reif JC, Hahn V, Ruckelshausen A, Weissmann EA, Maurer HP, Würschum T. Genetic dynamics underlying phenotypic development of biomass yield in triticale. BMC Genomics 15:458. doi: 10.1186/1471-2164-15-458.

- Minervini, M., Abdelsamea, M.M. and Tsaftaris, S.A. (2014). Image-based plant phenotyping with incremental learning and active contours. Ecological Informatics, 23, 35-48.

- Mochida K., Koda S., Inoue K., Hirayama T., Tanaka S., Nishii R., Melgani F. Computer vision-based phenotyping for improvement of plant productivity: a machine learning perspective. Gigascience. 2019:8.

- Navarro PJ, Pérez F, Weiss J, Egea-Cortines M. 2016. Machine learning and computer vision system for phenotype data acquisition and analysis in plants. Sensors (Basel).16(5). pii: E641. doi: 10.3390/s16050641.
- Neilson, E. H., Edwards, A. M., Blomstedt, C. K., Berger, B., Lindberg Møller, B., and Gleadow, R. M. (2015). Utilization of a high-throughput shoot imaging system to examine the dynamic phenotypic responses of a C4 cereal crop plant to nitrogen and water deficiency over time. J. Exp. Bot. 66, 1817–1832. doi: 10.1093/jxb/eru526
- Paine CE, Marthews TR, Vogt DR, Purves D, Rees M, Hector A, Turnbull LA. 2012. How to fit nonlinear plant growth models and calculate growth rates: an update for ecologists. Methods Ecol Evol. 3: 245-256.
- Pound MP, French AP, Fozard JA, Murchie EH, Pridmore TP. 2016. A patch-based approach to 3D plant shoot phenotyping. Machine Vision and Applications 1-13. doi.10.1007/s00138-016-0756-8
- Ranjan R, Chopra UK, Sahoo RN, Singh AK, Pradhan S. 2012. Assessment of plant nitrogen stress through hyperspectral indices. International Journal of Remote Sensing 22: 6342-6360.
- Rodriguez, D., Sadras, V. O., Christensen, L. K., & Belford, R. (2005). Spatial assessment of the physiological status of wheat crops as affected by water and nitrogen supply using infrared thermal imagery. Australian Journal of Agricultural Research, 56(9), 983-993.
- Sahoo Rabi N , Viswanathan C., Krishna Gopal , Das Bappa, Goel Swati , Dhandapani Raju, Kumar Sudhir, Viswakarma Chandrapal , Swain P and Dash S K 2019, Next generation phenotyping for developing climate resilient rice varieties, Oryza 56 :92-105.
- Sahoo, Rabi N., S. S. Ray, and K. R. Manjunath. (2015). Hyperspectral remote sensing of agriculture. Current Science 108:848-859.
- Singh A, Ganapathysubramanian B, Singh AK, Sarkar S. 2016. Machine learning for high-throughput stress phenotyping in plants. Trends Plant Sci. 21: 110-124.
- White, J. W., Andrade-Sanchez, P., Gore, M. A., Bronson, K. F., Coffelt, T. A., Conley, M. M., & Jenks, M. A. (2012). Field-based phenomics for plant genetics research. Field Crops Research, 133, 101-112.

Imaging for Document Analysis

Collaborating Organizations:

IIT Jodhpur

IIIT Hyderabad

IIIT Bangalore

ISI Kolkata

CDAC Noida

User Agency: National Digital Library of India

Industry Partners: VideoKen, KritikalVision.AI
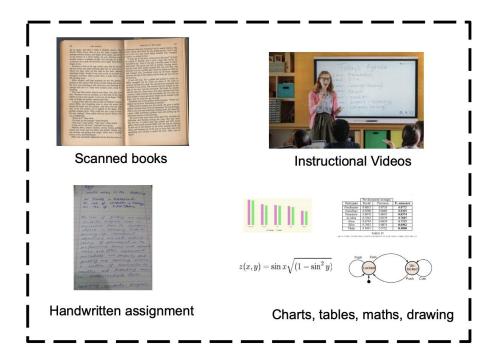
1. Introduction

Figure 1: Optical character recognition (OCR) has manifested itself into various challenging avatars beyond just reading cleanly scanned books, for example, interpreting instructional videos, assessing handwritten assignments, reading not just text but charts, tables, mathematical expression, and drawing.

Large document collections are becoming omnipresent. A number of them originate as images especially when they were captured with cameras in a physical application setting. Many of these are also stored as images since the image format preserves a number of visual attributes that text documents do not carry.

Digital library of books and manuscripts was one of the first to emerge in this space. They contain a large number of scanned/digitized books and manuscripts. One of the first challenges was to recognize the textual content in these collections. OCRs and recognizers emerged as a reaction to this, and they were later adapted to work in a wilder imaging setting. They were hard to develop since the collections had diverse types (eg. printed or handwritten) and document images in diverse languages (English, Hindi, etc.).

With digital document workflow systems becoming more and more popular, a number of use cases in business (banking, finance), administration (offices), health care (records), education (information sources) are demanding document analytics (or in general Visual Text Analytics) algorithms for better understanding and meeting user needs. Pure text recognition is no longer enough to build a deep semantic understanding of the document content. For example, one needs to understand the rows, columns, cells, and relationships to decode what a table describes. A number of visual representations (such as graphs and infographics) have become integral to the document understanding in these application domains.

This now demands a dedicated program or mission for enabling analytics over large collections of documents and videos. iHub Drishti has identified document image analysis as one of the verticals and the objective is to recognize the text and graphics components in images and to extract the intended information as a human would.

2. National Need and Impact

Understanding text in images has many potential real-world applications. Examples include, but are not limited to, e-governance, digitization of health records, education, and robotic process automation in large enterprises. In linguistically-diverse countries, like India, text understanding often comes with additional challenges with relation to processing rich linguistic structures and semantics. Moreover, one of the application areas that we would be focusing on to demonstrate the utility of document image understanding, is education. In India, technological development in scalable and online education is the need of the hour. India has an acute shortage of good quality teachers, especially in rural areas. Given this scenario, online recorded video lectures and digital libraries often come as a boon. However, this also comes with its share of challenges, such as the creation of- and efficient access to- large collections of these resources given our language diversity. Visual text analytics aims to provide technological solutions to these problems.

The Government of India has already identified the future need for online education. The National Digital Library of India (NDLI) a project under the Ministry of Human Resources and Development (MHRD) is already dedicated to the nation and function. The National Programme on Technology Enhanced Learning (NPTEL) is another initiative by MHRD with IITs and the IISc. As per Wikipedia, NDLI and NPTEL contain more than 45 million scanned books and 20K instructional videos, respectively. Majorly, these resources contain images or videos which contain text as important cues. In the context of such large-scale online resources that will continue to grow in the coming years, there are several concerns, for example, developing a

smart interface to these resources where a learner can efficiently perform interactive question answering, could revolutionize online learning.

3. Objectives

The major objectives of this project are grouped into the following three categories:

Core Research: Under the aegis of iHub Drishti, we aim to build core technologies for individual steps involved in the entire pipeline of document analysis.

- Spotting, matching and retrieval schemes
- Complex layout analysis
- Multilingual document image analytics
- Question answering and information retrieval
- Interpreting graphical components
- Text-aware video and image compression schemes

Applications: The core technologies developed in this vertical can be applied to a number of areas that require document analysis. Given the current situation due to COVID19 and post-pandemic predictions, efficient methods of delivering and facilitating online education has become one of the top priorities. Currently, one of the application areas in this vertical is online education. In addition, we will partner with organizations that require technology solutions for text or document analysis. Some of the specific challenges related to this application are:

- Smarter interface to the digital library
- Video indexing and retrieval system
- Trustable and explainable automatic evaluation system

Education and Outreach: Generating trained manpower in the area of document image analysis is an important goal.

4. Research Challenges and Scope of work

Under this vertical of iHub Drishti, we plan to work on two broad modules: (i) Core technologies and (ii) System development in the education domain. In this section, we will provide detail of the research challenges and scope of our work. This vertical will work on a set of problems awarded through open calls for multi-institute consortiums.

Core technologies for Document Image Analytics

Spotting, indexing, and retrieval schemes: Word spotting, effective indexing scheme, and efficient retrieval for industry-scale data are open areas of research in the document image analysis community. Adopting modern representation learning techniques from computer vision for word images and representing them discriminatively is one of the research directions in this area. At our T-hub, we plan to work this important, highly explored but yet an open problem in document image analysis.
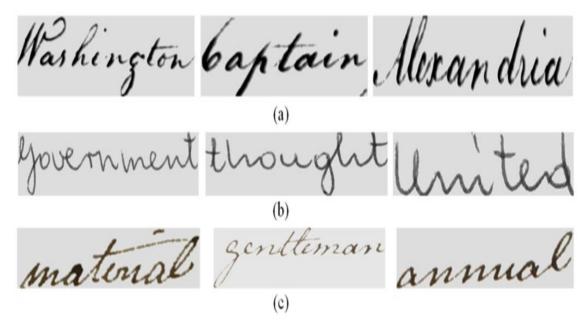
Figure 2: Images taken from publicly available handwritten datasets. Discriminative and universal representation for handwritten word images is an open area where our T-hub aims to contribute.

Complex layout analysis: Document images come with highly complex layouts which are often non-trivial to analyze. Few such examples are shown in the Figure. Semantic labeling of components in text and graphics and beyond (for example, title, caption, header, footer, logo, etc.) has an analogy with scene graph generation techniques for natural images in computer vision. We plan to explore these techniques for methodological developments on complex layout analysis.

Figure 3: Document images come with highly complex layouts which are often non-trivial to analyze. Few such examples are shown here. Such challenges call for investment in methodological developments on complex layout analysis.

Reading charts, drawing, tables and equations: Typical OCR systems often only read textual content and ignore charts, drawing, tables and equations, etc by considering them as graphics. We aim to fill this gap in OCR systems by developing robust readers for these elements. These unstructured elements present in books/slides are often information-rich and automatically converting these elements into a more structured form, for example, chart-to-

table (shown in Figure 4) can have many downstream applications, for example, explaining some chart to visually-impaired person, efficient access to scientific resources, etc. Though there are some works in this direction, a rigorous effort is missing in the literature. We plan to address this problem using recent advancements in sequence to sequence learning in the field of NLP and Computer Vision.
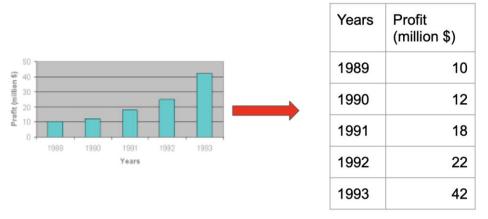
Figure 4: Typical OCRs often ignore charts, tables, equations, drawings, etc. These unstructured elements present in books/slides are often information-rich. Our goal is to fill this gap in modern visual text analytics by automatically converting these elements into a more structured form, for example, chart to table.

Multilingual document analytics: In language-diverse countries such as India, many languages and scripts exist. Therefore, many historical documents, office orders, etc. often contain text from multiple scripts. Semantically indexing multilingual documents, translation (pre-recognition), and cross-lingual access (e.g., seamlessly searching via English query to obtain relevant Hindi text) is one direction where our T-hub plans to work.
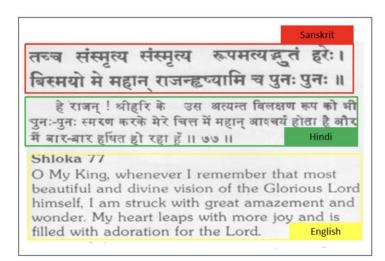
Figure 5: Being a multilingual country many document images in our country are multilingual in nature. In this example, three languages are shown, viz. Sanskrit, Hindi, and English. A rigorous effort around semantically indexing multilingual documents, translation, and cross-lingual access is needed. This is one direction where our T-hub plans to work. Figure taken from [Mathew et al., 2016].

Open-ended Question Answering on Document images and videos: Question answering on document images has been emerging as an important area in the ICDAR community. Accessing large-scale content e.g. scanned books and MOOC videos via answering natural language questions is an interesting outcome of this line of research. Consider, someone wants to learn the Binary Search Tree and ask a question: what *is the order of keys in an inorder traversal of a BST?* We aim to develop QA systems that can read and interpret slides/books in video lectures and not only answer the question but also point to specific frames where this concept is discussed in the video lecture. A couple of examples illustrating the utility of document image-based question answering are shown in Figure 6.
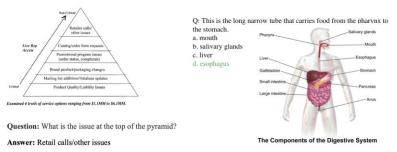
Figure 6: Document Image Question answering aims to answer natural language questions by reading and interpreting document images, and it has several applications in the field of online education [Mishra et al, 2019; The CVPR 2020 "Document Visual Question Answering" (DocVQA) challenge, Kembhavi et al, 2017]

Text-aware Image and Video Compression algorithm: With the advent of advancement in mobile-phone cameras, it has been used for various requirements; including imaging of the black/green board of the classroom, notice board, presentation slides during seminars, and screenshot during synchronous mode of e-teaching/learning. However, all of these images need skew correction and also need a compression algorithm that should be able to compress salient and non-salient parts differently, for a given file size requirement. As an example, the text is the most important component in the case of such images and hence it can be labeled as the salient region. Other examples can be advertisements hoardings where people, in general, capture its image, and the information therein is used at a later stage. Here also, the text including numbers (such as phone numbers) are salient regions. In this work, it is proposed to develop compression algorithms for compound images, having text and non-text regions, all related to e-learning. For compression, JPEG is the most popular generic algorithm for compression of images. A customized algorithm is expected to provide an improved compression ratio for the given quality requirement. Alternatively, such an algorithm is expected to provide a higher quality of image for a given compression ratio. In the proposed work, salient regions (text) will be identified and DCT coefficients of such regions will be decently quantized as compared to non-salient regions, for a given file size requirement. In case file size permits, salient regions, if small, can be compressed losslessly while other parts will be lossy compressed, resulting in a hybrid method of compression – lossless and lossy.

In the case of transmission of video lectures, the following scenarios can be thought of. During lecture delivery, some instructors use slides while others prefer tablets or other similar gadgets for writing the concepts to be delivered to the students. In both the cases of synchronous mode of teaching, instructors deliver lectures with the transmission of video and audio. In addition to this, the slide mode of instruction, i.e when the text content is prepared in advance,

instructors spend typically 2 minis per slide, unlike in case of seminars/conferences, where typically 1 min per slide is spent. Therefore, the development of compression algorithms needs to be adaptive in this sense too, it should estimate the activity level on the projected slide and take decisions on the required compression strategy. It is easy to visualize that video is not very important but the text and audio are the most important components in such mode of teaching/learning. Here also, identification of salient features (text part of a video frame) is important. While the lecture is going on, only still images of the instructor can be shown at one corner of the frame, in case of availability of the narrow bandwidth channel. In this case, text and audio must be given high priority and needs to be done adaptively, depending on the channel bandwidth.

The above background is with respect to the e-mode of teaching-learning but the application of such compression algorithms has a much wider scope mentioned at a later stage. We plan to address this problem by:
- Development of image compression algorithms for such images.
  a. Identification of text components of such images, preferably in the DCT domain, for careful quantization of such coefficients. This will enable the plugging in the algorithm in JPEG - the most popular image compression method in use.
  b. Labeling the saliency with different levels for adaptive quantization
- Development of video compression algorithms for online education purpose
  c. Since computational cost is an important issue in the online delivery of lectures, a careful study will reveal how much adaptive compression algorithms can be developed for video. Highly adaptive methods, in general, are computationally demanding and hence online delivery of lectures may be a matter of serious concerns
  d. A study may help in finding the relationship between achieved compression ratio and the computational complexity

Text-aware Compression methods have applications in other areas as well. For example, in case of compression algorithms for KYC documents, the text components of the document are important. KYC document uploading is one of the mandatory requirements for getting the benefits of several government welfare schemes. Some agencies, such as mobile communications service providers, also ask for KYC documents for other purposes. Since users themselves acquire the image of such documents, the image used to be (a) skewed, (b) more than required dimensions, and (c) file size also used to be larger than the one specified by the agency. Users try to correct these errors, but sometimes, they postpone the uploading to the later stage or give-up uploading due to cumbersome and time taking correction process. In this application, a system can be developed that will have a drop-down menu to choose the document and will accept the acquired document image as-it-is. The system does the necessary corrections at the backend. The document image will be skew corrected, dimensionally adjusted, and appropriately compressed. As the various KYC documents are different from each other; the proposed system will have document-specific-compression algorithms for different types of documents. The developed algorithm will judiciously compress the salient (text and numbers) and non-salient regions while meeting the desired file size requirement specified by the agency.

In addition to the above, the current trend includes the recording of lectures delivered by the speakers at conferences. This is generally done by mobile phones. Since there are storage

memory constraints in case of such recordings, the compression algorithm, if switched for this purpose, should record the audio and the text displayed on the screen. Maybe for an initial 1-2 mins, the video can also be recorded to refer to the speaker at the later stage.

System development for education domain

Online learning and education has seen a significant boost in the past few years. It has the potential to become the solution to scalable quality education in India which seriously lacks in quantity and good quality teachers, especially in rural areas. The current COVID scenario has further triggered the need to invest in technological development to support online education. There are two major components of an online education:

1. Efficient access of the content (e.g., satisfying information needs of students given large-collection of videos or scanned books)
2. Scalable evaluation schemes (in an online setting we are now talking about class size of 1000 or more. In such large classes, we need technological solutions to make evaluation automatic or semi-automatic).

In this context, *visual text analytic*s (i.e., interpreting text appearing in videos or images) can prove an important direction. Given large scale availability of scanned books, for example, the National Digital Library of India (NDLI) contains over 4 million scanned books, and Massive open online courses (MOOC), seamless access to the content has emerged as a challenge. Consider an example where students want a specific definition of some terminology in a book and its explanation in a video. Can we enable such access? At TIH-Drishti our one of the focus is to enable cross-language smooth access of education resources. The second major objective under this vertical will be to develop systems for automatic assessment, for example, assessing handwritten assignments automatically or semi-automatically.

In the TIH, we plan to work on system development specially online and scalable education as a use case. To this end, we plan to work towards developing the following two systems:

Smart interface to Digital Library: The number of Massive Open Online Courses (MOOC) and other online resources is increasing rapidly. In the context of such ever-growing searching and accessing *topics of interest* has become an important challenge. Consider the scenario where a student is trying to find a relevant video to learn any given topic. The modern search techniques will retrieve hundreds of videos with different content, presentation styles, duration, and teaching quality. Finding the right video that is most suitable to the student's background, learning goal and preferred learning style is a challenging problem. Moreover, directly accessing the topics of interest is only possible, if videos are indexed properly. Given the large number of education, automatic indexing video is a challenging problem. There is early progress in these lines of work, but this is far from being solved *[Mahapatra et al, 2018]*. This problem will be worked upon by the IIT Jodhpur team in collaboration with our industry collaborator, VideoKen.

Figure 7: Indexing video has emerged as an interesting problem in the context of online education. The problem here is to automatically create a table of content for the video lectures [Mahapatra et al, 2018].
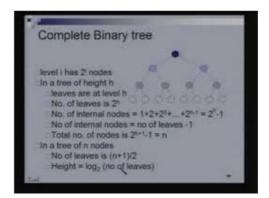
One example illustrating this problem is shown in Figure 7. The specific challenges in this direction are as follows:

- *Picking titles from Slides Videos:* VideoKen deals with a variety of slides frames, the current methods pick topics from slides well when consistent slides/template is used across a video but sometimes breaks when designer slides are used.
- Creating a table of contents (ToC) for Non-Slides Videos: Non-slides videos could have different types i.e. animation, Digital blackboard (for example, lectures from Andrew Ng or Khan academy), or Classroom Blackboard.

Trustable and explainable automatic evaluation system
Evaluation is an important component of education and e-learning. One of the system development efforts at T-hub will be to develop a trustable and explainable automatic evaluation system for online learning. Towards this end, the following two components will be explored:

Visual Question Generation: The goal of visual question generation (VQG) is to generate natural questions from a given image or a video. This can immensely help online scalable education by automatically generating questions for reading comprehension materials for online educational videos or presentation slides [Heilman and Smith, 2010]. Consider an example shown in Figure 8. Given a video lecture on a binary tree, can we develop a VQG system that learns to ask questions such as *How many nodes does the fourth level of a complete binary tree contain?* And predicts the corresponding correct answer. Other directions in this space are mapping Questions and Videos. Many organizations have a vast number of question libraries as well as recorded videos. For each video, we need to develop methods to pick the right set of questions that can be presented to the learner as well as vice-versa where for a given question.

Figure 8: Given a lecture video or images of slides, our goal is to generate relevant questions and answers automatically. Such automatic question generation ability can have a significant impact on online learning.

Automatic assessment of handwritten assignments: Scaling education to thousands of students requires large-scale assessment to be done with least or no manual effort. In this regard, automated assessment of handwritten answers is an important problem. This has significant importance in enhancing the throughput of instructors. The aim of this problem is to evaluate a student's handwritten answer by assigning an evaluation score that is comparable to the human-assigned scores. There are early studies to address this problem in the literature [*Rowtula et al, 2019;* Sharma and Jayagopi, 2018, Kanejiya et al., 2003; Mohler et al., 2011]. However, this problem is far from being solved in a rigorous real-life setting. Our aim is to develop automatic or semi-automatic systems for automatic assessment of one word, short, and long handwritten answers. Figure 9 illustrates the problem of the automatic assessment of handwritten assignments.
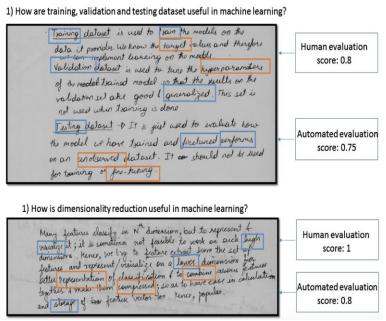
Figure 9: Automatically evaluating handwritten assignments has a lot of importance in scalable education. One of our aims is to develop a system that can automatically grade short handwritten given a model answer [Rowtula et al, 2019].

