**A Seminar Report On**

# 3D Human Sensing

**Submitted in partial fulfillment of requirements for the award of the degree of**

### MASTER OF COMPUTER APPLICATIONS

**Submitted by**

## G.Vamsi (22091F0060)

#### Under the Guidance of

**Dr.R.Raja Kumar M.Tech, Ph.D.**

**Professor, Dept. of CSE**

### DEPARTMENT OF MASTER OF COMPUTER APPLICATION

##### RAJEEV GANDHI MEMORIAL COLLEGE OF ENGINEERING & TECHNOLOGY AUTONOMOUS

**Affiliated to JNTUA Anantapuramu,** APPROVED BY A.I.C.T.E., NEW DELHI, ACCREDITED BY NAAC of UGC, NEWDELHI with “A+” grade, ACCREDITED BY N.B.A, NEWDELHI,

NANDYAL 518501, (Estd-1995)

**YEAR: 2023**-**2024**

## Rajeev Gandhi Memorial College of Engineering & Technology

##### (AUTONOMOUS)

APPROVED BY A.I.C.T.E., NEW DELHI, Affiliated to JNTUA, Anantapuramum, ACCREDITED BY NAAC of UGC, NEWDELHI with 'A+' Grade,

ACCREDITED BY NBA of AICTE, NEWDELHI NANDYAL-518501

**(ESTD -1995)**

### MASTER OF COMPUTER APPLICATION CERTIFICATE

This is to certify that the report entitled “**3D Human Sensing”** is a bonafied record of the technical seminar done by **G.Vamsi** (22091F00060) under my supervision and guidance, in partial fulfillment of the requirement for the award of Degree of Master of computer Applications from **Rajeev Gandhi Memorial College of Engineering & Technology,** Nandyal for the academic year 2023-2024.

**Guide Dr.R.Raja Kumar** Professor, Dept. Of CSE.

### HOD

**Dr. K. Subba Reddy M.Tech, Ph.D.**

Professor and HOD.

**Place:** Nandyal

#### Date:

***Candidate’s Declaration***

I hereby declare that the technical seminar entitled **“3D Human Sensing”** was carried out and written by me under the guidance of **Dr. R. Raja Kumar M.Tech Ph.D**, **Professor**, Dept. of Master of Computer Application, R.G.M. College of Engineering & Technology, and this technical seminar report is submitted in the partial fulfillment of the requirements for the award of the degree of “Bachelor of Technology” in **Computer Science & Engineering**. I have not submitted the matter presented in this report anywhere for the award of any other Degree.

###### G.Vamsi

Roll No.:22091F0060

Dept. of MCA, RGMCET.

# Preface

I have made this report file on the topic **3D Human Sensing,** I have tried my best to elucidate all the relevant detail to the topic to be included in the report. While in the beginning I have tried to give a general view about this topic.

My efforts and wholehearted co-corporation of each and every one has ended on a successful note. I express my sincere gratitude to **Dr. R. Raja Kumar M.Tech Ph.D.** who have assisted me throughout the preparation of this topic. I thank him for providing me the reinforcement, confidence and most importantly the track for the topic whenever I needed it.

G.Vamsi 22091F0060

### ABSTRACT

The development of 3D video in recent years realizes 3D surface capturing of human in motion as is. In this paper, we