5. Outcomes and Deliverables

1. Foundational Research: Address various fundamental research challenges related to visual text analytics. These research innovations will be submitted to top-tier publications, patent and open source codes
2. Technology Solutions:
   a. Technological solutions to scalable online education
   b. Systems for efficient access of online education resources
   c. Automatic assessment tools
3. Entrepreneurship and start-ups: The TIH will encourage students and scholars to become entrepreneurs by creating start-ups. It is expected that activities in the various projects undertaken in TIH will provide avenues for initiating start-ups.
4. Industrial collaboration: We have already collaborated with two startups in this vertical. As several of the projects will be in open call as consortium mode, we aim to build a collaboration of the TIH with several existing and new industrial partners.

6. Training and Outreach

The technologies developed as part of the document image analytics vertical can be applied to a variety of applications involving document analysis. We will organize courses to train students and professionals in this interesting and important area. The BTech in AI and DS at IIT Jodhpur also has a specialization in Language Processing. Some additional activities that we plan to pursue are:,
1. Capacity building of students and faculty by organizing summer and winter schools,
2. Skill development and training by offering certificate programs, and
3. Organizing national and international workshops.

7. Budget

This vertical will be primarily on an open call basis. We will partner with user agencies and define the request for proposals, and support academic institutions and startups to create and deliver technologies. IIIT Hyderabad, IIIT Bangalore, VideoKen and KritikalVision.AI are four of our major collaborators in this vertical. Projects teams, while contributing and demonstrating the technology on the demo system will also work on specific research and fundamental problems associated with this.

8. Workplan

Year 1				Year 2				Year 3				Year 4				Year 5			
Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
Staff and student Hiring																			
Design and open call for a consortium project																			
Data collection/annotation																			

Lab development and purchase																
		Core research														
		Development of Technology														
									Integration and testing							
		Education and training for developing manpower														

9. References

[Alaei et al., 2012] A. Alaei, U. Pal, and P. Nagabhushan (2012), "Dataset and ground truth for handwritten text in four different scripts," IJPRAI.

[Dutta et al., 2018] Kartik Dutta, Praveen Krishnan, Minesh Mathew and C.V. Jawahar (2018), Towards Spotting and Recognition of Handwritten Words in Indic Scripts, The 16th International Conference on Frontiers in Handwriting Recognition, USA

[Goyal et al., 2019] Yash Goyal, Tejas Khot, Aishwarya Agrawal, Douglas Summers-Stay, Dhruv Batra, Devi Parikh (2019): Making the V in VQA Matter: Elevating the Role of Image Understanding in Visual Question Answering. International Journal of Computer Vision 127(4): 398-414.

[Guo and Zhang 2010] Guo, C. and Zhang, L. (2010), "A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression", IEEE transactions on image processing, Vol. 19, No. 1, pp. 185–198, 2010.

[Hao et al., 2016] L.Hao, L.Gao, X.Yi, and Z.Tang (2016), A table detection method for PDF documents based on convolutional neural networks, in IWDAR.

[Heliman and Smith, 2010] Michael Heilman and Noah A. Smith (2010). Good question! statistical ranking for question generation. In Human Language Technologies: ACL, pages 609–617.

[Kanejiya et al., 2003] D. Kanejiya, A. Kumar, and S. Prasad (2003), "Automatic evaluation of students' answers using syntactically enhanced LSA," Workshop on Building Educational Applications Using Natural Language Processing - Volume 2.

[Kembhavi et al., 2017] Aniruddha Kembhavi, Min Joon Seo, Dustin Schwenk, Jonghyun Choi, Ali Farhadi, Hannaneh Hajishirzi (2017): Are You Smarter Than a Sixth Grader? Textbook Question Answering for Multimodal Machine Comprehension. CVPR: 5376-5384.

[Mahapatra et al., 2018] Debabrata Mahapatra, Ragunathan Mariappan, Vaibhav Rajan, Kuldeep Yadav, Seby A., Sudeshna Roy (2018): VideoKen: Automatic Video Summarization and Course Curation to Support Learning. WWW (Companion Volume): 239-242.

[Mitra and Veeraraghavan, 2015] Barua, S., Mitra, K., and Veeraraghavan, A. (2015), "Saliency guided Wavelet compression for low-bitrate Image and Video coding", pp. 1185–1189, 2015.

[Mohler et al., 2003] M. Mohler, R. Bunescu, and R. Mihalcea (2011), "Learning to grade short answer questions using semantic similarity measures and dependency graph alignments." Association for Computational Linguistics.

[Rowtula et al., 2019] Vijay Rowtula, Subba Reddy Oota, Jawahar C. V. (2019): Towards Automated Evaluation of Handwritten Assessments. ICDAR: 426-433.

[Sharma and Jayagopi, 2018] Annapurna Sharma, Dinesh Babu Jayagopi (2018): Automated Grading of Handwritten Essays. ICFHR: 279-284.

[Xu and Cohen, 2016] Xu, J., Joshi, R., and Cohen, R. A. (2016), "Overview of the emerging HEVC screen content coding extension", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 26, No. 1, pp. 50–62.

Industry 4.0 & Cloud Manufacturing

Collaborating   Organizations:
IIT Jodhpur
Industry   Partners: Kansara   Modler   Pvt.   Ltd.,   Jodhpur
Foreign Institutes: Texas A&M University, Ecole Nationale Supérieure d'Arts et Métiers, France

Introduction

Technological inventions played a major role in increasing industrial productivity and quality over the past few years. The inventions are demonstrated as previous three industrial revolutions, Mechanization (powered by steam engines in the 1800s) as first revolution, Mass production (powered by electricity and the assembly line in the early 1900s) as second revolution, and Automation (powered by computers in the late 1900s) as third revolution [Industry 4.0 working group, 2013]. The rapid developments in the Internet and embedded system technologies in recent years gave rise to a number of novel technologies, such as Cyber-Physical Systems (CPS), Internet of things (IoT), Cloud Computing, and Big Data Analytics. The advent of these new generation technologies enabled creation of smart, networked entities crafted with a certain degree of intelligence, and connected to each other. The widespread deployment of sensors in manufacturing, extensive application of software and internet connectivity between entities in industrial production brought physical and virtual worlds together gave rise to CPS. All these transformations mark the transition of current industrial production to the fourth stage -Industry 4.0, which is characterized by smartness and networking [Industry 4.0 working group, 2013].
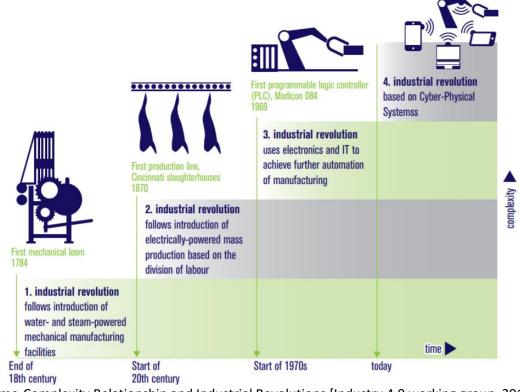
Figure 1: Time-Complexity Relationship and Industrial Revolutions [Industry 4.0 working group, 2013]

During the same time as the third industrial revolution, a need of increased sharing and collaboration among manufacturing resources was realized. Some of the studies highlighted the drawbacks with existing advanced manufacturing technologies and showed that the industrial resources are unable to collaborate and share the expertise [Tao et al., 2015]. The concept of cloud computing introduced a new way of delivering by sharing of computational resources. The combination of cloud idea and manufacturing resulted into a disruptive paradigm known as cloud manufacturing [Li et al., 2010]. Industry 4.0 and cloud manufacturing constitute two major initiatives seeking benefits of information technologies to promote new

generation manufacturing and transforming the entire manufacturing community. The TIH, iHub Drishti plans to conduct cutting edge research in the domain of Industry 4.0 and cloud manufacturing in the country. The hub aims to actively engage in the development of technical know-how, industry-deployable hardware and software, training modules and educational activities in the domain. It is envisioned to achieve the same in close collaboration and partnership with industries and academic institutions.

Background

The TIH iHub Drishti is a part of the National Mission on Interdisciplinary Cyber-Physical Systems focussing on Industry 4.0 and Cloud manufacturing. The HSRC Global Industry 4.0 Market and Technologies 2018-2023 report estimates the global Industry 4.0 market to reach $214 billion by 2023 [HSRC Report, 2018]. The majority of economies of the globe (both advanced and developing) are encouraging and adapting Industry 4.0 initiatives to increase manufacturing competitiveness of their organizations. For example, Malaysia, Thailand and Indonesia are implementing Industry 4.0 to truncate their journey to a higher base of economic benefits. Singapore launched the "Singapore Smart Industry Readiness Index" to provide a common framework to motivate industries to start, scale and sustain Industry 4.0 efforts as a hub for high-value manufacturing. China is a major economy accounting for about 25% of the world's manufacturing activity launched initiatives such as "Scientific Innovation 2030" and "Made in China 2025" to enable manufacturers in prioritizing Industry 4.0. The Government of India's push to manufacturing through the "Make in India" initiative and subsequently formulating National Policy for Advanced Manufacturing aims at increasing the contribution of manufacturing output by considering Industry 4.0. A PWC report published in 2016 indicated that about 80% Indian manufacturing industries expected more than 10% improvement in efficiency and over 60% of respondents expected 10% improvement in the additional revenue. The last few decades of the twentieth century brought a huge leap forward in the third industrial revolution due to the advent of computer systems and automation. India did not grab the opportunities due to ample availability of labour and limited access to software. The majority of Indian manufacturers were stuck at industry 2.5 with paper-based processes and heavy human dependency. In spite of poor resource engagement, Reserve Bank of India reported that 1,734 Indian manufacturing industries posted INR 47,100 crores in Q3 FY2019, a growth of 29.4% from the same period in previous year. The financial growth is an indicator of the strength and ability of Indian manufacturing to perform even at sub-optimal levels. If successful technological innovations are conducted in a planned manner, India can catapult directly into the next revolution i.e. industry 4.0.

Industry 4.0 will yield the benefit of a faster learning cycle and give an edge to Indian companies that are competing with legacy producers in developed countries. Therefore, it is pertinent to generate the competence and resources to undertake research and development activities in these. The TIH on Industry 4.0 and cloud manufacturing at IIT Jodhpur considers this as an opportunity at the right time to drive these activities through a consortium of experts within the institute accompanied by industrial and academic collaborators across the country.

National Need and Impact

The issues linked with productivity and quality have been the primary focus for manufacturing industries during each industrial revolution and resulted in multiple inventions. For example,

Ford conceived the concept of mass production assembly line to aid an individual in performing assigned tasks in a meaningful manner for realizing better quality products. Since the invention of the assembly line, many transformations occurred to enhance product quality e.g. replacement of manual processes with machinery and/or replacement of humans with robots. The product quality monitoring and control has been one of the cost intensive functions for the manufacturing industries encompassing almost 15%-20% of sales revenue. The cost levels as high as 40% of the total cost of manufacturing operation are also observed by many MSMEs [Matt et al., 2020]. SMEs and MSMEs manufacture components, sub-assembly and assembly level items for the Original Equipment Manufacturers (OEMs) with very limited focus on research related to processes and systems on the shop-floor.

The industries in the Indian MSME sector are the backbone of economic structure and have unremittingly acted as the bulwark for the economy, providing it resilience to ward off global economic shocks and adversities [CII, 2020]. MSMEs contribute around 6.11% of the manufacturing GDP, 24.63% of the service sector GDP and 33.4% of manufacturing output with ~63.4 million units throughout the country. MSMEs provide employment to around 120 million persons and contribute around 45% of the overall exports from India. The MSME sector has consistently maintained a growth rate of over 10% and about 20% of the units are based out of rural areas. The sector caters to the manufacturing needs of major industrial sectors within the country such as telecommunication, automotive, healthcare, electronic goods, etc. and has a geographically diverse presence [KPMG, 2017].

(a) Sector-wise Breakup   (b) State-wise Breakup
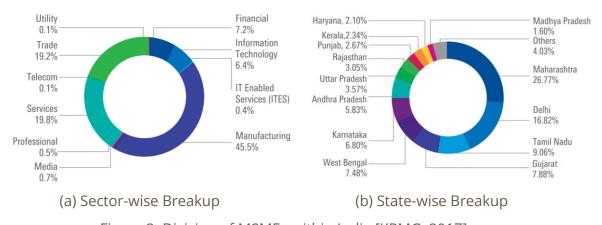
Figure 2: Division of MSMEs within India [KPMG, 2017]

Figure 2 shows sector-wise and state-wise distribution of MSMEs within the country. It can be seen that the manufacturing sector contributes almost half of the total output generated and it is of significant importance in the growth of the country. This indicates the deployment of a significant rural workforce in the MSME sector and is an exhibit to the importance of these enterprises in promoting sustainable and inclusive development as well as generating large scale employment, especially in the rural areas. The major inhibitors for the growth of MSMEs are cash flow, lack of capital and credit finance, inadequate infrastructure, use of outdated technology, lack of skilled manpower and problems of market connection [Kumararaja, 2020]. The financial and infrastructural issues can be addressed by interventions from state and central government ministries and other linked agencies. But, other problems require India-focussed technological innovations aiming at the success of the "Make in India" program. The COVID-19 crisis has merely increased the magnitude of existing problems for these industries.

A meaningful strategy to implement industrial automation will be able to handle the issues linked with manpower shortage and demand fluctuations effectively. It will also enable dealing with an irregular product mix providing increased productivity, flexibility and time efficiency. It is necessary to deploy affordable automation to support employees, simplify the process, and improve growth trajectory over sophisticated high-cost automation. Such technological interventions can be a great difference on shop floors of the MSMEs improving their ability to serve multiple customers catering to different markets. A smart SME architecture consisting of three core pillars namely, affordable and intelligent automation, real-world interface and compact resource planner connected to a seamless common data exchange network can be an ideal solution for these issues [Kumararaja, 2020]. Figure 3 shows an overall architecture for the development of smart MSME in an Indian context.
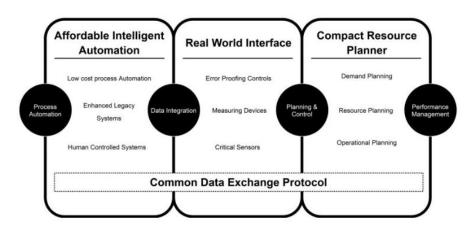
Figure 3: Architecture for Smart MSME [Kumararaja, 2020]

The TIH iHub Drishti at IIT Jodhpur will aim at developing vision-based automation solutions to improve operational efficiencies and global outreach of Indian MSMEs. This will be attempted through active research collaborations with MSMEs to carry out initial needs assessment and determine critical processes requiring technological interventions followed by connection with Original Equipment Manufacturers (OEMs) for integration of requirements and quality expectations. One of the major bottlenecks for industries in the MSME sector is the inability to detect a slight variance in production processes or material state which makes the entire production run defective. The inability to detect this result into multiple operational problems such as reduced productivity, quality non- conformance, delivery failures, and lack of flexibility to meet demand. The extremely thin profit margins and return on investment in the sector coupled with tough competition from the global manufacturers result in loss of business. One of the classical examples related to lack of automation in the domain is the manual inspection of raw materials, processed parts and finished components which is error-prone, cumbersome, expensive and human intensive. The manual inspection activity involves significant training for human experts to perform the task. The inspection activity is not scalable and requires further training with the change of product line. As the inspection activity is carried out offline, the feedback to the manufacturing operation is not possible. It will result in significant rejections as multiple items are manufactured due to continuous operations adding scrap costs to the manufacturing.

An automated inspection powered by machine vision emerged as a replacement for cumbersome and error-prone manual activities in the last few years. The machine vision has

the potential to be an integral part of many manufacturing applications to achieve autonomous manufacturing and inspection abilities on the shop floor with minimum human interference. The scope of automated inspection systems is majorly limited to tasks such as measuring/gauging and help in meeting stringent quality requirements. It has not been extended and conceived as a major option for manufacturing industries in the MSME sector as a viable solution to remove the bottlenecks. In recent times, manufacturing enterprises across the globe are undergoing major transformations due to evolution of Smart Manufacturing or Industry 4.0. The new manufacturing philosophy aims at IoT-based intelligent automation, improving standalone machines using network-connected entities based on machine-to-machine and machine-to-person communications, augmentation of artificial intelligence and machine learning in the process of decision making, implementation of cloud computing concepts in manufacturing, etc. The importance of machine vision is expected to grow multi-fold in manufacturing industries with the advent of Industry 4.0. The vision system is expected to play a critical role in the IoT layer to provide image data through a connected network to a big data cloud server. The application of machine vision can provide awareness to a machine about its surroundings through high-order image recognition and decision-making. The raw data captured from the vision hardware requires stringent mining and cleaning procedures to remove unnecessary information due to disturbances and outliers captured from the shop floor environment. The filtered information can be used as an input for artificial intelligence or machine learning driven techniques to detect and describe what happened to the product, determine why that happened, predict what may happen and prescribe which remedial actions must be taken [Pal et al., 2016; Zhong et al., 2017].
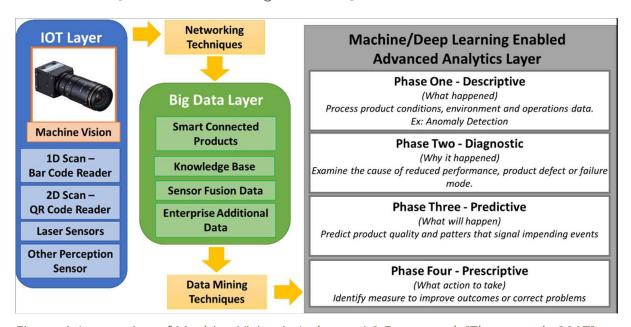
Figure 4: Integration of Machine Vision in Industry 4.0 Framework [Zhong et al., 2017]

Figure 4 shows a simplified diagram of a vision system in Industry 4.0 setting indicating the relevant components. The implementation of future machine vision solutions such as knowledge-driven decision-making, real-time control, online advanced analytics is challenging and requires diverse expertise. It may take at least a decade for industries to achieve a concrete degree of maturity in obtaining fully-operational functionalities [Leitao et al., 2016]. The implementation of machine vision systems requires a set of visual sensors such as cameras

and illumination systems accompanied by other sensors for enhanced data collection (hardware component) by registering event/s of interest. The application of newer set of imaging techniques such as hyperspectral imaging and polarization imaging are employed to broaden the scope of vision systems to identification of chemical composition [Warren and Cohn, 2017] and stress patterns [Richert et al., 2010] in the work material respectively. The hardware component is accompanied by an interpretation system that uses built-in intelligence (software component) for interpreting the event and acting on the same according to predefined instructions (computer logic). The interpretation system involves development of computational models and vital elements performing the majority of the work. The models decompose incoming visuals captured using a hardware system and process the same to extract relevant information or status of the manufacturing system. The involvement of new generation artificial intelligence and machine learning tools do not need them to be constantly programmed with explicit instructions for analysing the data. These tools broaden the limits of what a computer and camera/sensors can determine/predict/analyse and allow computerization of visual and inspections that previously required specific human expertise. The vision system can be augmented further with additional sensors to capture critical information such as the status of the component, machines and overall manufacturing system. The additional information related to these aspects can aid manufacturers in developing prognostic ability related to the components of organization which is costly otherwise. The development of such inspection and prognostic systems developed by an individual MSME will not be practical as it requires manpower with significant expertise and considerable financial commitments. The proposal aims to adapt the concept of cloud computing that has been successfully employed in sharing of computational resources. Cloud computing aims to realise the idea of offering computing resources as services in a convenient pay-as-you-go manner. This idea was expanded into the manufacturing realm by introducing the concept of cloud manufacturing [Li et al., 2010]. The most prominent feature of cloud manufacturing is the seamless and convenient sharing of distributed manufacturing resources and realising the idea of Manufacturing-as-a-Service. The concept aims to provide various services to manufacturing industries over the internet as conveniently obtaining water or electricity. The concept has the potential to revolutionize the MSME sector as the significant initial investment in hardware and software will not be essential.

The technological solutions developed as a part of the proposal will facilitate MSMEs to achieve an ability to overcome shortcomings mentioned earlier without major transformations in the set up. The TIH will enable the combination of knowledge from manufacturing domain, computer vision, sensors and IoT, artificial intelligence and cloud manufacturing to meet the deliverables. The technologies developed by TIH will help MSME to improve manufacturing quality and performance by eliminating defects, performing product verification, enabling tracking at every stage of the production process. The industries will realize the benefits in terms of lower manufacturing costs, higher customer satisfaction and wider reach through the technological solutions.

Planned Activities

This section summarizes the preliminary project proposals planned by TIH to achieve the deliverables highlighted in the previous section.

Developing Cloud Manufacturing architecture for MSMEs in Manufacturing Sector

Cloud manufacturing is defined as an integrated cyber-physical system that can provide on-demand manufacturing services, digitally and physically, at the best utilisation of manufacturing resources [Wang et al., 2014]. A cloud manufacturing system encompasses three broad categories of participants; an operator or cloud provider, resource and service providers, and resource and service consumers [Adamson et al., 2017]. Figure 5 highlights linkages among the participants and operating principle of a typical cloud manufacturing system. The operator manages and operates the cloud platform by delivering the required support and functions to resource providers and consumers. The operator maintains services and technologies required to run the system by finding, combining, controlling and coordinating the required services for fulfilling consumer requirements. The resource provider encompasses different industries offering various types of resources or services to a cloud manufacturing platform on the sharing basis. The cloud manufacturing provides on-demand manufacturing services to consumers e.g. OEMs by submitting service requirements to the cloud platform on pay-per-use basis.
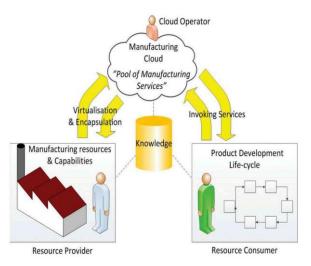
Figure 5: Operating Principle of Cloud Manufacturing [Adamson et al., 2017]

There are a number of research initiatives at an international level related to cloud manufacturing with academic and industrial participation. The ongoing development activities related to cloud manufacturing are in its formative stages and have challenges linked to conceptual development, technology expansion, standardization of information, etc. Some of the underlying challenges for the successful implementation of cloud manufacturing are listed below which requires significant efforts and technological developments.

1. How to achieve the integration of various technologies such as cloud computing, IoT, Semantic Web, high-performance computing, embedded systems etc.
2. How to bring various types of resources and capabilities to the cloud as services, especially the implementation of Hardware-as-a-Service (HaaS) (i.e. knowledge-based resource clouding)
3. How to achieve overall management and control of clouds (including service composition, collaboration between cloud manufacturing applications, open communication standards, distributed control and coordination of manufacturing equipment, and user interfaces in cloud environments)

The iHub Drishti aims at the development of unique and indigenous cloud architecture for MSMEs in the manufacturing sector. The proposed cloud-based architecture will not only enable MSMEs as a service provider on the internet based platform but also permit seeking of need-based services from third-party technology related experts via modules. The MSMEs suffer from the severe shortage of manpower expertise and adequate technical resources due to various cost constraints. It results in inferior products, use of sub-optimal processes, increased rejection rates, larger product development cycles and higher manufacturing costs. Therefore, it is essential to embed assistance features to the cloud in the form of pay-as-you-use to enhance the competitiveness of MSMEs. The platform will offer a set of critical services which can be realized at much lower cost by multiple MSMEs through the cloud manufacturing philosophy. Some of the of important and useful functions for MSMEs to be embedded into the proposed cloud architecture are listed below;

1. Design Expert: It will offer services related to the design function through a shared library of CAD tools with adequate expertise as design consultant.
2. Analysis Expert: It will offer services related to the analysis function through a shared library of analysis tools with adequate expertise as analysis consultant.
3. Manufacturing Expert: It will offer services related to selection and usage of appropriate manufacturing processes and parameters through the use of computational tools and experts.
4. Tooling Development Expert: It will offer services related to the design and development of appropriate tooling required to manufacture the product.
5. Quality Control Expert: It will offer services related to inspection and quality control of products manufactured within the organization.
6. Logistics Expert: It will offer services related to the optimal selection of logistics services in guaranteeing just-in-time delivery of products to service requires.
7. Maintenance and Service Expert: It will offer services related to the online monitoring of machines and generate feedback related to maintenance and service related issues.

It is expected that design and manufacturing resources will be provisioned as services from the cloud, for all phases of the product development lifecycle with evolution of Industry 4.0 and cloud manufacturing concepts in coming years. The traditional manufacturing workflow consisting CAD, CAM, CAPP combined with CNC running on local hardware (Figure 11(a)) will be transitioned to a cloud ecosystem where various elements run on cloud computers (Figure 11(b)). The users will be able to seamlessly connect with this system from anywhere in the world for on-demand design and manufacture of parts for a variety of manufacturing machines. The primary challenge associated with the development of such systems will be ability to capture manufacturing process and systems knowledge as software, standardization for information exchange by developing interoperability standards, communication protocols, cybersecurity issues, transmission of signals in real-time, etc. The TIH will also attempt developing solutions for these issues along with implementation of above-mentioned architecture and computational tools.

Vision-based Inspection and Prognostic Solutions for Manufacturing Operations

According to the Automated Imaging Association (AIA), machine vision encompasses all industrial and non-industrial applications in which a combination of hardware and software provide operational guidance to devices in execution of their functions based on capture and processing of images. Though industrial computer vision uses similar algorithms and approaches as academic/educational and government/military applications, the constraints are considerably different [Cognex, 2020]. Industrial vision systems demand robustness, reliability, and stability accompanied by much lesser cost than those used in governmental/military applications. Machine vision systems rely on digital sensors protected inside industrial cameras with specialized optics to acquire images which are further processed by software for analysing and measuring various characteristics for decision making. The system may also require integration with additional sensors to enhance the decision making ability. Figure 7 shows a vision based bottle fill-level inspection system employed in the pharmaceutical industry. Each filled bottle passes through an inspection system that triggers a vision system to flash a strobe light and take an image of the bottle. After acquiring an image, the vision software processes and analyses the information to issue a pass-fail response based on fill-level of the bottle. More sophistication in the system can be added to detect the type of fluid in the bottle, weight of the fluid, quality of labelling, guidance of bottle on different routes etc. Machine vision systems have also been expanded for objective measurements e.g. dimensional measurements or guiding robots to align parts in an assembly operation.
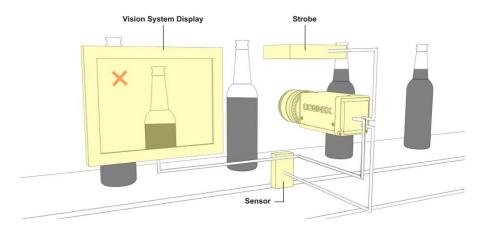
Figure 7: Bottle Fill-level Inspection Using Vision System [Cognex, 2020]

Machine vision is an essential element of manufacturing automation and its application domain is expanding further due to the evolution of Industry 4.0. The vision systems are designed for conducting inspection activity which is one of the crucial parts directly related to productivity, profitability and competitiveness of organization. However, penetration of machine vision in Indian MSME sector is very low due to legacy practices involving manual workers performing inspection tasks. The majority of MSMEs across the globe face challenges related to lack of skilled workforce and reduced labour productivity. These problems are

aggravated in developed countries due to aging of population and lower birth rates while India is facing a crisis linked with labour migration accompanied by shortage of skilled workforce. Therefore, it is essential to improve productivity through automation and staying competitive by adopting technological trends delivering smartness. The MSMEs are positioned at a lower layer of the supply chain in manufacturing with labour-intensive approaches orientated towards short-term operations rather than long-term strategic issues. The industrial policies for implementing smart factories are encouraged at policy level by government ministries but little research has been carried out on challenges for MSMEs in the participating Industry 4.0 ecosystem. It is realized that the MSMEs lack motivation for adopting and implementing emerging technologies due to lack requisite specialist and scepticism towards the technology. Table 1 summarizes some of the strategic goals essential for MSMEs that can be ideally met by employing machine vision systems.

The TIH at IIT Jodhpur aims at bridging this gap by providing much needed motivation and boost to MSMEs by developing technological solutions linked with vision applications in the smart factory context, skilling of manpower, establishing start up ecosystem to support innovation in the domain of vision based system, refresher and short courses for awareness in the employee, collaboration with foreign experts in the domain for know-how exchange, etc. The TIH plans to be a state-of-the-art facility in the country to conduct research in the domain of machine vision and Industry 4.0 by developing solutions for smart factories and Industrial IoT relevant to MSMEs. The TIH aims at developing vision based solutions to assist manufacturing industries in achieving five broad requirements; guidance, identification, gauging, inspection and prognosis. Figure 8 shows key points involved in the development of a typical machine vision system.

Table 1: Strategic Goals of MSMEs and Machine Vision

Strategic Goal of MSME	How Machine Vision can help?
Improved Product Quality	Inspection, Gauging, Measurement, Assembly verification
Improved Manufacturing Productivity	Replacement of manual tasks with machine vision applications
Improved Product flexibility	Programmable system for measuring, gauging and identification
Lesser set up time and inspection down time	Programmable system and possibility of computer integration
Comprehensive information and process control	Recording of the information on computer
Lower capital equipment cost	Less possibility of obsolescence due to software upgrades for capability
Lower production cost	Single vision system against many human operators, Early detection of defects
Reduction of scrap rates	Early detection of defects
Integration with inventory control	Integration with optical character recognition (OCR) and identification
Reduces floor space	Single system for multiple operations
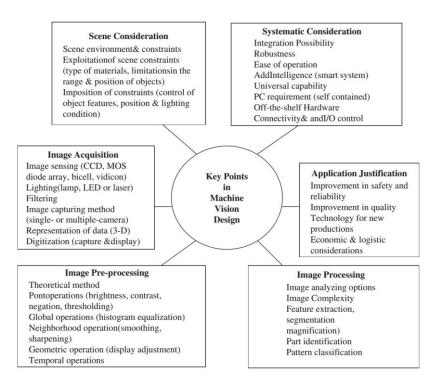
Figure 8: Key Points in Design and Application of Machine Vision

The implementation of vision based guidance aims to focus on developing solutions for MSMEs involved in the assembly work and related activities. The solutions will focus on automated location (position and orientation of a part) determination, comparison with specified tolerance level, and ensuring correct assembly. The information generated from guidance can be used to report the location and orientation of part in 2-D or 3D space with robot or machine controller. It is reported that vision-based guidance achieves far greater speed, accuracy and consistency than manual positioning in tasks such as arranging parts on or off pallets, placements of parts on and off the conveyor belt, part alignment in assembly or removing parts from bins [Liu et al., 2015]. The guidance can also be used for alignment with other machine tools as parts may be presented to the camera in unknown orientation during the production process. Figure 9 shows architecture for a typical vision-based guidance system proposed for assembly of narrow cabinet products [Liu et al, 2015]. The guidance systems can help in implementing robot-based manufacturing and assembly operations reducing human involvement in repetitive work. The TIH aims at developing vision-based manufacturing processes solutions (involving hardware and software) for autonomous applications e.g. in-situ repair or hazardous environments.
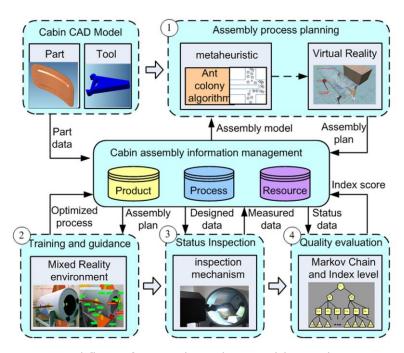
Figure 9: Workflow of Vision-based Assembly Guidance System
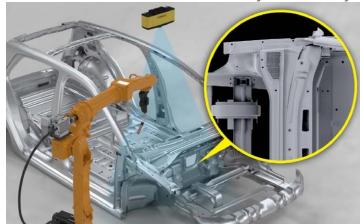
Figure 10: Vision-based Robotic Welding

A large number of MSMEs are involved in doing job work for OEMs and produce components at a higher production rate. It is necessary to manufacture products meeting global standards and requirements to be competitive in the market which necessitate checking of each component. The system is to be designed for binary decision making related to accept or reject based on template matching and does not require complex mathematical models to determine the dimensions. The system should be quick in data processing, robust, low-cost and scalable to adapt changes of product line. The TIH aims at the development of cost-effective modular solutions combining cloud manufacturing philosophy and vision-based systems. It is also envisaged to scale these solutions to a variety of manufacturing sectors such as automotive, footwear, textile, leather etc. by collaborating with academic and industrial partners.

Outcomes and Deliverables

The TIH vertical in the domain of Industry 4.0 and cloud manufacturing will plan and execute the activities to meet the following deliverables;

Basic and Translational Research: India needs strong fundamental research in the domain of Industry 4.0 and cloud manufacturing to generate significant technical output and creation of skilled manpower with adequate expertise. The domain requires subject knowledge of multiple domains ranging from manufacturing processes and systems, vision system algorithms and applications, artificial intelligence and machine learning, big data and data analytics, sensor and IoT implementation. The TIH will engage collaborators from these domains in meaningful discussions with industrial collaborators, international experts and MSMEs requiring technological solutions. The efforts will be made to conduct basic research for generating technical know-how and concept development. The activity involves engagement of research students to generate much required manpower with expertise for industrial employment generation. The concepts developed at laboratory scale will be fine-tuned through engaging discussions with industrial partners and MSMEs. The metric for quantification of deliverables will be in the form of peer-reviewed publications in international journals and conferences, patents and presentations in workshops.

Multidisciplinary Research: The domain requires expertise from multiple domains of engineering to solve problems that are of immediate needs to industries which cannot be solved with research in silos. The TIH will act as a vehicle in the domain for collaborative research combining expertise from within the institute, collaboration of multiple institutes and industry-academia collaboration with an objective to share expertise and resources to solve problems in the domain of Industry 4.0 and cloud manufacturing. The quantification measures will commence joint-projects in the domain that aims to develop innovative solutions adaptable by industries.

Application Domains: The objective of TIH is to facilitate activities in the diverse domains linked with Industry 4.0 and cloud manufacturing to develop industrially relevant technological solutions. The activities within TIH aims to conceptualize the development of hardware and software that can be deployed for smart factories. The TIH will act as a hub for the development activities with collaborators and user industries as participants. The TIH aims to develop technological solutions in the following broad domains linked with Industry 4.0 and cloud manufacturing;

- Vision-based technologies for GIGI (Guidance, Identification, Gauging and Inspection) for reduced human intervention and operational efficiencies
- Vision-based technologies for Manufacturing Prognosis
- Cloud-based framework for resource sharing in manufacturing industries
- Cloud-based assistance framework for assisting MSMEs in manufacturing sector

Curriculum Development: In spite of significant thrust on embedding concepts of automation and smartness in manufacturing industries, the educational curriculum does not contain adequate exposure to the same. The TIH will engage with collaborators in developing curriculum and courses for engineering graduates to provide adequate exposure of the domain. The TIH would also be conducting skilling programs to train industry personnel for the knowledge upgradation.

Entrepreneurship and start-ups: The TIH will devise a mechanism to encourage and support students, research scholars, interns, etc. to venture into entrepreneurship activities by formulation of tech start-ups. As TIH is aiming to develop technological solutions that are of long-term interest to industries by engaging in meaningful research, it is expected that many participants will be able to promote technology startups. The research ecosystem within TIH

will promote the "Make in India" initiative of the government through inclusive development within the nation.

Industrial collaboration: Another important feature of the TIH is to promote strong linkages between industry and academia through intensive collaborations in developing the solutions. This is an essential feature to conduct the research which is relevant to the market and benefitting the user industries. There is tremendous thrust to manufacturing activities to boost GDP growth and promote inclusive development. There are not many collaborative platforms in the country which will assist development activities in the domain of Industry 4.0 and cloud manufacturing at this stage. The TIH will act as an enabler for concepts in these domains in achieving long-term sustainability goals related to the development of smart factories.

Collaboration and Stakeholders

The TIH aims to collaborate with the following stakeholders from government:

- Ministry of Heavy Industries
- Ministry of Micro, Small and Medium Enterprises (MSME)
- Confederation of Indian Industries (CII)
- Ministry of Textile

Training and Outreach

The TIH aims to develop courses, curriculum and modules for diverse topics related to fundamentals as well as applications in the domain of Industry 4.0 and cloud manufacturing. A significant benefit is expected by creating an awareness about the implementation of technology on the shop floor. This will be attempted through development of awareness modules and success stories explaining the potential benefits derived from implementation of technology and embedding smartness into manufacturing activities. This will be achieved by active collaboration with government ministries and CII as these can attract large participation which is necessary for the awareness. The TIH would also plan to organize industry sponsored workshops and upskilling programs for the knowledge upgradation and training of existing employees within the organization.

References

Adamson G, Wang L, Holm M, Moore P, Cloud manufacturing - a critical review of recent development and future trends, International Journal of Computer Integrated Manufacturing, 30(4–5), 347–380, 2017.

Ari I, Muhtaroglu N, Design and implementation of a cloud computing service for finite element analysis, Advances in Engineering Software, 60-61, 122-135, 2013.

Baumann F W, Roller D, Additive manufacturing, cloud-based 3-D printing and associated services - overview, Journal of Manufacturing and Materials Processing, 1, 1-60, 2017

Caggiano A, Cloud-based manufacturing process monitoring for smart diagnostic services, International Journal of Computer Integrated Manufacturing, 31(7), 612-623, 2018.

Chen T, Strengthening the competitiveness and sustainability of a semiconductor manufacturer with cloud manufacturing, Sustainability (Switzerland), 6(1), 251–266, 2014.

CII, Report on Micro, Medium and Small Scale Industries, 2020, accessed from www.cii.in on 7-06-2020

Cognex, Introduction to Machine Vision – A Report, 2020

Czimmermann T, Ciuti G, Milazzo M, Chiurazzi M, Roccella S, Oddo C M, Dario P, Visual-based defect detection and classification approaches for industrial applications – a survey, Sensors, 20, 14591-25, 2020.

HSRC Report, Industry 4.0 Market Will Reach $214B by 2023, Washington DC, 2018 https://homelandsecurityresearch.com/industry-4-0-market-will-reach-214b-2023/ - accessed from on 7-06-2020

Hung M H, Lin Y C, Huang H C, Hsieh M H, Yang H C, Cheng F T, Development of an advanced manufacturing cloud for machine tool industry based on AVM technology, IEEE International Conference on Automation Science and Engineering, 2013.

Industrie 4.0 Working Group, Securing the Future of German Manufacturing Industry-Recommendations for Implementing the Strategic Initiative, Acatech, Munich, Germany, 2013

Jin X, Yu S, Zheng P, Liu Q, Xu X, Cloud-based approach for smart product personalization, Procedia CIRP, 72, 922-927, 2018.

KPMG, Report on Catalysing MSME Entrepreneurship in India, 2017, accessed from https://assets.kpmg/content/dam/kpmg/in/pdf/2017/07/Catalysing-MSME-Entrepreneurship-India.pdf on 7-06-2020

Kumaraja V, SME Automation – Key to growth in India, accessed from https://www.themachinemaker.com/manufacture/sme-automation-key-to- growth-in-india

Leitao P, Colombo A W, Karnouskos S, Industrial automation based on cyber-physical systems technologies: Prototype implementations and challenges, Computers in Industry, 81, 11–25, 2016.

Li B H, Zhang L, Ren L, Chai X D, Tao F, Typical characteristics, technologies and applications of cloud manufacturing, International Journal of Computer Integrated Manufacturing Systems, 18(7), 1345–1356, 2012.

Liu Y, Li S, Wang J, Zheng H, Lu J, A computer vision-based assistant system for assembly of narrow cabinet products, International Journal of Advanced Manufacturing Technology, 76, 281-293, 2015.

Malamas, E N, Petrakis E G M, Zervakis M, Petit L, Legat JD. A survey on industrial vision systems, applications and tools, Image and Vision Computing, 21(2), 171-188, 2003

Matt D T, Modrak, V, Zsifkovits H (Editors), Industry 4.0 for SMEs – Challenges, Opportunities and Requirements, Palgrave Macmillan, 2020.

Pal A, Dasgupta R, Saha A, Nandi B, Human-Like Sensing for Robotic Remote Inspection and Analytics, Wireless Personal Communications, 88(1), 23–38, 2016.

Papacharalampopoulos A, Stavridis J, Stavropoulos P, Chryssolouris G, Cloud-based control of thermal based manufacturing processes, Procedia CIRP, 55, 254-259, 2017.

Richert M, Xavier O, De Martino A, Imaging Polarimetry for the determination of stress constraints in transparent solids, The European Journal Physical Conference, 2010.

Tao F, Cheng Y, Zhang L, Nee, A Y C, Advanced Manufacturing Systems: Socialization Characteristics and Trends, International Journal of Intelligent Manufacturing, 28, 1079-1094, 2015

Wang J, Ma Y, Zhang L, Gao R, Wu D, Deep learning for smart manufacturing: methods and applications, Journal of Manufacturing Systems, 48-C, 144-156, 2018.

Wang L, Gao R, Ragai I, An integrated cyber-physical system for cloud manufacturing, Proceedings of ASME 2014 International Manufacturing Science and Engineering Conference, Michigan, USA, 2014.

Wang X V, Givehchi M, Wang L, Manufacturing system on cloud: a case study on cloud-based process planning, Procedia CIRP, 63, 39-45, 2017.

Warren R E, Cohn D B, Chemical detection on surfaces by hyperspectral imaging, Journal of Applied Remote Sensing, 11(1), 015012, 2017.

Zhang J, Starly B, Cai Y, Cohen P H, Lee Y S, Particle learning in online tool wear diagnosis and prognosis, Journal of Manufacturing Processes, 28(3), 457-563, 2017.

Zhong R Y, Xu X, Klotz E, Newman S T, Intelligent Manufacturing in the Context of Industry 4.0: A Review, Engineering, 3(5), 616–630, 2017.

Augmented & Virtual Reality

Collaborating Institutes

IIT Jodhpur

IIT Delhi

IIT Bombay

KLETech Hubli

DTU, Delhi

QUT, Australia

University of Toronto

Industry Partner: Vizara Technologies Pvt. Ltd.

Summary

Augmented Reality (AR) and Virtual Reality (VR) are at the forefront of interaction and immersive content delivery across a range of application domains like training and education; experiential culture, heritage, and preservation; physical and mental health care; computer-aided-design; remote collaboration; and games, movies, and social media. While VR aims to replace our sensory inputs from the real world with entirely their virtual representations, AR augments real world presence with virtual additions.
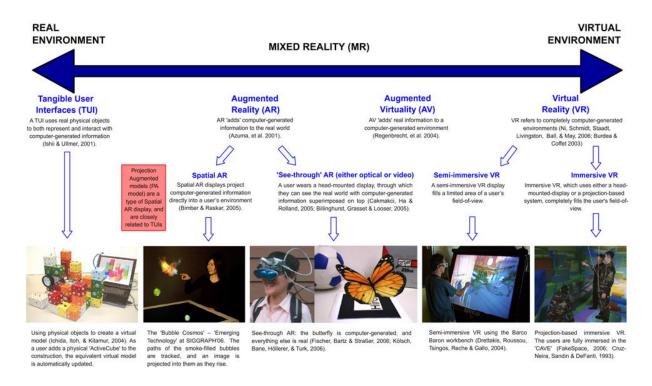
Figure 1. The Reality-Virtuality Continuum [Milgram and Kishino, 1994]. Source: https://en.wikipedia.org/wiki/File:Adapted_milgrams_VR-AR_continuum.png

Myriad mixtures of real-world and virtual-world have become popular in specific domains (see Figure 1) and terms like mixed-reality and eXtended-reality (XR) are gaining traction. VR space is seeing rapid world-wide expansion due, in part, to substantial ongoing advances in GPU-based computational capability, low-latency tracking and displays, efficient and realistic computer graphics, computer vision and machine learning algorithms. Hardware capabilities range from low-cost and consumer-grade to highly-specialized and medical-grade. These advances, and their massive potential in revolutionizing the quality of life, have transformed AR and VR from fancy gadgetry to hotbeds of research and development around the world.

This TIH aims to foster cutting edge research in AR and VR in India. Our goal is to become the hub for AR/VR technical know-how, AR/VR algorithm and system development, training and education in AR/VR technology, development of AR/VR applications in close collaboration with domain experts, and evolving metrics, standards and methodologies for design and evaluation of AR/VR-based applications and systems, and their utility.

Among the biggest hurdles in proliferation of VR/AR applications for situations that demand them is the development of appropriate virtual content. Currently, there are myriad hardware

devices that make up the AR and VR landscape. Targeting content to one of these is difficult enough. Making the content ubiquitously available is even harder. One of the core goals of this TIH is to research and develop frameworks that allow for rapid development, prototyping, and testing of AR/VR applications. We aim to make developing solutions involving AR/VR easier for everyone. This ranges from researching, testing and deploying current algorithms to interfacing with multiple hardware devices.

Figure 2: Source: Tracking haptic gloves in VR, Oculus Research, 2018.

AR/VR applications try to computationally mimic all sensory inputs including visual (using computer graphics), tactile (using haptics) and auditory (using sound rendering), and even olfactory in certain contexts. Research in these aspects is crucial to immersion and believability of VR. Additionally, AR requires accurate sensing, tracking and reconstruction of the real world using computer vision and other technologies. Figure 2 shows a potential solution being researched by a commercial VR research lab, where markers are used to track the hands and translate them to the movement of virtual hands displayed on a HMD. The gloves on the hand also relay haptic signals to mimic the sensation of touch according to how the virtual hands interact with the virtual object being shown. The TIH also aims to support a thriving collaborative consortium that looks into all these related aspects of AR/VR. We also aim to improve existing hardware devices with software middleware to make them more functional and user friendly.

Background

The Technology Innovation Hub at IIT Jodhpur is part of the National Mission on Interdisciplinary Cyber-Physical Systems focussing on Augmented and Virtual Reality.

Recently there has been a great interest in virtual and augmented reality, presumably seen as the potential technological breakthrough to compete with or even supersede smartphones as a dominant platform and interface accessing digital contents [Azuma, 2019]. It is therefore pertinent to give the required attention to generate competence and resources to undertake research and development activities in these areas. The proposed TIH on Augmented and Virtual Reality at IIT Jodhpur considers this as an opportunity at the right time to drive these activities through a consortium of experts in different institutes across the country. The TIH offers a vision to create a world-class facility in the area of augmented and virtual reality pursuing cutting edge research.

The present consortium of experts has many members who have worked in several collaborative projects in the past with successful completion. Two of these projects have had components on VR/AR related activities, which demonstrate adequate technical competence and ability among the partner institutes to undertake objectives envisioned by this TIH. Both these projects were coordinated and led by the Director IIT Jodhpur during his stay at IIT Delhi.

One of the projects "Programme for Autonomous Robotics (PAR)" funded by BRNS had five sub-projects. All sub-projects were successfully executed and deliverables were shared with BARC. Some which were deployed in their labs. One of the sub-projects "Immersive Environments for Teleoperation", dealt with developing data acquisition, processing and rendering of 3D immersive viewing of remote environments, and developing haptic interfaces for manipulation of objects at the remote side. Figure 3a shows a snapshot of the functioning of the interface.

In another project, called Indian Digital Heritage Project, we have designed and developed a virtual tour system for the Vittala Temple in Hampi, which can also be used for other heritage sites. Our system uses a unique technique to blend physical models with virtual models. This technique is then used to create a simple user interface, which offers a high degree of immersion without compromising on the richness of experience. It also allows multiple people to participate simultaneously. In our system (Figure 3b), we have created a tabletop replica of the entire temple complex by employing 3D printing technology. (More examples may be from Vizara).

Figure 3a: Haptic Teleoperation

Figure 3b: Mixed Reality Presentation

Figure 4 3D Printed Model with Holographic Projection

In terms of technology development, the members of the team have worked on fundamental computer graphics, geometry, image processing, computer vision, and machine learning algorithms. These have included techniques for interactive rendering, efficient model

representation and geometric computation, animation, haptic rendering, 3D reconstruction, text and language processing, simulation and visualization, user interfaces.

National Need and Impact

It is increasingly -- and painfully -- apparent that while physical distances may sometimes be mandated by geographical constraints or pandemics, the work of the world must go on. We must teach our children and train our workforce. We must be able to communicate, tell our stories, get access to essential life services, and to entertainment. We must find ways to recreate our social interactions remotely when needed. AR and VR are particularly unique in their ability to assist with and help achieve these goals. There is therefore a dire global need, and a sense of urgency, for rapid progress in this area, and many countries in the world are seriously exploring AR and VR to gain a strong foothold in the cyberworld. Of course, the immersion afforded by virtual reality has implications far beyond a 'lock-down' scenario.

India must invest in research in this domain to stay competitive and in step with the rest of the world. We cannot stay reliant on agencies from other countries for providing core technologies that have the potential to affect so many aspects of our lives. Therefore we must not only develop the core AR/VR technology stack suitably augmented by research in the related disciplines of computer graphics, computer vision, haptics and machine learning, we must also nurture and develop the many application domains that AR/VR can be applied to.

Education and training is an important application domain, which we must invest in to leverage and foster India's human resource supremacy. VR/AR have proven to be excellent tools for teaching skills where personal experience is critical to flatten the learning curve. Further, lesson retention is much improved with VR. Online teaching and training for students and the workforce is also crucial, because of its scalability. AR/VR have the demonstrated ability to make this more participative and interactive for the students. We must develop methods that are relevant to our internet backbone capability, and enhance our way of teaching, without demanding too much in terms of setup or equipment on both the trainee and the trainer sites.

Digital heritage is another such domain that has the potential to benefit immensely from AR/VR research. Indian cultural heritage enriches the lives of her citizens and it is also the world's window into India's history, philosophy, science and her people.

It has significant economic impact as it attracts millions of visitors every year from around the world. Foreign tourist arrivals (FTAs) in India have grown steadily over the past few years. The number of Foreign Tourist Arrivals (FTAs) in India during 2018 increased to 10.56 million as compared to 10.04 million in 2017. The growth rate in FTAs during 2018 over 2017 was 5.2% as compared to 14.0% during 2017 over 2016. In 2018, foreign exchange earnings (FEE) from tourism were US$ 28.59 billion as compared to US$ 27.31 billion in 2017, registering a growth of 4.7%. A large proportion of the jobs in the tourism sector are linked directly or indirectly to our cultural heritage.

Today, our heritage, both tangible and intangible, faces a very real risk of being eroded from our physical and mental landscape due to natural and manmade threats. Artefacts and monuments are being degraded and damaged, customs and practices are being lost, languages and craft forms are dying. Concerted efforts are needed to preserve, understand, strengthen and adapt Indian heritage in our modern times. Augmented and Virtual Reality can

support a broad scientific undertaking to reincarnate our heritage in the digital realm and better engage Cultural Heritage consumers using immersive and interactive interfaces.

Similarly, there is untold potential for improved outcomes in health and wellness with the help of VR. It can not only allow remote consultation that feels more personal, but also train health professional and emergency response teams. Further, VR has been successfully used for mental and physical health therapies (Streicher 2018, Boeldt et al. 2019, Riva et al. 2019), and even for meditation (Kosunen et al. 2016, Navarro-Haro et al., 2017). Due to the ability of VR to immerse a patient in severe life-like situations, a wide array of what-if situations, or managed serene environments, multiple studies have shown improved outcomes for pain management, phobia management, muscle retraining, etc.

Objectives

1. AR/VR Framework. We plan to develop a core framework for making research and development in AR and VR more accessible and efficient. The framework will also allow our applications to interface with and be deployed on various hardware and software platforms.

2. Core Technology. We want to foster research areas that are critical to AR and VR. All this research will contribute to the core framework. This research will enable content creation, manipulation and augmentation for AR/VR.

3. Applications. Although the choice of initial prototype applications and their detailed design necessarily needs to evolve with time, and depends on additional stakeholders who will inevitably connect with the hub, we feel a few core applications will be important to develop -- both because of their general usefulness and their coverage of the types of applications we seek to foster. These include

   ● Experiential AR/VR applications for cultural heritage.

   ● Immersive AR/VR applications for both synchronous and asynchronous teaching.

   ● AR/VR application for training primary healthcare professionals.

   ● 2D and 3D content creation and authoring applications for AR/VR.

   ● Presentation and storytelling applications in AR/VR.

   ● Digital twins or avatars for collaborative, remote social interactions.

   ● AR/VR assisted training for manufacturing and industry involving simulations of dynamic processes and operation of complex machinery.

   ● AR assisted indoor and outdoor navigation.

   ● AR/VR applications for retail.

   ● Assistive, rehabilitation and therapy applications using AR/VR.

4. Teaching and Outreach. One of our important goals is to target the development of human-resource in VR and AR technology and applications.

State of the Art

AR and VR technology is expected to totally change the way we connect with each other and take us a step closer to the 'digitally augmented' space. As per IDC predictions, the AR/VR market was enjoying about 60% of the total world spending in 2018. But it is expected to hit 85% of the total world spending by the end of 2021, with retail, transportation, manufacturing and public sectors on the top of the chart [Nancy Gupton et al., 2020]. Likewise, SMEs are expected to grow their involvement in the AR/VR market with a CAGR of 66.40% during the period 2018-2025.

AR and VR have been extensively used in gaming applications. They have also been used for applications as diverse as tele-collaboration, physical therapy, and mental health intervention. The critical ingredients of any AR/VR application include situation modeling, motion tracking (of hands, fingers, body, head, and eye), interaction techniques (including haptics, gestures, and devices), display devices (head mounted, wall, mobile), and computer graphics and computer vision techniques. All these areas have seen significant research, including by members of this TIH team. We mention below some of the ones more relevant to this TIH.
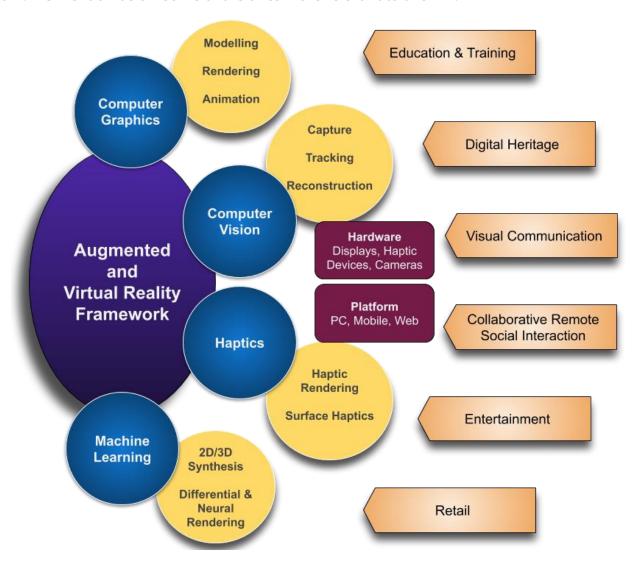
Figure 5. Research domains, challenges and application areas.

AR-VR Framework

In order to experience AR on mobile devices and the PC, one has to download and install the various software and applications to get into the augmented and virtual world. Now WebXR has come into existence [W3 WebXR, 2020]. As depicted from the name, this will facilitate users to interact with augmented and virtual reality devices in the webspace. It will enable them to get exposure to AR on Google Chrome, Mozilla Firefox, or other such browsers regardless of which device they run upon, resulting in lower barriers to its usage and higher overall experience. Additionally many AR frameworks like ARKit [Apple ARKit, 2020], ARCore [Google ARCore, 2020], Vuforia [Vuforia Engine, 2020] are now supported by mainstream real-time development engines like Unity [Unity, 2020] and Unreal Engine [Unreal Engine, 2020]. These can be deployed on a variety of hardware ranging from PCs to mobile devices to game consoles. VR frameworks like SteamVR [SteamVR, 2020] also allow the setup of VR systems, help in development of VR applications by providing relevant SDKs and run virtual marketplaces to sell VR apps.

Core Technology

[Sutherland, 1968] developed the first tethered head mounted display for VR. Since then the VR headset has shrunk in weight and size, while improving vastly in performance and fidelity as evidenced by the Oculus and Vive HMDs. Research in VR has ranged from talking issues of navigation and redirected walking [Nilsson et al., 2018] to near eye lightfield displays [Lanman and Luebke, 2013]. AR research has been bolstered by advances in tracking and SLAM [Mur-Artal and Tardos, 2017], 3D reconstruction [Agarwal et al. 2011] and devices like the Microsoft HoloLens.

Touching an object causes rich haptic cues that enable us to understand the object's physical properties and adeptly control the interaction. Although human experience centers on physical contact with tangible items, few computer systems provide the user with high-fidelity touch feedback, limiting their intuitiveness. Haptic interfaces are mechatronic systems that modulate the physical interaction between a human and his or her tangible surroundings.

[Adams and Hannaford, 1999] propose an artificial link between the haptic display and a virtual world, to include both the impedance and admittance models of haptic interaction. However, current haptic interfaces provide only single finger interaction with the virtual environments. For realistic immersiveness in the AR and VR environment, a lot more needs to be done on the device design side. There has been some progress in this direction, for example, hand-held haptic controllers have been proposed for virtual reality systems in [Sun et al., 2018, 2019, Kwon et al., 2019]. These controllers mainly provide only vibration feedback, hence do not provide a complete realistic immersive experience in AR and VR systems. Advancement in the fields of surface haptics and mid-air haptics has promised the solution. The field of surface haptics deals with providing the "sense of touch" on the touch screen itself [Varder et al., 2018] while mid-air haptics focuses on providing the same experience in the air [Carter et al., 2013, Long et al., 2014].

Machine learning for VR and AR has improved the quality of images by allowing rapid on-the-fly antialiasing [Bako et al., 2017], foveated rendering [Kaplanyan et al., 2019], physics based character control for virtual avatars [Won et al., 2020] and clothed human body reconstruction from images [Huang et al., 2020].

Content creation for AR/VR has also been a long standing research problem. Models created for such purpose have to be both realistic for immersion and simple for rendering. That is why a popular technique in VR applications is to simply display pre-recorded or pre-rendered video. That is, however, usually not sufficient for immersion and interaction -- two components that are important in many AR/VR applications.

Modeling techniques range from ab-initio design, to assembly from a library of models, to 3D reconstruction of actual locations. Image based modeling and rendering have also been used. Automated content creation has also been reported, mainly in non-VR contexts. Such schemes include procedural model generation [Getto et al., 2020], sketch-based recognition and creation of models [Bhattacharjee and Chaudhuri, 2020], and language or text based guidance for selection of models [Seversky & Yin 2006, Fremont et al. 2019]. Many of these techniques focus primarily on the design of language and its translation. There remain significant gaps in the need to efficiently determine what the content itself should be. To this end, a recommender based strategy is promising [Savva et al., 2017].

For a broad-based, comprehensive, and practical scheme, these various techniques need to be combined and augmented with behavioral and semantic models. There also is a need to incrementally populate, modify and enhance scenes with new or matching content, with different properties for different needs. Finally, there is a need to improve visual interfaces for such design -- possibly a VR based interface itself.

Application Areas

There are many applications of AR and VR. Since its appearance, VR has been used in different fields, as for gaming, military training , architectural design, education, learning and social skills training, simulations of surgical procedures assistance to the elderly or psychological treatments. A recent and extensive review of [Slater and Sanchez-Vives, 2016] reported the main VR applications, including weakness and advantages, in several research areas, such as science, education, training and physical training. Furthermore, another by [Freeman et al., 2017] focused on VR in mental health, showing the efficacy of VR in assessing and treating different psychological disorders as anxiety, schizophrenia, depression, and eating disorders.

Although AR is a more recent technology than VR, it has been investigated and used in several research areas such as architecture, maintenance, entertainment, education, medicine, and psychological treatments. More in detail, in education several AR applications have been developed in the last few years showing the positive effects of this technology in supporting learning, such as an increase on content understanding and memory preservation, as well as on learning motivation. For example, [Ibáñez et al., 2014] developed a AR application on electromagnetism concepts' learning, in which students could use AR batteries, magnets, cables on real superficies, and the system gave a real-time feedback to students about the correctness of the performance, improving in this way the academic success and motivation). AR systems allow the possibility to learn how to visualize and act on composite phenomena that traditionally students study theoretically, without the possibility to see and test them in the real world.

Efforts were also made by researchers to create AR/VR interfaces to rejuvenate digital heritage. The Virtual Angkor project aims to recreate the sprawling Cambodian metropolis of Angkor at the height of the Khmer Empire's power and influence in the twelfth century. Collaborating across disciplines and technologies, it has been built from the ground up by a team of Virtual

History Specialists, Archaeologists and Historians at Monash University in Melbourne, Flinders University in Adelaide and The University of Texas at Austin across a period of more than ten years. Indian Digital Heritage is a unique initiative by DST (Department of Science and Technology) supporting collaborative projects between researchers in the areas of Technology and Humanities for digital documentation and interpretation of our tangible and intangible heritage. The project highlights art, architecture, and cultural legacy of the world heritage site Hampi in Karnataka, the medieval capital of the Vijayanagara dynasty. IIndian Digital Heritage (IDH) research project helped generate technologies required for digital preservation and dissemination of the elements of digital heritage. The first two research translational projects completed under the Indian Heritage in Digital Space scheme of the DST, were showcased at the National Museum, New Delhi during the 1st International Heritage Symposium and Exhibition (IHSE) 2020. These include some of the following:

3D Virtual Models of Monuments and Sites

Digital models of monuments from laser-scanned data (point clouds) were created, using CAD modeling and mesh generation. Using 3D graphics, visualization and modeling, realistic virtual replicas of monuments are rendered and used as assets in virtual walk-throughs and VR experiences.

3D Digital Reconstruction of Damaged Parts

A collaborative approach for conjectural reconstruction of broken parts of monument structures, sculptures and artefacts, was used, wherein historians and conservation architects provide inputs to modelling teams and fabrication engineers. Digital merging of these conjectural reconstructions with existing structures provides validations for physical conservation efforts.

3D printed true-scale Monument replicas

3D printing was done to reproduce true-scale replicas of monuments and artefacts. These replicas can be reproduced at any scale or size – from a room-scale model of a monument site to a table-top reproduction of a structure to a hand-held small replica - with perfect scaled-down dimensions.

Interactive Museum Installations

Digital heritage museum installations including kiosks with projection systems, vision equipment, interaction tools, 3D printed sites offering mixed-reality e-heritage experiences such as virtual tours, projection mapping shows, gaming interfaces and holograms.

VR Heritage experiences

VR experiences of many heritage sites including Hampi temples, Kashi Vishwanath temple, Taj Mahal and Sabarmati Ashram, which include walk-throughs with a gaming interface in VR headsets such as HTC Vive, as well for Samsung gear and Google cardboard. The content includes real-life 360 VR videos and computer generated animations with true-scale 3D models.

AR Apps for Museum and On-site tours

Working with 3D content of the monument sites, mobile AR apps for on-site and museum tours are created. A tourist can use the app to point her phone towards an artefact or structure and

get information and trivia about it. She can also see how a damaged structure may have looked in the past, and also other similar pieces of art and architecture of the same style.

Basic Research Challenges

1. AR/VR Framework

This framework will implement an AR and VR technology stack and connect to all the relevant specialization areas that AR and VR build upon. This will set up transparent and standardized interconnects between various algorithms, devices and data, and make it available to researchers as well as industry-based developers of applications. The underlying framework will also include a categorization of VR applications, and guidelines and standardisation of hardware setup and software development of applications of various categories. The main challenge in this here is to create a unified framework that is able to adapt to different VR contexts, interaction strategies, hardware, number of users, scale, etc. This framework will incorporate the core technology we will develop at this TIH. It will also include a ground-up and well-integrated module for evaluation of applications developed using the frame-work. This would involve developing a mechanism for specifying objective functions for an application and measuring users' involvement, engagement, and improvement with respect to these objectives.

2. Core Technology

Research challenges in the core technology areas of AR/VR span across a range of specialized domains and aid in content creation, manipulation and augmentation. Some of these research challenges are outlined below:

1. ***Capture of real world environments using high fidelity laser scanners, images and videos.***

Laser Scanning allows enhancing the design process, speeds up and reduces data collection errors, saves time and money, and thus makes it an attractive alternative to traditional data collection techniques. It produces high-density point clouds or polygon meshes which can be inspected directly to a CAD model, geometry or manufacturing drawings. However, there are two major issues with laser scanning of real world sites

- Point Cloud data is very large to manage and process.
- No standards for laser scanning of heritage sites are defined.

These guide us to the following research problems to solve:

a. Defining standards of the extent, scope and data formats for data acquisition: These are required to avoid incomplete and erroneous data acquisition. For e.g. the use of the laser scanned data should be defined prior to doing the scanning. There should be focus on capturing each and every detail of a site or a monument if immersive, realistic walk-throughs are expected as output.

b. Development of automatic tools for point cloud and 3D mesh data processing. Cleaning up the 3D point-cloud data and cutting it into smaller parts to enable processing is done manually and is a tedious and time-consuming task. Point cloud data processing tools

which can use data analytics and artificial intelligence to provide auto-cleaning and optimal divisions of the huge point cloud heritage data are needed.

2. Creation of replica and novel geometry

The pipeline from Point-cloud data to 3D printing files involves complex meshing and sculpting processes, which can be automated. Since monuments consist of reliefs, sculptures and geometric intricacies, a tool is required for automation of the enhancement of sculptures and geometric details. This is currently done manually by 3D artists. On the other hand, generating a full-fidelity replica of real scenes is not always required. This is particularly true in applications to train emergency response professionals, or clearing operations for navigation of hazardous scenes. In these cases, the structure of the virtual scene is important, but appearance might not be. In other situations, appearance may be useful for immersion, but geometry may be simplified.

We will investigate such application-sensitive modeling, where the axes of realism of gross geometry, geometric detail, and appearance can be explored to evaluate and specify the requirements of an application using a grammar, coupled with an automatic or semi-automatic generation of a virtual scene that satisfies that grammar. This model will be generated starting with images, video, point-clouds, or CAD models of the scene. Among the research objectives are to:

a. Generate parameterized CAD models from images, videos and point clouds

b. Extend single-view reconstruction techniques to determine the appearance of the surfaces in these CAD models

c. Create editing primitives to modify an existing CAD model to generate a family of models, for similar or different applications

To this end, we will devise an extensible shape and behavioral model, so that a designer can complete content creation in phases. For example, one can assign properties like heavy or light furniture, portraits that stay on walls, foliage and its reaction to force, roads and their slickness, buildings with or without interior, rooms with entrance and hallways, Objects with augmented visualization properties, or their interaction with other objects in the scene. Embedding such meta-level information, some specified some pre-computed, in a semantic graph would also help improve rendering time algorithms (like collision detection, visibility computation, global illumination, audio rendering, haptic response, etc.) Furthermore, suggestions for these components can be generated by a learning based system based on the current context and history. The designer would then be able to incrementally add components in a plug-and-play fashion.

3. Creation and animation of avatars to populate virtual worlds.

Rapid creation of Virtual avatars of real users are needed to inhabit worlds. These need to be created from various capture modalities like images/video and motion capture data. The problem is especially challenging in the presence of deformable objects like clothing and complex movement.

4. Acquisition of Intangible and Ecological Heritage.

Reliable digitisation of intangible heritage like performing arts requires acquisition of 3D/ 360 degree multimedia data. Design of appropriate acquisition architecture is a challenge as there exists a need to capture data at multiple levels of detail simultaneously and dynamically to

document nuances of a particular art form. Further, documentation of background knowledge adds another level of complexity to the problem. Establishment of standard protocol for acquisition of digital data of intangible heritage is a challenging but interesting problem to address in the Indian context.

Ecological heritage digitization would imply creation of digital documentation of the ecosystem to the level of detail at which we can digitally simulate the dynamics of the environment at least with certain assumptions. This will also imply capturing of geo-physical details, complete surround video and, if possible hyperspectral data, soundscape data, flora and fauna. Representation of multimodal data in a linked framework such that data can be exploited for experiential virtual tour as well as simulation of the ecosystem is a research challenge.

5. Coherent rendering for augmented reality.

Synthetic objects that are augmented into the real world need to react to real world illumination conditions. This forms the basis for the problem of coherent rendering. We want to investigate real-time and offline algorithms for determining real world illumination to allow photoreal content augmentations to happen.

6. Physics-based simulations of real world phenomena.

The virtual world needs to replicate many real world processes to be immersive and believable. Accurate and efficient simulation and rendering of real world phenomena like fluid flow, clothing, fire and explosion hazards, soft and rigid objects collisions, anatomically based modeling of human movement are all complex research challenges that can enrich VR and AR content. We want to investigate many of these algorithms as required by our applications.

7. Haptics Interaction and Interfaces.

Haptic interfaces will typically involve mechanical, electrical, and computational layers that work together to sense user motions or forces, quickly process these inputs with other information, and physically respond by actuating elements of the user's surroundings. Well-designed tactile feedback can greatly increase the realism of virtual worlds. We shall work on building realistic haptic interfaces for VR applications. A novel multi-model AR and VR interface will be built for interaction with virtual or remote objects. Mid-air haptics focuses on providing the experience of touch in the air. This technology may be used in conjunction with a holographic projection for better immersive experiences. In surface and mid-air haptics, our skin is directly used for interaction with the virtual objects, which is in contrast with the existing haptic interfaces. Both fields have a lot of potential to contribute to AR and VR environments.

8. Machine Learning for AR and VR.

Deep learning has shown great promise in the understanding of visual data. Current research is directed toward synthesis of 3D and 2D models, images and video from neural networks. These naturally lend themselves to content generation for AR and VR. Various ML techniques have also been used to clean up visual data, to aid in tracking and reconstruction and even to parse complex scenes. Hence the TIH will invest in cutting edge research in machine learning applied to understanding, processing and synthesis of multi-sensory data with special emphasis on visual data.

Computational displays are aimed at bridging the gap between synthesized images and physical reality through the joint design of optics and rendering algorithms, as informed by our

ever-expanding knowledge of the human visual system. Today's head-mounted displays (HMDs) present a means to more closely approach this goal than prior direct-view displays, depicting accurate perspective, shading, binocular, and motion parallax depth cues. End-to-end convolutional neural networks (CNN) and generative adversarial networks can be designed to efficiently solve the full range of computational tasks for emerging near-eye displays. These networks synthesize defocus blur, focal stacks, multilayer decompositions, and multiview imagery—the critical inputs required for all major variants of varifocal, multifocal, and light field displays.

9. *Managing Data for AR/VR.*

Often data in digital space needs to be 3D. Some challenges in creating and managing this data involve transition of objects from real to virtual, interplay of geometry and semantics for AR/VR, organization of Large 3D/2D content along with metadata, handling 3D Data in AR/VR over the internet and mobile devices and presenting this data and content to the general public.

Application Areas

a. Digital Heritage

   i. Creating a virtual environment for Heritage-like domains requires different design approaches for meeting requirements of education, exhibition enhancement, exploration, reconstruction, and virtual museums. Our focus will be on education (enabling users to learn the historical aspects of tangible and intangible CH) and exploring historical and current views of CHs to discover, interpret, and acquire new insight and knowledge.

   ii. We shall explore use of AR for virtual restoration of museum artifacts as well as built structures for experiential rendering for the end user. As an example, we shall first look at building an AR interface to construct value added visualisation of manuscripts. This can be extended to sculptures, paintings and other CH artifacts.

Cultural heritage (CH) can use AR to provide enriched, interactive and digitally managed information. Virtual restoration of historic sites can be done by using 3D model reconstruction presented on in situ by using AR. We propose to present a mobile augmented 3D visualization system that captures the geographic information when placed in front of any site of historic importance. The 3D reconstructed models are registered with the geographic information to obtain geo-registered point clouds for learning about monuments history and purpose.

Virtual tourism also benefits from immersive walkthroughs of heritage sites. We use VR and AR to create walkthroughs accompanied by stories about events and characters related to those sites. Here storytelling in VR is also a promising research challenge. We hope to address related research challenges like how to develop authoring tools to tell stories in VR, study effects of long term immersion in VR environments and their implications on storytelling, and develop new rendering, modeling, and animation tools to better render the real world in VR.

Some of the interesting add-ons which can be provided through VR/AR in the context of tourism are: 3D Conjectural reconstructions & digital restorations of artefacts in real time which can add value to tourists' experience of the exhibit; AR based techniques to respond to the tourist's context with reference to an exhibit; storytelling through holographic visual avatars

A recurring problem in the tourism industry has been maintenance of heritage & cultural sites in tandem with managing millions of visitors per year. Need of the hour would be to reduce maintenance costs & loss of history to damage while maintaining or increasing interest in the site. interactive user-curated or virtually guided tours on 3D printed replicas with real-time AR content and multi-user experience is a possible alternative to work on..

b. Education and Training

The immersive nature of VR will help teachers in engaging students with learning material in various new and exciting ways. It will improve the chances of teaching distance learners in VR classrooms; preventing them from missing out on the benefits of learning in a collaborative environment. Whereas, AR will add flexibility to the on-the-job learning experience. It will offer them an opportunity to get access to the information required in real-time; enabling them to polish their skills and do their job simultaneously.

A large part of engineering, computer science, and even medical education involves understanding of processes and procedures. This may be in the form of the visualization of a machine, animation of an algorithm, or exploration of a system. A visual representation of these is well regarded as a potent learning tool. We will investigate how well virtual immersion adds to the learning in this context. Furthermore, in line with our theme to simplify the building of virtual worlds, we will develop techniques to generate 3D models and animation from an organized and annotated text, flow-chart, or a high-level visual language. These techniques can be further expanded in the future to create immersive environments from the plain-text description of system, processes, and algorithms. Thus the project will entail the development of mechanisms, frameworks, and AR/VR simulations for evaluation of their potential in education that may be online or imparted live at home.

One of the best examples towards this includes Case Western Reserve University (CWRU), which has helped all their medical students learn through home. The university is using a Microsoft Hololens to teach human anatomy. Introducing such technology in our nation will improvise the skill sets of every individual to a greater extent. Lessons in History uses AR to help rewrite history books in the classroom and inspire the leaders of tomorrow by featuring stories of powerful women. Through your smartphone we can scan over any portrait of any person in the school textbook and the history associated with the portrait will be narrated making learning in reality a new paradigm [World Economic Forum, 20th April 2020].

Training is a pervasive requirement in any skill based business. This includes up-skilling of professionals for a new job, or on-the-job training. It is not surprising that VR is being recognized as a critical tool for training and teaching skills. The problem is that there may be as many scenarios as there are jobs. These scenarios are also often quite comprehensive and elaborate. Generating a new experiential training application for each scenario is prohibitively expensive and counter-productive. We plan to develop the technology for quickly creating the training application for a given scenario. This requires two important steps: how to create a specification for any scenario, and how to translate the specification to an application. We will develop visual or textual grammar for such specifications. This will be followed by automatic asset generation schemes for quick rig-up of training application along with variability and what-if scenarios.

Also, mixed reality applications are commonly necessary for many scenarios. For example, a real tool may need to be held, or force required may be more than what hardware can sustain.

In such situations, a mixed reality application is called for. Determining the right mix of real objects and tools and virtual scenery for any given application is itself a tough challenge. We will develop guidelines and metrics for determining the basic VR/AR mix of different classes of applications.

The operators who will work in a virtual environment before operating the physical equipment, this shall circumvent damage to the equipment to a great extent. The operators are geared to avoid hazardous situations and possible safety violations on the physical equipment. The service personnel who work on large equipment are prone to accidents or possible damage to equipment, training the service personnel in VR labs on the same equipment shall reduce the possibility of damage and accidents. The access to expensive equipment in the VR environment can be made available to a larger number of people even in small cities. This will lead to democratization of skill development.

c.   Visual Communication

Visual communication includes advertising, storytelling, presentations and films. All these benefit from the immersive visualizations made possible by VR. The presented material can not only surround the user but also let them interact with elements which were earlier only meant for passive consumption. A particular example of this are interactive digital signage and kiosks. We plan to develop enhanced interaction technologies like mid-air haptic interfaces for this purpose.

Mid-air haptic interfaces that couple computer vision and affective computing with haptic feedback through the use of focused ultrasound. Mid-air haptic interfaces can aid in defining a gesture set, affect the emotional state of users, significantly increase user engagement with the interactive content, increase brand engagement, dwell time and brand recall while also enhancing the overall user experience. In this project we shall address challenges in the design and implementation associated with the next generation of mid-air haptic interfaces for digital signage and kiosks. Such public displays will not face issues with hygiene and cleaning requirements, and would be robust against extended use and potential damage.

d.   Collaborative and Remote Social Interaction

Terms like remote collaboration and tele-presence have been increasingly inserting themselves in the business lexicon, for they can save both time and money spent on travel. Remote collaboration is also potentially available 24x7 at short notice. Further, Covid-19 may have forever changed the dynamics on travel and physical interaction. On the other hand, humans cannot do without social interaction, be it in informal and for pleasure, or for business. Virtual interaction may be the answer, going forward. There has been significant advances in tele-presence in recent times, but the technology is not yet a replacement for physical interaction. We need to focus on creating more responsive avatars -- representing, capturing, and transmitting higher level information to overcome bandwidth and latency constraints. Incremental motion detection and isolation techniques will help derive this high level information.

Beyond technological advances, research is also required to determine how people interact in virtual settings, and how effective such collaboration is. In this context, the role of such tele-presence in remote teaching has not been explored. Given the current situation, we will need to increasingly rely on remote teaching and instruction. We need to develop experiments to

isolate, observe, measure, and learn different aspects of designing virtual tele-collaboration in different application settings. Such research can also be re-targeted to understand, intervene or highlight social mal-interaction, like bullying or abuse.

e. Assistive Technology

VR offers a great potential for individuals with different abilities. There are already attempts using this for developing social skills for individuals with autism. VR provides a safe environment to develop and practice a skill. The challenge here is to generate a realistic VR environment that relates to the real environment so that training in a VR environment actually benefits the individual in the real world. Further the current VR tools may not be fully accessible and compatible with people having different abilities. This may require to develop necessary software tools and hardware adjustments to expand the accessibility of the present VR tools.

f. Entertainment and Sports

Computer games and visual special effects are increasingly relying on advanced computer graphics and vision for enhanced interaction and photorealism. VR is the perfect medium for research in these domains as it not only offers immersion into these virtual worlds, it also makes possible simultaneous interaction with this content. Also there is an ever growing market now for games and movies in VR. For this content has to be developed not only for VR, but also in VR. At present there are very few frameworks or authoring environments that allow creation of such content. AR is also widely being deployed for hybrid games that merge synthetic elements into the real world. The most well known example of this is Pokemon Go which became an instant worldwide hit. We plan to develop capability to easily create and deploy such content.

Virtual cinematography is becoming common in digital filmmaking where the entire filmmaking process is deployed and carried out while being immersed in a virtual environment. This allows pre-planning of complex shots where rendered effects and real world actors can inhabit the same space in real-time. This allows the saving of large costs involved in filmmaking and will be very beneficial to our mutli-faceted film industry.

Broadcasts of sporting events have long benefited from AR whether to show the line of a cricket ball delivery to decide a crucial LBW decision or to visualize the offside line in a soccer match to decide on a contested goal. These capabilities are only being enhanced each day as capture and rendering technologies improve. We also plan to include enhanced bio-mechanical simulations of sportspersons in VR so that training decisions can be made wisely. Pose and form can be studied more closely in an immersive environment leading to better performance and therapy for injured athletes and sportspersons.

g. Retail

Virtual try-on for garment retail is an immediate application of AR/VR to retail where a customer can virtually try any item of clothing and see how they look in it. Computer graphics and simulation methods have improved to an extent where they can not only depict how the character will look in the garment but also give an idea of comfort and fit. Furniture retail also benefits from AR/VR where the consumer can place the item to be purchased in their existing environment to see how the new item fits in. Both these applications require developing improved capture, enhanced rendering and simulation capabilities.

h. Navigation

One of the emerging trends in AR/VR space is the use of AR for indoor and outdoor navigation. GPS directs us towards a shop, hospital, or other places. Still people have a difficult time contextualizing maps with their immediate surroundings. Augmented reality-powered navigation systems will assist people indoors and outdoors by displaying navigation directions on real world surfaces. This is easier to read and understand. They will instruct them on the path towards any particular product, software, or location in an engaging and effortless way; giving them an optimal experience.

The advent of smart city initiatives by governments and business, and technology advocates around the world are also leading the charge for increased AR/VR adoption and market growth. But even beyond smart cities, immersive technology appears poised to become a bigger part of day-to-day life for citizens around the world. The power to spatially visualize data that helps us navigate and connect better to our surroundings will be crucial to these cities of the future.

i. Manufacturing and Industry

The MSMEs of India need skill training more than anybody else. Our goal is to train machine operators in the VR environment before the MSME invests in expensive machines so that the operations can get a running start when the manufacturing line is in place. The efficiency improvements, safety drills and lean manufacturing activities can be verified and validated in the VR environment before implementing these improvements on the line. This saves time and money.

Product Verification and Validation plays a vital role in the product design and development cycle. Computer aided engineering (CAE) tools are used to conduct simulations and computer aided design (CAD) software tools are used to create 3D models, however these by themselves will not be sufficient to verify serviceability, interference or fouling between parts, manufacturability, dynamics and fluid interactions with the product. A VR environment would be the right solution to overcome the hurdles. However, the SMEs and MSMEs cannot afford to invest in a VR wall or a VR cave, we plan to address this gap with our VR setup and provide the visualisation services to customers.

j. Digital Twin

A digital twin is a virtual model of a process, product or service. This pairing of the virtual and physical worlds allows analysis of data and monitoring of systems to head off problems before they even occur, prevent downtime, develop new opportunities and even plan for the future by using simulations. Digital twin technology can be used to provide VR and AR based services through the mechanism of Technology as Service (TAAS) to the MSME sector.

In the context of the Industry 4.0 and of the digital factory, digital twin and virtual reality represent key technologies to design, simulate and optimize cyber-physical production systems and interact with it remotely or in a collaborative way. In order to satisfy these needs and to create a dynamic and immersive virtual environment, it is therefore necessary to combine the capacities of the digital twin to perform simulation of the production system with the capacities of the immersive virtual environment in terms of interactions. Digital twin will provide a data driven visual simulation of the entire physical and digital infrastructure, which continuously changes with the updates in the real system. TIH will work on developing digital twins for smart cities addressing specific problems of city governance.

Health care is one of the sectors which is going to benefit the most from the concept of digital twin. The emergence of smart wearable devices, organized storage of health data of individuals and societies, need for personalized and targeted medication and interweaving of engineering and medical disciplines are the major driving forces for the realm of smart and connected health. Explicit models of human organs can help us to understand, analyse and predict how disease progression impacts healthy organs. Visual depiction of the same through AR/VR technology provides an user-friendly interface to the medical professionals.

k. Telemedicine

With telemedicine becoming popular AR/VR interface for telemedicine is now becoming increasingly common. Through AR and use of remote manipulators doctors can examine a patient remotely. For example, mobile based sensors and specialized stethoscopes can provide useful information. AR interface can indicate the position of the sensors on the body of the patient on a mobile based image presented to a remote doctor. Doctors can as well instruct the patient to position sensors correctly. In a special situation, a remote robot can impersonate doctors and follow doctor's instruction for examination. In COVID-19 scenario, doctors are even exploring options of operating portable USG and X-ray machines through remote robotic interfaces.

A typical telemedicine system will consist of a human system interface connected to an operator, a communication medium, and the teleoperator/robot. The human system interface has input devices for position and orientation sensing, and output devices for displaying multiple modalities (i.e., vision, auditory, haptics, etc.). An operator is connected with the human system interface and commands the position/velocity of the teleoperator while observing the remote environment through the multi-modal feedback. The communication medium is used to transport the multimedia streams bidirectionally.
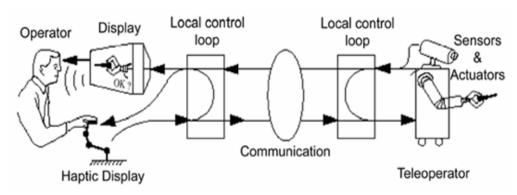
Figure: A Telepresence and Teleaction Scenario. Figure source: Hinterseer et. al 2008

There are various interesting and challenging research problems to realize realistic and immersive telesurgery systems, such as the good design of manipulators, kinesthetic and tactile sensing, data communication, delay compensating controllers, and immersion into the virtual workspace. The data communication module requires meeting quite a restrictive quality of service guarantee. In line with this vision, we, at IIT Jodhpur, will work on designing solutions to the above mentioned research activities under the proposed TIH.

Digital twin of a patient can be a ready reference for a doctor. All lab and radiological data are recorded on the digital twin with appropriate semantic model to characterize each patient.

Doctors make decisions based on detailed contextual analysis of the patient condition with reference to observations made over a telehealth platform.

Tele-consultation for psychological problems is also possible. VR based diagnostics have been designed for detecting mental health issues. Further, some of these problems can be treated with VR/AR interactive procedures. A doctor can remotely observe patient behavior during such therapy sessions and otherwise before taking decisions for further treatment.
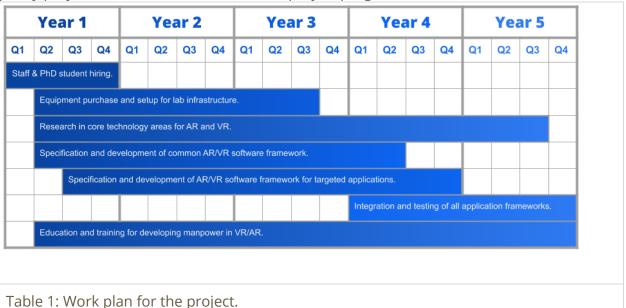
l.   Virtual Reality for Wargames

Simulation of military action between two or more hostile forces using the corresponding rules, data and procedures can be used to train commanders and their staff in planning and conduct of offensive and defensive operations in varied terrains.

VR/AR will combine man-in-the-loop aspects of live simulation with the actual operational battlespace of constructive simulation, to provide the most realistic possible training environment, Providing VR support in military simulations is an excellent method to improve accuracy of data and scenario authenticity.

Work Plan

A yearly work plan for the duration of the project is given in Table 1. Additionally, we will have yearly project review milestones to assess project progress.

	Year 1				Year 2				Year 3				Year 4				Year 5			
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
Staff & PhD student hiring.	■																			
Equipment purchase and setup for lab infrastructure.		■	■	■	■	■	■	■	■											
Research in core technology areas for AR and VR.		■	■	■	■	■	■	■	■	■	■	■	■	■	■	■				
Specification and development of common AR/VR software framework.		■	■	■	■	■	■	■	■	■	■									
Specification and development of AR/VR software framework for targeted applications.			■	■	■	■	■	■	■	■	■	■	■							
Integration and testing of all application frameworks.													■	■	■	■	■	■	■	■
Education and training for developing manpower in VR/AR.		■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■

Table 1: Work plan for the project.

Outcomes and Deliverables

- Foundational Research: One of the objectives of the project is knowledge creation in the areas of Augmented and Virtual Reality addressing their research challenges. This will in particular involve the core technology areas which cover several specialized domains such as creation of believable VR contents, visual and behavioral realism, seamless augmentation with multi sensory inputs. The deliverables will be in the form of publications in internationally reputed forums,

position papers in conferences and workshops. This may also lead to emergence of new open problems.

- Collaborative Research: The TIH will primarily rely on the collaborative research that brings together two and more academic units in the same institute, partnership of two or more institutes, industry academia cooperation with a goal to share knowledge and resources to solve specific problems in AR/VR. The deliverables will be starting new collaborative projects in particular themes, adopting an interdisciplinary approach combining strengths of different domains for innovative solutions. The collaborative research may be showcased through publications and demonstrations.

- Development of Applications: The TIH proposes a diverse range of applications which may be developed. Each application may require specific inputs and development of techniques that suit the best for the particular application. The proposed AR/VR framework will guide to decide the hardware set up and choice of software development environment for a particular application. This will also offer a set of tools and APIs that may be used for developing a new application. The TIH will create a repository of code base, libraries and software tools that can be made available for the collaborators, start-ups and other stakeholders for delivery of products

The TIH aims to develop frameworks for the following application areas

- Digital Heritage
- Remote Collaboration
- Education and Training
- Sports and Entertainment
- Rehabilitation & Therapeutics
- VR for Combat simulator
- Digital Twin Applications
- AR/VR for Telemedicine

The frameworks will materialize into concrete applications depending upon government or industry partners. The TIH may delve into developing frameworks for other applications if suitable partners are interested.

- Infusion in Curriculum: In the current scenario where the youth is introduced every other day with VR/AR contents through games and other entertainment medium, it is important to provide an appropriate academic discourse at different levels for the audience. This TIH will provide avenues to discuss and create appropriate courses in AR/VR for the curriculum at different levels from school to pre-doctoral level. This may help in capacity building in the area of AR/VR for the future.

- Entrepreneurship and start-ups: The TIH will encourage students and scholars to become entrepreneurs by creating start-ups. It is expected that activities in the various projects undertaken in TIH will provide avenues for initiating many start-ups. The parent institute of a student working in the project may provide extra support and assistance from their technology incubation cells. Some of the start-ups may be further formed as Section-8 companies. This will also promote the "make in India" motto of the nation.

- Industrial collaboration: One of the main emphasis of this hub will be to create a strong link between academia and industry. This is important to make the research more relevant. As there is an enormous interest in industry in AR/VR, the TIH will offer the right platform for collaborative activities both in the core technology and applications of AR/VR. The TIH will also provide technology consultancy and transfer to the relevant industries.

Collaboration and Stakeholders

Government Collaboration

- India Digital Heritage Project Partners
- BARC
- MInistry of Culture
- Ministry of External Affairs (MEA)
- ISSA, DRDO, Delhi

Application Problem of MEA

A very important application domain to be pursued is management of overseas Indian cultural assets using AR, VR technology supported by MEA. Key objectives of the programme :

1. to share the digital documentation of important indian cultural heritage with our partner countries

2. to enable Indians anywhere to enjoy indian cultural heritage abroad from the comfort of their homes

3. to share India's technological expertise with other countries and provide business opportunities for Indian firms.

The selection of cultural assets may be undertaken by MEA using an appropriate mechanism (high-level expert committee). After selection of projects is done by MEA, the TIH would provide the scheme architecture for the execution of the project, which include the data capture (using laser technology), digital documentation, content and creative design of the user experience, and implementation of a technologically-appropriate platform. Suggested projects which may be taken up could be:

Cultural asset	Outcome of DDCH	Remarks
Vat Phou cultural landscape, Laos PDR	Provide digital documentation for the temple complex which can be co-owned /shared with Laos. Laos scientists may be trained in DDCH technologies Indian citizens may experience virtually the Hindu temple architecture Indian museums can use digital data to recreate the Vat phou complex accurately in India using 3d printing.	Project would be similar to the Vittala temple project in Hampi. Execution likely to take about 300-360 working days. Proven technology. After Japan (who has done this for Angkor Wat), India would be the next major player in this area.
Indian Hospice, Jerusalem	DDCH of this building will provide valuable technological cooperation for Indian academia with Israeli academia. Indian citizens can experience the 800 year old hospice building, while interacting with modern day people using this building.	Technologically easy to accomplish.
India Arrival Day (various LAC countries)	This will be DDCH of a valuable intangible cultural asset .i.e., the migration of indentured labour in the 19th century Valuable diaspora-connect.	This project will require more creative content design and recreation of historical assets like the ships which were used to travel, commemorative statues etc.
Ramayana Circuit	The DDCH of the entire Ramayana cultural landscape including paintings, architecture, artforms, and buildings spread out in south and south-east Asia.	This project will require a mix of technologies to capture the different types of cultural assets, as well as inter-agency cooperation to bring together the different stakeholders like M/o tourism, M/o Culture and various ramayana research institutes.

VR requirement of ISSA

VR/AR Requirements in Military Simulation Applications

i. A complete indigenous AR/VR solution is required for seamless 3D visualization with VR effects of battlefield scenarios prevailing in military simulation applications. The

proposed solution should provide all software applications / tools, 3D object models, data models and hardware equipments.

ii. There can be a VR Terminal corresponding to each user of military simulation application. The VR Terminal should generate realistic images, sounds and other sensations that simulate a user's physical presence in a virtual or imaginary environment. Such terminals will be used for lower echelon players and will have Head Mounted Displays (HMD), data gloves and other sensors for user position, orientation, distance & direction cue, touch etc. The viewing area for VR can be selected either by area marking on the map or selected as the person moves his head in a particular direction. This will help the user in better visualization, understanding and appreciation of the battlefield.

iii. A Unified Virtual Environment is also required for displaying output of all other VR terminals along with other displays. This environment will be for higher commanders and preferably, without HMD. These may be 3D projections seen through windscreen or onto the walls of the room.

iv. 3D Objects of soldiers, military weapons, equipments and vehicles are to be developed and provided as part of solution.

v. Terrain, weather and meteorological aspects are important part of such simulations. Terrain visualization models should include desert, semi-desert, plains and mountainous terrains. Also, all effects of weather, day-night, moon-phases, time-of-the-day and other relevant meteorological parameters should be incorporated.

vi. All data models to be provided, as part of solution, in a form that can be interfaced with other simulation processes.

vii. The required solution should be available for desktop, client-server as well as for 3-Tier web-based architecture.

viii.The solution should exploit latest hardware / software technologies like GPU-based shaders and efficient rendering algorithms.

ix. The solution should be indigenously developed, cost-effective and maintainable for future requirements/enhancements.

The concept for required solution is depicted in figure below. Simulation systems for training in naval domain and for analysis requires to provide the user a realistic depiction of the environment, along with the various systems and entities involved in a particular scenario being simulated. This helps in better visual appreciation, and contributes substantially in carrying out the training activities in an effective manner. It also provides a method to validate the models and the simulation.

We are looking forward to the development and integration of following VR and 3D capabilities in context of over the Sea and underwater Scenario Simulation:

UNIFIED VIRTUAL ENVIRONMENT

Military Simulation Application

VR/AR Terminals

VR/AR Terminals

VR/AR Terminals

Figure: Proposed incorporation of VR technology in Military Simulation Applications

3D/VR Requirements for Naval Scenario Simulation

Industrial collaboration

- Samsung
- TCS
- Qualcomm
- nVIDIA
- Vizara Technologies Pvt. Ltd.

Training and Outreach

We will develop short and long course modules on different topics related to VR and AR. In addition, we will develop measurement tools and metrics to assess students and trainees of VR technology, to evaluate the efficacy of VR applications. This measurement of efficacy would essentially involve ingraining into the VR/AR application itself, modules for evaluation and tracking of its usage, and recording of user responses. This would help assess the application as well as the user, particularly in the context of training based VR/AR applications.

Recently, AR or VR has been increasingly used in industry and research. This has led AR or VR to extend in the education and training space as well. The past class-centered learning with or without digital media being used in classroom education and MSME training would now be replaced to promote constructivist-oriented and student-centered learning experience with the impact of the AR/VR on institutes. Transformative learning using such technologies is definitely happening in the near future and is possible, but requires a good understanding of the potential design, technology and practitioner, and most importantly, by host institutions. Once the potential benefits are made evident via outreach, then interests by all stakeholders drive the labs towards training via learning platforms either in-person or online-training. The idea is to create opportunities for integrating art, science and technology, enhancing the ability to capture young people who are educated and motivated to learn and grow. The agenda would be to facilitate training to faculty, staff, undergraduate, postgraduate students and research scholars. Tap on the gaps in learning/industrial training and partner with industries to provide the virtual training simulations and focus on increasing the accessibility, affordability and

availability. On the outreach front, the aim is to operate more AR/VR centers, programs, and initiatives meeting business and community needs across the nation.

These labs would provide a space to students undertaking their projects involving AR and VR to use these labs for designing and testing their VR and AR applications. The VR and AR courses along with allied courses that are multi-disciplinary would be designed and incorporated in all the undergraduate, postgraduate and research scholarly curriculum in order to engage in the future of AR and VR to solve the community problems from these labs. Students from across the universities, institutes, colleges and schools use these labs to enhance knowledge about historical sites, machinery, and conduct their research using virtual and immersive environments. These labs would become a hub for industry tie-ups, entrepreneurs and corporate training.

The VR lab can also be a resource for training operators, service personnel and production workers, who do not have access to expensive machines as stated earlier. Training young engineers and students in VR technologies and their applications shall open new employment opportunities for them. Creating awareness among MSMEs about the benefits of VR, shall help the small businesses to utilize VR resources to solve their design, manufacturing and operational problems. Training technicians to service, repair and maintain VR equipment can open new avenues for skilled workers. By developing VR software indigenously, we can reduce the overall cost of VR implementation and thus make it affordable for MSMEs in India.

Conclusion

This TIH on Augmented and Virtual reality will fill an important gap in the nation's infrastructure, and will be an important tool in making India increasingly self-reliant in this space. It will foster an ecosystem for increased activity and improved outcome in many areas requiring training, be it for up-skilling and reskilling, for healthcare and emergency responders, for security and defence, or for basic education. The TIH will also augment and enhance virtual tourism in support of digital preservation and amplification of cultural heritage. With immersive tele-collaboration, the TIH will also help increase productivity in a variety of research and industrial domains, and many other economic activities.

References

- A New Generation Video Capture System, https://www.4dviews.com/.

- Adams R.J., and Hannaford, B. (1999). Stable haptic interaction with virtual environments. In IEEE Transactions on Robotics and Automation, vol. 15, no. 3, pp. 465-474, June 1999. [Adams and Hannaford, 1999]

- Agarwal, S., Furukawa, Y., Snavely, N., Simon, I., Curless, B., Seitz, S. M., & Szeliski, R. (2011). Building Rome in a day. *Communications of the ACM*, *54*(10), 105-112. [Agarwal et al., 2011]

- Apple Developer Augmented Reality SDK - ARKit (2020). https://developer.apple.com/augmented-reality/ [Apple ARKit, 2020].

- Azuma R, The road to ubiquitous consumer augmented reality systems (2019), Special Issue Article, Human Behavior and Emerging Technologies, 1:26–32 https://doi.org/10.1002/hbe2.113 [Azuma, 2019]

- Bako, S., Vogels, T., McWilliams, B., Meyer, M., Novák, J., Harvill, A., ... & Rousselle, F. (2017). Kernel-predicting convolutional networks for denoising Monte Carlo renderings. *ACM Transactions on Graphics*, *36*(4), 97-1. [Bako et al., 2017]

- Bermejo, C., & Hui, P. (2017). A survey on haptic technologies for mobile augmented reality. arXiv:1709.00698 [cs]. http://arxiv.org/abs/1709. 00698 arXiv: 1709.00698.

- Bhattacharjee, S. and Chaudhuri, P. (2020), "A Survey on Sketch Based Content Creation: from the Desktop to AR and VR", Eurographics STAR 2020. [Bhattacharjee and Chaudhuri, 2020].

- Carter, T., Seah, S.A., Long, B., Drinkwater, B., & Subramanian, S. (2013). UltraHaptics: multi-point mid-air haptic feedback for touch surfaces. In Proceedings of the 26th annual ACM symposium on User interface software and technology, 505-514.

- CHESS (Cultural Heritage Experiences through Socio-personal interactions and Storytelling) - http://www.chessexperience.eu/ [CHESS, 2016]

- Concannon, D., Flynn, R., & Murray N. (2019). A quality of experience evaluation system and research challenges for networked virtual reality-based teleoperation applications. In Proceedings of the 11th ACM Workshop on Immersive Mixed and Virtual Environment Systems (MMVE '19). ACM, New York, NY, USA, 10-12.

- D. TuckI. Kuksa, "Virtual Heritage Tours: Developing Interactive Narrative-Based Environments for Historical Sites", Second Joint International Conference on Interactive Digital Storytelling -ICIDS, 2009. [D.TuckI, 2009]

- D. Stricker, A. P., M. Zöllner, "In-Situ Visualization for Cultural Heritage Sites using Novel Augmented Reality Technologies", International Conference on Graphic Archeology and Cultural Heritage,. [Didier Stricker et al., 2009]

- Dou, M., Khamis, S., Degtyarev, Y., Davidson, P., Fanello, S., Kowdle, A., Escolano, S., Rhemann, C., Kim, D., Taylor, J., Kohli, P., Tankovich, V., Izadi, S. 2016. Fusion4D: Real-time Performance Capture of Challenging Scenes. ACM Trans. Graph. 35, 4, Article 114 (July 2016). [Dou M et al., 2016]

- Freeman, D., Reeve, S., Robinson, A., Ehlers, A., Clark, D., Spanlang, B., et al. (2017). Virtual reality in the assessment, understanding, and treatment of mental health disorders. Psychol. Med. 47, 2393–2400. [Freeman et al., 2017].

- Fremont, D. J., Dreossi, T., Ghosh, S., Yue, X., Sangiovanni-Vincentelli, A. L., & Seshia, S. A. (2019, June). Scenic: a language for scenario specification and scene generation. In *Proceedings of the 40th ACM SIGPLAN Conference on Programming Language Design and Implementation* (pp. 63-78).

- Getto, R., Kuijper, A., & Fellner, D. W. (2020). Automatic procedural model generation for 3D object variation. *The Visual Computer*, *36*(1), 53-70.

- Google Developer SDK for Augmented Reality - ARCore (2020). https://developers.google.com/ar [Google ARCore, 2020].

- Huang, Z. Xu, Y. Lassner, C. Li, Hao and Tung T. (2020), CVPR 2020. [Huang et al., 2020].

- Indian Tourism Statistics 2019, India in Business, Ministry of External Affairs. http://tourism.gov.in/sites/default/files/Other/India%20Tourism%20Statistics%202019.pdf

- ITACITUS, 2008. "Intelligent tourism and cultural information through ubiquitous services." http://www.itacitus.org. [ITACITUS, 2008]

- Ibáñez, M. B., Di Serio,Á., Villarán, D., and Kloos, C. D. (2014). Experimenting with electromagnetism using augmented reality: impact on flow student experience and educational effectiveness. Comput. Educ. 71, 1–13.

- Jong-Gil Han, Kyoung-Wook Park, Kyeong-Jin Ban, Eung-Kon Kim, "Cultural Heritage Sites Visualization System based on Outdoor Augmented Reality", AASRI Procedia, Volume 4, 2013. [Jong-Gil Han et al., 2013].

- Kaplanyan, A. S., Sochenov, A., Leimkühler, T., Okunev, M., Goodall, T., & Rufo, G. (2019). DeepFovea: neural reconstruction for foveated rendering and video compression using learned statistics of natural videos. *ACM Transactions on Graphics (TOG)*, *38*(6), 1-13. [Kaplanyan et al., 2019]

- Kwon, J.H., Ryu, S., Ko, G., Lee, J., Suh, K., & Nam, S. (2019). Design of a VR Prototyping System Based on a Haptic Controller. In IEEE 6th International Conference on Industrial Engineering and Applications (ICIEA), pp. 41-44.

- Lanman, D., & Luebke, D. (2013). Near-eye light field displays. *ACM Transactions on Graphics (TOG)*, *32*(6), 1-10. [Lanman and Luebke, 2013]

- Litts K.B., & Lewis E.W. (2019). Mobile Augmented Reality: Exploring a new genre of learning. GetMobile: Mobile Comp. and Comm. 22, 3, 5-9.

- Lin, M. (Ed.), Otaduy, M. (Ed.). (2008). Haptic Rendering. New York: A K Peters/CRC Press.

- Long, B., Seah, S.A., Carter, T., & Subramanian, S. (2014). Rendering volumetric haptic shapes in mid-air using ultrasound. ACM Trans. Graph. 33, 6, Article 181.

- Milgram, Paul; H. Takemura; A. Utsumi; F. Kishino (1994). Augmented Reality: A class of displays on the reality-virtuality continuum. In Proceedings of Telemanipulator and Telepresence Technologies. pp. 2351–34.

- Microsoft Mixed Reality Lab (2020), https://www.microsoft.com/en-us/mixed-reality/capture-studios,

- Mur-Artal, R., & Tardós, J. D. (2017). Orb-slam2: An open-source slam system for monocular, stereo, and rgb-d cameras. *IEEE Transactions on Robotics*, *33*(5), 1255-1262. [Mur-Artal and Tardos, 2017].

- Nancy Gupton. September 21, 2017, What's the difference between AR, VR, and MR?, The Franklin Institute. Updated by Patrick J. Kiger, January 6, 2020. [Nancy Gupton et al., 2020]

- Nomura, R., Unuma, Y., Komuro, T., Yamamoto, S., & Tsumura, N. (2017). Mobile augmented reality for providing perception of materials. In Proceedings of the 16th International Conference on Mobile and Ubiquitous Multimedia. ACM, 501-506.

- Nilsson, N. C., Peck, T., Bruder, G., Hodgson, E., Serafin, S., Whitton, M. & Rosenberg, E. S. (2018). 15 years of research on redirected walking in immersive virtual environments. *IEEE computer graphics and applications*, *38*(2), 44-56. [Nilsson et al., 2018].

- Perkins Coie LLP, 2020 Augmented and Virtual Reality Survey Report, Industry Insights into The Future of Immersive Technology, Volume 4, March 2020. [Perkins Coie LLP, 2020].

- Sadzak A, S.Rizvic, C.Dalton, A.Chalmers., "Information Perception in Virtual Heritage Storytelling Using Animated and Real Avatars", Spring Conference on Computer Graphics, 2007. [Sadzak et al., 2007]

- Salisbury J.K., & Srinivasan M.A. (1997). Phantom-based haptic interaction with virtual objects. In IEEE Computer Graphics and Applications, vol. 17, no. 5, pp. 6-10.

- Salisbury, J.K., Conti, F., & Barbagli, F. (2004).Haptic rendering: introductory concepts. In IEEE Computer Graphics and Applications, vol. 24, no. 2, pp. 24-32, 2004.

- Savva, M., Chang, A. X., & Agrawala, M. (2017). Scene Suggest: Context-driven 3D scene design. *arXiv preprint arXiv:1703.00061*.

- Selma Rizvic, Aida Sadzak, VedadHulusic, AmelaKarahasanovic, "Interactive Digital Storytelling in the Sarajevo Survival Tools Virtual Environment", ACM Digital Library, SCCG 2012. [Selma Rizvic et al., 2012].

- Seversky, L. M., & Yin, L. (2006, October). Real-time automatic 3D scene generation from natural language voice and text descriptions. In *Proceedings of the 14th ACM international conference on Multimedia* (pp. 61-64).

- Shrikant Srivastava, AR/VR Trends and Predictions for 2020 & Beyond, https://appinventiv.com/blog/ar-vr-trends, April 30, 2020. [Shrikant Srivastava, 2020]

- Slater, M., and Sanchez-Vives, M. V. (2016). Enhancing our lives with immersive virtual reality. Front. Robot. AI 3:74. [Slater and Sanchez-Vives, 2016].

- Sjenica S. Rizvic, I. Skalonjic, "Reconstructing Cultural Heritage Objects from Storytelling - Virtual Presentation of Sultan Murat's Fortress", in Proceedings of 2nd International Congress on Digital Heritage 2015. [Sjenica et al., 2015]

- Steam VR (2020). https://store.steampowered.com/steamvr, [SteamVR, 2020]

- Sun, Y., Yoshida, S., Narumi, T., & Hirose, M. (2018). Handheld Haptic Interface for Rendering Size, Shape, and Stiffness of Virtual Objects. In Proceedings of the 2018 ACM International Conference on Interactive Surfaces and Spaces, 411-414.

- Sun, Y., Yoshida, S., Narumi, T., & Hirose, M. (2019). PaCaPa: A Handheld VR Device for Rendering Size, Shape, and Stiffness of Virtual Objects in Tool-based Interactions. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. ACM, New York, NY, USA, Paper 452, 12 pages.

- Sutherland, I. E. (1968, December). A head-mounted three dimensional display. In *Proceedings of the December 9-11, 1968, part I* (pp. 757-764). [Sutherland, 1968].

- Unity Real-time 3D development platform (2020). https://unity.com/, [Unity, 2020].

- Unreal Engine - Real-time 3D development platform (2020). https://www.unrealengine.com/en-US/ [Unreal Engine, 2020].

- Vardar, Y., Güçlü, B., & and Basdogan, C. (2018). Tactile Masking by Electrovibration. In IEEE Transactions on Haptics, vol. 11, no. 4, pp. 623-635, 1 Oct.-Dec. 2018.

- Volumetric Capture Setup. https://www.dimensionstudio.co/.

- Vuforia Engine Augmented Reality Software Platform (2020), https://engine.vuforia.com/engine [Vuforia Engine, 2020].

- W3 Web XR Device API, https://www.w3.org/TR/webxr/ [W3 WebXR, 2020].

- Wagner, D., Billinghurst, M. and Schmalstieg, D. "How real should virtual characters be?" In Proceedings of the 2006 ACM SIGCHI international conference on Advances in computer entertainment technology. [Wagneret al., 2006].

- Wang, C.P., Wilson, K., Snavely, N., "Accurate geo registration of point clouds using geographic data." In: 3DV (2013). [Wang et al., 2013]

- Won, J. Gopinath, D. and Hodgins, J. (2020), "A Scalable Approach to Control Diverse Behaviors for Physically Simulated Characters", In SIGGRAPH. [Won et al., 2020].

- Zöllner, M., et al. An Augmented Reality Presentation System for Remote Cultural Heritage Sites. in The 10th International Symposium on Virtual Reality, Archaeology and Cultural Heritage VAST. 2009. [Zöllner et al., 2009]

National Level State-of-the-Art Facilities Development and

Budget Justification

iHUB Drishti proposes to set up following national level state-of-the- art facilities:

Real Time Computer Vision and Sensing Lab
Data Collection and Annotation Lab
Vision based Autonomous Systems Lab
Mixed Reality Lab
Trustable Visual Intelligence Lab
Visual Computing Lab

These labs will be available physically as well as remotely to all academicians, research labs, MSME's including Start-up's and other industries for technology development, characterisation and testing.

—

Real-time Computer Vision and Sensing Lab

Under this horizontal, a state-of-the-art Real-time Computer Vision Lab and Sensing lab equipped with both software development tools and hardware platforms will be established. This facility will be accessible to researchers and industries. This lab will be also involved in the development of a new generation of image sensors.

The lab will provide computer vision-based very low computational complexity algorithms for performing the perception-action action cycle in real-time. Various equipment like Xilinx, Zynq, Versal AI core series etc will be used for the same. In addition to this, the lab will support the characterization of CMOS image sensor, to implement different algorithms on hardware and verified using commercial off-the-shelf components and mimic the performance with the real world using state of the art facilities involved in this lab. Using the state-of-the-art characterization facilities including dynamic signal analyser for low noise measurement, low noise preamplifier, spectrum analyser, high resolution data acquisition system etc help us meet the industry standards, which will result in Intellectual Properties generation and more employment opportunities. This lab will be the first in India where hardware and software facilities related to imaging are available in a common platform. List of Capital Equipment and Software is as follows:

Xilinx Alveo: Adaptable accelerator cards for Data Center workloads	IoT networking Edge Gateways/Routers (Various Types)
Versal AI Core Series VCK190 Evaluation Kit	DC light source
Versal Prime Series VMK180 Evaluation Kit	Dynamic signal analyser
Zynq UltraScale+ MPSoC ZCU106 Evaluation Kit	Optical Accessories (including C-mount filters, optical filters, alignment and testing accessories, imagine lens, integrating sphere, frame grabber, resolution chart, polarizers, optical bench and other related accessories)
Zynq UltraScale+ MPSoC ZCU104 Evaluation Kit	
Vitis Unified Software Platform (Vitis AI Model Zoo, AI Optimizer, AI Compiler, AI Library, AI Profiler, AI Quantizer)	Logic Analyser and pattern generator
SDAccel and SDSoC Environments	Spectrum Analyser
Model Composer	Mixed signal oscilloscope
Vivado Design Suite - HLx Editions	Picoammeter
Intel® Arria® 10 FPGA Platforms	Low noise preamplifier
Intel® FPGA Programmable Acceleration Card (Intel® FPGA PAC)	Data acquisition system
	Ancillary Equipment (power supply, function generator, multimeter)
OpenVINO™ toolkit	DJI's Spark with Intel MOVIDIUS VPU
Qualcomm Snapdragon	Manifold 2-C (DJI)
Latte Panda	EINSCAN-PRO 2X 3D Scanner
NVIDIA Jetson Tx2/2i Dev Kit	Event based camera (DAVIS-346)
NVIDIA Xavier Dev Kit	High speed camera (i-SPEED 510)
NVIDIA Nano Dev Kit	Light Field Camera (Lytro Illum)

Google Corel Development Board (TPU) / Corel Mini Corel USB Device (TPU) Google SOM (TPU) Khadas VIM2 (NPU Based) Raspberry PI 4B HiKey 970 Dev Board (NPU based) Intel Neural Stick 2 PCs/Laptops & Printer Rugged Laptops for Field Testing Azure Kinect	Intel® RealSense™ Depth Module D430 + Vision Processor D4 Board SPAD Asynchronous Camera (SPCM-AQRH-16) FLIR T540 Thermal Imaging Camera Canon EOS R+Lense High speed Signal Processing units with GPU (TI) Hikvision PanoVu Series DS-2CD6924F-IS 8MP

Budget: INR 4.75 crores for the lab

Recurring cost:

IC fabrication	3	UMC/AMS	80 lakhs
Packaging and wire bonding	25	UMC/AMS	17 lakhs
Soldering units			3 lakhs
Total			100 lakhs

Data Collection and Annotation Lab

Building CV, AR and VR systems for unconstrained applications require a large amount of labeled data. The availability of correctly labeled and validated data is a major bottleneck in building accurate trustable systems. For several problems, classification engines can be designed by performing convenience sampling and collecting data from the web. However, building trustable and explainable systems require annotations not only with respect to the output class, but also with respect to attributes, semantic segmentation, and marking the regions of importance. While data annotation for object recognition can be outsourced, annotation for specialized applications like the ones in the healthcare vertical requires close collaboration with domain experts. To facilitate this activity, we plan to set up a data collection and annotation lab. Nowadays, there are several companies that provide data collection and annotation as a service. While many of these companies outsource the tasks to Indian manpower, there are very few companies that provide this service. By the year 2026, the global data annotation tools market is predicted to be more than a billion dollars. It is to be noted that this is only the outsourced marked, and does not include the internal resources spent on this process. Therefore, it is envisioned that this lab will not only serve as support for the hub, but it can also be a source of revenue generation for the hub. The items to be purchased in this lab include:

- Workstations with large monitors
- Storage facility
- Software licenses

Budget: INR 2.5 crores for the lab

Vision based Autonomous Systems Lab

We propose to create a national level state-of-the-art vision-based autonomous systems lab equipped with hardware for algorithms implementation and demonstration towards semi to full autonomy will be established. This facility will be accessible to researchers and industries. Details of the Vision-based Autonomous lab facilities and targeted projects will be part of vertical on Computer vision for Autonomous Systems. List of Capital Equipment and Software:

Outdoor All-terrain Ground Vehicles with a collaborative manipulator, 3D Vision and portable control unit Indoor differential drive Mobile manipulator robot with 3D vision Micro Robots (flying robots, quadrabots, robofly/robobee, legged robots) Robot buddies Autonomous Cart for warehouse application Autonomous CAR Quadcopters with 3D vision CPU, GPU, RTK, GPS, IMU, 4K Camera and remote control MicroLab Box	Engine based helicopter with 3D vision and other sensors portable control unit 3D Laser Scanner for 3D vision Dual Arm Robotic Arm with 3D vision Teleoperation module with AR VR, haptic feedback, force torque sensors and Exoskeleton Customised Autonomous Robotic Systems Robotic Arm with 3D vision for Industry 4.0 Manufacturing Applications CNC Lathe with 3D vision for Industry 4.0 Camera-based system with software for Industry 4.0

Budget: INR. 7.75 crores for the lab

Mixed Reality Lab

The life of a typical VR/AR application initially comprises three broad components: the design of its content and interfaces, the development of the application including the interplay of all included devices, and its deployment and use. Each has somewhat different infrastructure requirements. The ultimate success of a VR/AR application depends on the availability of proper infrastructure for each component. These can range from a simple mobile phone at the affordable end of the spectrum to dedicated experiential space with multiple wireless display devices, wall-sized wrap-around displays, full body tracking, eye tracking, lighting devices. 3D interaction devices with haptics, high-end graphics workstations, software setup, production studio, etc.

This TIH aims to develop, first of all, a national VR/AR resource. This includes a physical space with the requisite hardware, software, and other infrastructure to develop and experience VR/AR. This will be available for demonstration, training and development to qualified researchers and developers. The TIH will design guidelines for anyone to set up entry-level, medium-level and advanced-level VR/AR labs in a manner that they can be networked with others to produce a common inter-operable and collaborative VR/AR fabric -- a true national facility. In addition to these guidelines, TIH will initially set up 3-4 such labs across the country, to boot-strap and demonstrate the concept. Specifically, our goals include:

- Setting up national level lab infrastructure at the Hub that has access to cutting edge compute and display hardware. These will include workstations with professional grade

GPUs, AR HMDs like HoloLens and Magicleap, room-scale VR setups like the HTC Vive Pro and Valve Index, immersive CAVE infrastructure, full body scanner, multi camera motion capture, six degree-of-freedom haptic devices.

- Medium scale lab infrastructure at the project partner locations with access to cutting edge compute facilities like workstations with consumer grade GPUs, AR and VR HMDs like HoloLens, HTC Cosmos and Oculus Rift. Specialized facilities like motion capture, laser scanning, volumetric reconstruction rigs may also be set up.

- Small scale lab infrastructure that can be replicated at multiple locations in the country using cost effective hardware for mobile AR/VR like Google Daydream, Oculus Go, Oculus Quest and consumer grade workstations.

These labs will have the following (an indicative set).

- 10 active stereo projectors (2560×1600 each)
- Digital edge blending
- Free-standing structure with front-surface mirrors
- Raised floor surface
- Custom projector positioners
- [Active Stereo](#) and 10 pairs of shutter glasses
- Video, projector control, projector sync, and power cabling
- Installed and calibrated on-site by Visbox, Inc.
- Optical Tracking
    - ARTTRACK5 system (4 high resolution cameras)
    - Tracked head, tracked wand (controller)
- Image Generator
    - Graphics workstation with NVIDIA Quadro graphics
    - 19" APC rack and all cabling
    - Control console stand with monitor/keyboard/mouse
    - Software demos, applications, and development libraries
- Surround Sound System
    - Surround speakers: 5 satellite and subwoofer (5.1)
    - Surround system processor/amplifier
- MiddleVR for Unity
- Cost of such system will be around Rs. 2.5 crores

Laser scanner	Audio solution
Motion capture	SoftwarePhantom Premium 1.5 HF
Full-body scanner	Haptics Devices (inclusive of assembly)
High resolution Video+RGBD cameras	Interaction devices/Controllers
Room-size tracking solution	Maglev 200
Wireless Head mounted displays	CyberGrasp System
Stereo Graphics Work-stations	CyberTouch II

iHub Dishti will support setting up of the national level lab at IIT Jodhpur and three medium scale labs for collaborators at the cost of Rs. 14.5 crores. Other supported institutes (DTU) will set up a basic lab with Rs.50 lakhs.

Budget: 11.25 Crores

Trusted Visual Intelligence Lab

The lab aims to discover diverse approaches towards the achievement and integration of fairness, robustness, explainability, and accountability - across the entire life-cycle of system design. Potential work trajectories under the lab would be devising: a) mechanisms to quantify vulnerabilities of underlying system architectures; b) processes for bias detection and mitigation across datasets and models; c) strategies to induce different levels of governability, explainability, and interpretability into the foundational models; and d) user-interface design for ease in communication with- and functional evaluation of- explainable systems. The items to be purchased in this lab include workstations and software.

Budget: INR 1 Crore for the lab

Visual Computing Lab

Modern computer vision-related research requires high-end compute and storage support. Due to the specialized nature of the problems in computer vision, augmented reality and mixed reality domains, specialized graphical processing unit based computing facilities are required. Therefore, we propose to establish the Visual Computing lab, which will serve as a central hub supporting research activities of all the collaborators in the proposed TIH. Specifically, following are the objectives of this lab:

1. Enable high-end computing and large storage required for executing the next generation vision-driven applications.
2. Share very large-scale datasets with national as well as international collaborators.
3. Serve as leading computing hub in the country and support academic and research requirements of IIT Jodhpur and nearby Institutes.
4. Act as a computational hub for development, training, and testing of algorithms that involves computation in the petaflop range.
5. Support the startups for implementing their initial prototypes.
6. Facilitate training the skilled human resources required by country to support the make in India initiative,
7. Provide training related to usage of computing infrastructure and deep learning libraries to students and researchers by conducting workshops.
8. Serve as a center supporting research activities in the proposed TIH.

Budget: INR 12.75 Crores

3 DGX3	Ethernet Switch
200 TFlop HPC Nodes	KVM Console with switch
Master Node	Rack 42U with PDU
Head Node	Installation and commissioning
Storage (2 PB)	Laptops
Backup node	Workstations

Backup Software Capacity (6TB) Switches-DGX HPC connectivity	Software Licenses (Cloud, Storage, Security, Networking, SDE)

Budget

Table 1. Yearly Budget Requested from DST

	Year 1	Year 2	Year 3	Year 4	Year 5	Total
Technology Development	60,000,000	70,000,000	60,000,000	60,000,000	50,000,000	300,000,000
Manpower, HRD & Skill Development and TIH Manpower	34,827,600	45,703,600	49,224,800	51,531,000	44,713,000	226,000,000
Research and Development Labs	200,000,000	200,000,000				400,000,000
Innovation and Start-up ecosystem	2,500,000	10,000,000	24,000,000	26,500,000	27,000,000	90,000,000
International collaborations	0	3,800,000	3,900,000	3,900,000	3,900,000	15,500,000
Capex	50,000,000					50,000,000
Consumable, Contingencies, Utilities and Rentals	4,500,000	11,500,000	12,500,000	12,500,000	11,500,000	52,500,000
Int Travel	0	4,000,000	4,000,000	4,000,000	4,000,000	16,000,000
Total	351,827,600	345,003,600	153,624,800	158,431,000	141,113,000	1,150,000,000

Total Budget requested from DST (a): Rs. One Hundred Fifteen Crores
Funds planned to be generated from external sources (b): Rs. Fifty five crores
Overall Budget (a + b): One Hundred Seventy Crores

Table 2. Overall Recurring and Non-Recurring Budget from DST

Budget Head	Budget in Rs Crores					
	1st Year	2nd Year	3rd Year	4th Year	5th Year	Total
Recurring	11.75	12.00	13.00	14.00	12.25	63.00
Non-Recurring	7.50	15.00	23.00	4.00	2.50	52.00
Total	19.25	27.00	36.00	18.00	14.75	115.00

	Year 1	Year 2	Year 3	Year 4	Year 5	Total
Recurring	91,827,600	127,503,600	138,624,800	143,431,000	128,613,000	630,000,000
Non-recurring	220,000,000	257,500,000	15,000,000	15,000,000	12,500,000	520,000,000
Total	311,827,600	385,003,600	153,624,800	158,431,000	141,113,000	1,150,000,000

Table 3. Technology Development

	Year 1	Year 2	Year 3	Year 4	Year 5	Total
IITJ and Collaborators	30,000,000	40,000,000	30,000,000	30,000,000	20,000,000	150,000,000
Open Call	30,000,000	30,000,000	30,000,000	30,000,000	30,000,000	150,000,000
Total	60,000,000	70,000,000	60,000,000	60,000,000	50,000,000	300,000,000

Table 4. HRD Skill Development

Fellowship Name	Yearly Count	Year 1	Year 2	Year 3	Year 4	Year 5	One time development cost (in Year 1)	Total
CHANAKYA-GI (Graduate Internships)	46	2,300,000	2,300,000	2,300,000	2,300,000	2,300,000	5,000,000	16,500,000
CHANAKYA Postgraduate Fellowships	9	1,800,000	1,800,000	1,800,000	1,800,000	1,800,000	5,000,000	14,000,000
CHANAKYA Doctoral Fellowships	24	2,678,400	10,713,600	11,404,800	12,096,000	6,048,000		42,940,800
CHANAKYA Post-Doctoral Fellowships	0		0	0	0	0	0	
CHANAKYA Faculty Fellowships	3	1,200,000	1,800,000	1,800,000	2,400,000	1,600,000		8,800,000
CHANAKYA Chair Professors	3	900,000	1,800,000	2,700,000	2,700,000	1,800,000		9,900,000
CPS-Upgrading PG Programme	1						7,500,000	7,500,000
Total		8,878,400	18,413,600	20,004,800	21,296,000	13,548,000	17,500,000	99,640,800

Table 5. TIH Manpower

	Year 1	Year 2	Year 3	Year 4	Year 5	Total
CEO	1,350,000	2,820,000	2,940,000	3,060,000	3,120,000	13,290,000
CTO	1,080,000	2,280,000	2,400,000	2,520,000	2,640,000	10,920,000
Company Sec.	720,000	1,500,000	1,560,000	1,620,000	1,680,000	7,080,000
A/C & Store & Purchase	420,000	870,000	960,000	990,000	1,050,000	4,290,000
CA and Audit (outsource)	169,200	450,000	480,000	510,000	540,000	2,149,200
Executive Assistant	360,000	740,000	780,000	805,000	835,000	3,520,000

Network Engineer	540,000	1,200,000	1,320,000	1,440,000	1,500,000	6,000,000
Junior Assistant	210,000	450,000	480,000	510,000	540,000	2,190,000
Research Scientist	1,200,000	8,400,000	8,400,000	8,400,000	8,400,000	34,800,000
Integration Engineer	960,000	4,080,000	5,400,000	5,700,000	6,000,000	22,140,000
Hardware and Network Engineer	480,000	3,060,000	3,060,000	3,240,000	3,420,000	13,260,000
Coordinators Honorarium	800,000	1,200,000	1,200,000	1,200,000	1,200,000	5,600,000
Project Director's Honorarium	160,000	240,000	240,000	240,000	240,000	1,120,000
Total	8,449,200	27,290,000	29,220,000	30,235,000	31,165,000	126,359,200

Table 6. Lab Budget

Labs	Amount
Visual Computing Lab	12,75,00,000
Real Time Computer Vision and Sensing Lab	4,75,00,000
Data Collection and Annotation Lab	2,50,00,000
Vision-based Autonomous Systems Lab	7,75,00,000
Mixed Reality Lab	11,25,00,000
Trustable Visual Intelligence Lab	1,00,00,000
Total	40,00,00,000

Table 7. Innovation and Startups

	Year 1	Year 2	Year 3	Year 4	Year 5	Total
Translations Research	0	2,500,000	4,000,000	4,000,000	4,500,000	15,000,000
Engineering Accelerator	1,500,000	2,500,000	5,000,000	5,000,000	5,000,000	19,000,000
Startup and Entrepreneur	1,000,000	5,000,000	15,000,000	17,500,000	17,500,000	56,000,000
Total	2,500,000	10,000,000	24,000,000	26,500,000	27,000,000	90,000,000

Table 8. OpEx - Rentals, Workshops, Consumable

	Year 1	Year 2	Year 3	Year 4	Year 5	Total
Rental, Electrical,	3,500,000	7,500,000	8,500,000	8,500,000	8,500,000	36,500,000

and Utilities						
Consumables and Contingencies	500,000	2,000,000	2,000,000	2,000,000	1,500,000	8,000,000
Workshops and Outreach	500,000	2,000,000	2,000,000	2,000,000	1,500,000	8,000,000
Total	4,500,000	11,500,000	12,500,000	12,500,000	11,500,000	52,500,000

Overall Budget Justification

Technology Development: iHub Drishti can be self-sustainable through technology development and licensing. The technology will be developed by providing research grants to different researchers across the country. Substantial funds will also be used by partners for technology development. Filing patents in India and abroad also requires funds. Developing patent portfolios and managing portfolios also require funds. Overall, technology development funds will support the research ecosystem which will make the hub self-sustainable.

Manpower, HRD and Skill Development: iHub Drishti aims to engage with the next generation technologists and train at least 1000 students and professionals in 5 years. Funds are sought to undertake Chankya's programs, nurturing mentee institutions by training their students and faculty, support manpower at the TIH, and creare infrastructure for skill development activities.

Startup ecosystem: The Hub will promote start-ups, promote innovation and encourage Entrepreneurship for emerging applications. Funds required for supporting start-ups and product development to achieve transition from TRL 3-4 to TRL 6-7 and beyond. Funds needed to run engineering accelerator programs for startups. The hub will also provide some common facilities for startups. The hub will also have schemes to encourage and host women entrepreneurs. The hub will also organize hackathons for promoting innovation.

International Collaborators: The Hub has several international collaborators. For effective contributions of these collaborators, funds are required to host them in India at the TIH.

International Travel: The HUB requires travel related to Technology marketing. The HUB will also host several visitors and partners for various activities.The proposed HUB personnel will also have travel related expenses. Travel contingency expenses will also be met through this budget.

CAPEX and OPEX: Capital expenditure is required to set up the hub. For operational expenditures, including electricity and utility expenses, rental, contingencies, consumables, workshops and outreach a separate operational budget is required. IIT Jodhpur is kindly considering waiving off the first three years of the rental charges.

Management Plan

ihub -Drishti will set up a lean and agile management structure to facilitate directed research, translation, technology delivery and manpower development. for the stake-holders. The research plan proposed in the DPR will form the research charter of the hub. Each vertical will be considered as an independent functional unit. iHUB for each functional unit will prepare a detailed project management plan to achieve objectives and produce deliverables as per the charter. Each functional unit will be treated as a cost centre and will function with a clear yearly performance target of tangibles including revenue earning. In terms of revenue planning each functional unit will work out a long term plan for becoming financially self-sufficient in 5 years. The guiding principle of the management plan of the TIH i will be the following:

- To facilitate collaboration between individual research and technology development partners

- To come up with a flexible and agile architecture to support multiple different activities under the purview of the hub.
- Facilitating and enabling technology development, translation, incubation and industry interface
- Supporting project execution for line ministries
- Implement policies and programmes for human resource development in this area.

National Collaborators: Our collaborators are segregated as research collaborators, mentor-mentee institutions, and domain partners. Following activities will be undertaken to support:

- Joint projects with industry partners
- Ph.D. student exchange program
- visit of partnering students and faculty to/from IIT Jodhpur.

International Collaboration: We have identified theme areas of collaboration for international experts and we plan to collaborate with them via the following three methods.

- Support joint projects with industry partners
- Ph.D. student exchange program
- Support visit of student and faculty to the research lab directed by the collaborators
- Travel support of collaborators within India and hosting them at IIT Jodhpur

The institute has been actively pursuing collaborative works with many relevant industries in a broad domain of CPS and in particular Computer Vision, AR and VR. The Hub will actively work with user agencies and industries to solve technological challenges. The following are the main approaches iHub Drishti will pursue while engaging with industries:

- Joint projects with iHub Drishti and industry partners. The funds from the technology development budget will be used to develop the technology. The generated IPs will be shared based on contributions and prior agreement with the concerned Industry Partner.
- Industry supported Ph.D. scholars: This program will provide dedicated students who will work on broad domains of Computer Vision, AR, and VR related problems. These students will be part of a big problem and different students will contribute to solving their assigned problems.
- Joint bidding for different projects with industry partners: The Hub will also participate with industry partners to bid large projects floated by different governments/Ministries.

Organizational Structure

iHub Drishti is a Section-8 company and it's mandate includes research activities, training and capacity building, and technology generation to ensure self-sustainability. We envision this to become a national level entity partnering with the experts of CV and ARVR both nationally and internationally. To achieve these objectives, different entities are required. All the primary objectives are interlinked to each other and to achieve these objectives successfully, the organization is designed to have a close collaboration among different entities. The primary entities in the hub are as follows:

- Hub Governing Body
- Academic Research Team and Collaborators
- iHub Drishti Staff
- Industry Partners
- User Agency
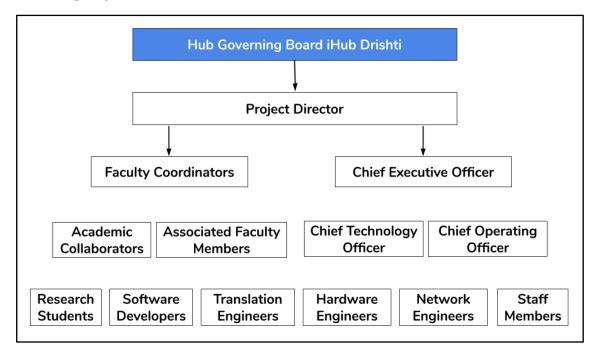
Figure: Schematic Diagram of the Organizational Structure of iHub Drishti.

TIH Governing Board

The TIH will be guided and governed by the Hub Governing Body, chaired by the Director, IIT Jodhpur. We have set up the Board of the Hub and the composition is listed below.

- Prof. Santanu Chaudhury, Director IIT Jodhpur, Chairperson
- Prof. Mayank Vatsa, IIT Jodhpur, TIH Project Director (convenor)
- Prof. Richa Singh, IIT Jodhpur
- Prof. Surajit Ghosh, Dean of R&D, IIT Jodhpur
- Prof. Kaushal A Desai, Associate Dean (Outreach and International Relations), IIT Jodhpur
- Prof. Anil Tiwari, IIT Jodhpur
- Prof. C. V. Jawahar, IIIT Hyderabad
- Prof. Vijay Chandru, Strancl-LS, Bangalore
- Dr. Manish Gupta, Google Research India, Bangalore
- Dr. Kaushik Saha, SRI, Delhi
- Dr. Murali Mohan, Mission Coordinator, DST

The Hub Governing Body (HGB) will be the apex body for the overall policy and strategic plan of iHub Drishti. HGB will also monitor the progress of the TIH and take major decisions regarding policy and overall functioning of iHub Drishti.

Academic Research Team

The academic research team comprises faculty members, post-doctoral fellows, doctorate and graduate students. These members will be from IIT Jodhpur, it's national and international collaborators. These members also include the mentee organizations that have been identified, the faculty members and students who will be working with the TIH. Our current list of collaborators is provided below, however, this will keep growing depending on the problem statements the TIH takes up and the researchers who partner with us to complete those. It is our vision that iHub Drishti will become the hub to connect all the researchers working in the area of computer vision, augmented reality and virtual reality.

National Collaborators	
Indian Institute of Science, Bangalore	IIIT Delhi
Indian Institute of Technology, Bombay	IIIT Hyderabad
Indian Institute of Technology, Delhi	Indian Statistical Institute, Kolkata
Indian Institute of Technology, Kharagpur	Indian Institute of Technology, Ropar
Indian Institute of Technology, Madras	Indian Institute of Technology, Hyderabad
Indian Institute of Technology, Kanpur	Indian Institute of Technology, Guwahati
IISER Bhopal	IIIT Bangalore
CSIR-CEERI, Pilani	CDAC Noida
CSIR-IGIB	IIT Indore
ICAR-CMFRI	ICAR-IARI
Mentee Institutions	
Delhi Technological University, Delhi	LNMIT, Jaipur
Punjab Engineering College, Chandigarh	MNIT, Jaipur
KLE Tech Hubli, Karnataka	NSIT, Delhi
International Collaborators	
Johns Hopkins University, USA	University of Warwick, UK

University of Maryland, College Park, USA	University of South Florida, USA
University of Lincoln, UK	Columbia University
Georgia Institute of Technology, USA	POSTECH, South Korea
University of South Florida, USA	Hong Kong Polytechnic University
SUNY Buffalo	Purdue University
TEES Institute for Manufacturing Systems, Texas A&M University	Laboratoire Mécanique, Surfaces, Matériaux et Procédés (MSMP), Ecole Nationale Supérieure d'Arts et Métiers, France
Queensland University, Australia	University of Toronto

TIH Staff

The employees of the TIH will be responsible for managing all the operations of the TIH, starting from planning different projects, technologies, interfacing between different agencies and institutions. The management will be governed by Chief Executive Officer, Chief Technology Officer and Chief Operating Officer along with the Project Director who will be faculty of IIT Jodhpur. Along with these, all the verticals identified in the TIH will have faculty coordinators who will be closely involved in planning and coordinating the research and development efforts across individual efforts.

iHub Drishti will also hire translational engineers, hardware engineers and software developers to take the research technologies developed by academic collaborators to the higher technology readiness levels that are suitable for deployment. The hardware engineers and network engineers will also manage the research labs and computing resources developed as part of the hub. The staff members will be responsible for managing the day-to-day operations related to finance, accounting and human resources.

Verticals	Coordinators	Co-Coordinators
Vision for Autonomous Systems	Prof. Suril Vijaykumar Shah	Prof. Amit Bharadwaj, Prof. Rajendra Nagar
Computer Vision for Better Living	Prof. Richa Singh	Prof. Deepak Mishra
Imaging for Document Analysis	Prof. C. V. Jawahar (IIIT Hyderabad) Prof. Anand Mishra	

Industry 4.0 and Cloud Manufacturing	Prof. Kaushal A. Desai, Prof. Santanu Chaudhury	
Dependable AR & VR for X	Prof. Subodh Kumar (IIT Delhi)	Prof. Parag Chaudhuri (IIT Bombay), Prof. Santanu Chaudhury

Industry Partners

Taking technology from the lab to the field requires a lot more than core research in the subject matter. The industry partners are essential to build the ecosystem of transferring the technology from research to successful deployment. This is where the hub will collaborate with our industry partners, both startups and industries, depending on the mutual interest and expertise. The institute has been actively pursuing collaborative works with many relevant industries in a broad domain of CPS and in particular Computer Vision, AR and VR. The Hub will actively work with industries to solve technological challenges. **The startups involved in the iHub Drishti may also be incubated at the Technology Innovation and Startup Centre (TISC) at IIT Jodhpur.** Following Start-ups /MSME and industries have agreed to partner with *iHub* **Drishti.**

Startups/MSME		
Vizara Technologies Pvt Ltd	SS Innovations	Kansara Modler Pvt. Limited
DependableVision.AI	Videoken Inc.	Kritikal Solutions Pvt. Ltd.
Algo8.AI		
Industries		
Google Research India	NVIDIA	NetApp
Samsung	JCB	Entuple Technologies Pvt. Ltd.
Tata Consultancy Services	National Instruments	Qualcomm

User Agencies

Based on the problems already identified under the five verticals, we have established confirmed partnerships with the following user agencies. These user agencies have agreed to partner with us and discuss the requirements. Many of them have also committed their support towards validating and implementing the technology solutions created by the iHub Drishti to fulfill user level requirements.

User Agencies	
DRDO CAIR	PGIMER Chandigarh

DRDO R&D Engineers Pune	AIIMS Jodhpur
DRDO Young Scientist Labs - AI	TeleRadiology Solutions
DRDO Young Scientist Labs - AT	CARING, Mahajan Labs
DRDO DTRL	Medanta Hospital
DRDO ISSA	Max Hospital
DRDO DLJ	Tata Memorial Hospital, Kolkata
CDAC Noida	Himachal Pradesh Forest Department
BPRD, Ministry of Home Affairs (Tentative)	Ministry of External Affairs (Tentative)
Ministry of Culture	National Digital Library of India
Ministry of Agricultural Research	Ministry of External Affairs

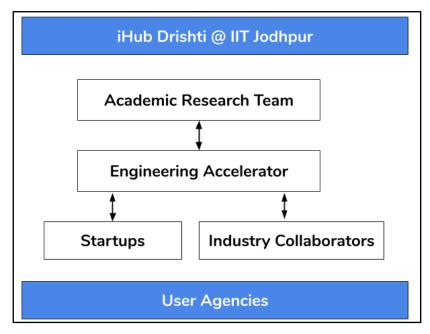Figure: Schematic Diagram of the linking of iHub Drishti's academic research team with startup and industry via an engineering accelerator.

Membership Model

All academic researchers, industry whose research agenda includes research and technology development in computer vision and AR-VR applications are invited and taken on board into TIH as members.

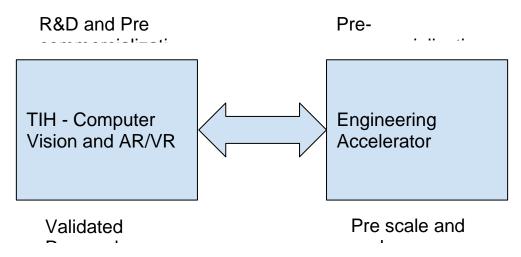The following SOPs are provided as annexures.

1. Annexure 1_SOP - Associate Member in HUB - Enrolment of Start-ups, Industry, R&D organizations as "Associate Members"
2. Annexure 2_SOP - faculty_experts from other entities as member in HUB - Enrolment of experts/ technologists/ faculties from entities external to IIT Jodhpur as "Honorary Member"
3. Annexure 3_SOP - IITJ faculty as member in HUB - Enrolment of IITJ faculties as "Honorary Member"

Entrepreneurship and Startup Ecosystem

The ecosystem at IIT Jodhpur essentially reflects freedom to encourage innovative ideas, R&D intensity, industry interaction, IP regime and protection, incentives for faculty/students to ideate/innovate, institutional setup to handhold/innovate entrepreneurs, startup policies, equities and investments, networks with other incubators, investors, angels and entrepreneurs. IIT Jodhpur Innovation Complex has already been inaugurated by Dr Arabinda Mitra, Scientific Secretary, Office of the Principal Scientific Advisor, Govt. of India on the occasion of the Institute Foundation Day on 2nd August, 2019. As a part of this, IIT Jodhpur has already setup a Technology Innovation and Startup Center (TISC) which is a registered section-8 company incorporated on 3rd July, 2019 under the companies Act, 2013. IITJ-TISC is focusing on **Deep Tech** to promote startups/companies founded on scientific discovery or meaningful engineering innovation to solve the big issues that really affect the world around us through transformative technologies. It basically aims at nucleating a cluster of new age ventures around the focal theme "Artificial Intelligence of Things (AIOT)", which is thematically unique in the country. Deep tech fields of TISC interests include: New materials, specially materials intelligence**,** Artificial Intelligence**,** Precision Medicine & Multi-omics, Cyber-security, Digital economy, Robotics, Advanced Communications, Quantum Computing, etc. Possible fields for deep tech applications include: Agriculture, Food (including processing, analytics and computing), Life sciences, Aerospace, Energy, Defence, etc. This incubator has also been recognized in the year 2019 for support by the Ministry of MSME and Ministry of Electronics and Information Technology, Government of India. A rich infrastructure is being built to support around 50 incubates at a time. Presently around 15 students startup ideas are in the process for their incubation and financial support by the Ministry of MSME. The activities of TIH will be linked with the IITJ-TISC.

Engineering Accelerator

One of the critical Facilities for this TIH will be the Engineering accelerator. Engineering accelerators bring together critical aspects of Data, Talent, Domain and Entrepreneurship to accelerate research into real time systems that can create impact on the ground.

R&D and Pre
~~commercialization~~

Pre-

TIH - Computer Vision and AR/VR	Engineering Accelerator

Validated

Pre scale and

While the Innovation Program of the TIH will mostly be led by Academia and other Research Scientists the Translation program will be led by the Engineering Accelerator. This Accelerator will act as a Bridge between Academic Research and Start-up's. Once a Technology has reached TRL 2-3 levels it will move to the Engineering Accelerator which will aim to take it to the TRL 4-5 level.

IIT Jodhpur will have a strategic tie-up with AI foundry (www.aifoundry.ai) which is a private company to execute this Engineering Accelerator. The Accelerator will be physically located on both IIT Jodhpur premises and at the Bangalore offices of AI foundry.

Legal Framework

The overall ambit of the legal framework pertaining to iHub Drishti, including IPR, patent, revenue sharing, is to be governed under the existing legal framework of the Government of India. The hub will enter into a memorandum of understanding with IIT Jodhpur and other collaborating institutions/agencies/industry with the broad outlay as follows:

- *The Purpose and Activities:*. iHub Drishti and IIT Jodhpur intend to discuss and explore the development of joint collaborations between themselves in the areas of research, education and training. The Parties understand and acknowledge that any specific activities and programs implemented pursuant to this MOU shall be subject to: (a) availability of funds; (b) the review and approval of each Party's authorized representatives; and (c) the execution of further written agreements such as those necessary to undertake sponsored research, or research collaborations. Until the execution of such further written agreements, each party is responsible for its own costs and expenses incurred under this MOU.
- *Compliance with Laws*: Each Party shall comply with the laws and regulations, including, without limitation, export control, non-discrimination, data privacy, immigration and sanctioned parties or transactions.
- *Ownership of Project Work Product/Intellectual Property:* The parties will enter into a separate IP agreement for each of the projects. Broadly, IIT Jodhpur and iHub Drishti will jointly own any and all work product and/or Intellectual Property developed jointly (e.g., to the extent the parties would be considered joint inventors and/or joint copyright holders, as applicable, under relevant intellectual property laws) under this Agreement ("**Joint Intellectual Property**"). Each party will retain its rights in any work product and/or Intellectual Property developed prior to and/or outside the scope of this Agreement. In case of any externally funded project on which IITJ and iHub Drishti jointly work, the IP agreement of the funding agency will be applicable. Such details will be part of separate IP agreements along with a Statement of Work.
- *Commercial Licensing Opportunities Available to iHub Drishti and Revenue Sharing*. The hub will have the first right of commercialization for the projects supported by the hub. Generally, the revenue generated via the IP will be jointly shared. However, details of the licensing process, sub-licensing, exclusivity, and revenue sharing will be part of a separate agreement.

iHub Drishti will enter into similar agreements with other partnering institutions, user agencies, industry including startups, and other hubs based on mutual understanding and negotiations.

Environmental Impact

iHub Drishti is a computing hub in nature and thus, there will not be any environmental impact. Since the hub will be hosted at IITJ existing infrastructure, there are no new constructions, land acquisitions, dislocations and other environmental clearances are required.

a. Land acquisitions are not involved
b. Environmental clearances are not involved
c. Forestry clearances are not required
d. Wildlife clearances are not required.

Cost-Benefit Analysis

While there is no direct quantitative method for performing cost-benefit analysis for establishing the hubs, qualitative and unquantifiable benefits are as follows:

1. iHub Drishti will build the ecosystem of bringing the national and international experts in computer vision and ARVR together.
2. The network and interconnect group of faculty, students, and researchers associated with the hub will build a strong presence in the CV & ARVR community.
3. The publications and know-how generation will be an explicit and direct measurable asset which will bring international visibility.
4. The hub will produce the next generation of skilled researchers in CV & ARVR. Furthermore, the educational programs of skilling and reskilling will benefit thousands of students nationwide.
5. The hub will contribute towards the "Make in India" and **Atmanirbhar Bhara**t initiative. It will also directly help towards National Data Sharing and Accessibility Policy (NDSAP) and Open Government Data (OGD) Platform India.
6. The hub will nurture and foster a startup ecosystem and contribute towards the socio-economic growth of the country. The hub will be a bridge between the industry, academia and domain user agencies. It will open the plethora of opportunities towards different business areas and services related to CV and ARVR domain.
7. Government consulting and participation in policymaking in the domain of CV and ARVR will be an additional tangible benefit of setting up the hub.

Risk Analysis

Any research initiative always has an element of risk of developing and working with new ideas. The fast pace of research in CV and ARVR coupled with advancement in hardware technology increases the chances of risks. Unanticipated internal or external risks pose challenges towards the growth of the hub. There is always a risk of project failures or the products not being commercially successful or can have human factors for unsuccessful endeavor and complexities. Finally, there are risks around intellectual property in the modern technology-rich world. It needs to be acknowledged that there will not always be a return on the investment in research and it is possible to have major setbacks in certain cases. On the other hand, returns from such a research initiative can be also considerably greater than for other investments. Overall, taking a comprehensive view, investment in building a research ecosystem such as iHub Drishti should be viewed as an investment in the future.

Outcome

As delineated in the DST's document, iHub Drishti will strive to meet the goals such as manpower training, research publications, technology development, and startup ecosystem. Here, we provide an overview of our planned outcomes.

Educational Outcomes

With the changing landscape of the industry, the job requirements are also changing. Today, Artificial Intelligence, Machine Learning, and Computer Vision are one of the most sought after

educational technologies. It is not only required and desired in industry jobs, but the research positions in this area are also increasing significantly with increasing research thrust by the Government and Industry. Indian government has recently launched a 400+ crores *Future Skills Programme for Reskilling Upskilling of IT Manpower for Employability (PRIME).* iHub Drishti will collaborate with IIT Jodhpur and other partner institutions to offer educational programs on computer vision, augmented reality, virtual reality, and allied areas.

- The recent growth trend of participation in computer vision conferences globally is astounding. Within a few years, the top conferences in the area like CVPR, ICCV, ECCV and WACV have reached from receiving a couple of hundred papers to several thousands, and the attendance is also increasing. The AR/VR global summit had an attendance of more than 11,000. All of these events show major participation from the industry leaders as well. This shows the increasing interest and requirements of trained professionals in the area. IIT Jodhpur is currently running ***MTech and MTech-PhD programs on Artificial Intelligence*** and courses related to Computer Vision are integral to the program. These programs are designed to promote research and development activities in areas related to computer vision. These programs will produce 40-50 students on a yearly basis, out of which around 50% will get expertise in CV, AR and VR. These students will contribute to developing the research and development ecosystem in the country. Further, TIH will also focus on training Doctoral students of IIT Jodhpur around the identified research areas and inspire them to take to careers in technology development.

- IIT Jodhpur has recently launched the ***BTech in AI and DS program*** as well, with a large number of courses related to CV, AR and VR. The program also offers two ***BTech specializations in Visual Computing and Socio-Digital Reality*** that include courses like Computer Vision, Digital Image Processing, Computer Graphics, and AR-VR as compulsory courses. These are supported by a long list of elective courses such as Computational Photography, Medical Image Analysis, Visual Perception, Animation, and Video Processing and Analysis, Introduction to Haptics, and Human-Computer Interaction. Courses related to designing Dependable AI and Vision systems are an integral part of the program.

- Aligned with the vertical on Human Health, IIT Jodhpur in collaboration with AIIMS Jodhpur, has started a graduate program on ***Medical Technologies***. This program will provide a wonderful opportunity for doctors and engineers to work together towards creating solutions and technologies for improvement in the healthcare facilities.

- Several job market analysis reports have projected that AI specialists are among the top three job profiles emerging in the country. Considering these requirements, IIT Jodhpur has started an ***MTech Program*** in AI for working professionals and Computer Vision is one of the primary focus areas of the program.

- We will also develop course modules, both short-term and long-term on CV, AR and VR. These course modules will help students and professionals in both upskilling and reskilling themselves. A recent report by IBM on Indian entrepreneurship highlighted that 70% of India's venture capitalists surveyed indicated that startups are experiencing difficulties scaling due to challenges obtaining employees with the right skills. Such programs should enable them to learn more recent technologies and upgrade their knowledge and skills.

- IIT Jodhpur also offers **Continuing Education Programs** including degree programs and certificate programs such as MTech in Cyber-Physical Systems. We will also offer programs related to CV, AR, and VR, including focused programs for researchers and developers from different user agencies such as DRDO and law enforcement officials.

- The TIH would conduct short term courses and skilling programs for the upgradation of existing manpower employed within manufacturing industries across the nation in collaboration with CII and associated ministries.

- We will organize **workshops and conferences** involving academic and industry leaders to share the developments which are happening in this area, across the globe. IIT Jodhpur will be hosting the **International Conference on Computer Vision, Graphics, and Image Processing 2020** which is one of the prestigious conferences in India and the **IEEE International Conference on Face and Gesture Recognition, 2021** which is a flagship conference of the IEEE Computer Society and IEEE Biometrics Council.

Research Outcomes

Along with research publications in top journals and conferences, patent and open source codes, and joint collaboration with industries and end users for translational research, iHub Drishti will focus on following research outcome:

1. Efficient sensing framework of seeing through visual clutter

2. Recognition systems for identification of objects, humans and movement in unconstrained setting

3. Vision-based autonomous navigation capability in unstructured terrains and unknown environments

4. Vision based reasoning capability for decision making and acting in the environment

5. Fair, explainable, robust, trustworthy and secure Computer Vision and ARVR technology.

6. Advanced image based diagnostic and prognostic tools for dealing with trauma, chronic and infectious diseases.

7. Machine learning techniques for radiomics and radio-genomics to support individualized treatment plans.

8. Low cost techniques for point-of-care pathology assessment, minimizing requirement of labs for initial diagnosis.

9. Image-based analysis tools for ophthalmologic diagnostics, automated diagnosis and grading of diabetic retinopathy, detection of individuals suffering from silent hypoxia, better management of pregnant women, malnourishment of children and cardiac care at PHC's.

10. *AI based platform for responding to medical and other national emergencies – RAKSHAK platform.*

11. Technology Solutions for scalable online education, efficient access of online education resources, and Automatic assessment tools

12. Technologies for long-term sustainability and development of smart factories.

13. Technology for ARVR for Digital Heritage, Remote Collaboration, Education and Training, Sports and Entertainment, Combat simulator, and Digital Twin Applications

14. ARVR Technology for Rehabilitation & Therapeutics and Telemedicine

15. Vision-based technologies for GIGI (Guidance, Identification, Gauging and Inspection) for reduced human intervention and operational efficiencies

16. Vision-based technologies for Manufacturing Prognosis

17. Cloud-based framework for resource sharing in manufacturing industries

18. Cloud-based assistance framework for assisting MSMEs in manufacturing sector

IP, Industry Linkages, User Agencies, and Startup Outcomes

1. Development of end user centric solution and demonstration of technology feasibility

2. Establish linkages with industries in the area of Computer Vision and ARVR for upscaling of technology from TRL 2-3 to TRL 5-6 levels

3. User agency centric translation research which can be used by the agencies for solving real world problems.

4. Patents, copyright, industrial design rights in technologies developed under the sub areas of sensing, recognition, navigation and doing of Vision based systems.

5. Provide new and emerging technology and compatible businesses with an environment that would support startups and increase their prospect of success;

Evaluation

As per the overarching mandate of the TIH prescribed by the National Mission, the mandate of iHub Drishti includes research, educational activities for up-skilling and reskilling programs, technology and IP generation, and self-sustainability. For the iHub Drishti, we have segregated these mandates and responsibilities into the following four dimensions:

- Research and development,
- Training and capacity building,
- Technology generation,
- Self-sustainability

There are several quantitative indicators to measure different kinds of activities, outputs and results. However, we believe that they can only act as an aid in the peer-review process. The quality of activities, outputs and results cannot be reduced to quantitative elements. Value or quality should be based on observable facts, including quantitative data, through analysis, discussion and interpretation. The evaluation should be performed considering the context, identity and missions of the TIH. With this understanding, we provide the performance metrics for each of the four mandates.

Research and Development	Training and Capacity Building
- Publications: articles in peer-reviewed journals, books, chapters, papers in conference proceedings - Research grants received and successfully completed - Prizes and distinctions awarded to members of the entity - Participation in national and international collaborative research projects - Participation in editorial committees, in the scientific committees of symposia or conventions, scientific review bodies	- Reskilling and up-skilling programs - Manpower trained, both at the host institute and collaborating institutions - Educational programs in collaboration with academic institutions - National and International events organized - National and international network
Technology Generation	**Self-sustainability**
- IP and patent portfolio - New products or market-ready technologies developed - Spectrum of technology development according to different readiness levels	- Technologies commercialized - Start-ups incubated and graduated - Transactions with industry - Revenue generated: turnover and profit

Tentative Implementation Action Plan for iHub Drishti

MAJOR ACTIVITY	YEAR 1				YEAR 2				YEAR 3				YEAR 4				YEAR 5			
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
Establishment of TIH	X	X	X																	
Building interior and utility arrangement	X	X	X	X	X															
Staff recruitment	X	X	X																	
Procurement of equipment and consumables, lab establishment	X	X	X	X	X	X														
Preparing Guidelines	X			X	X															
Distribution of Initial Projects		X	X	X	X	X														
Call for proposals			X	X	X	X	X	X	X	X	X	X	X	X	X	X	X			
Assigning/sanctioning projects			X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X		
Review of Projects			X	X				X	X				X				X			
Yearly review of TIH progress			X		X			X	X				X				X		X	
Mid Term Review			X						X											
Preparation and Publication of progress reports			X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X

Planned target achievements plan (cumulative):

Time in Months	6	12	18	24	30	36	42	48	54	60
Technology Development	2	5	10	20	30	40	55	70	85	100
Entrepreneurship Development	-	-	10	25	35	45	60	75	90	100
Human Resource Development	-	10	15	20	30	40	50	70	85	100

| International Collaboration | | 25 | 25 | 50 | 50 | 75 | 75 | 100 | 100 | 100 |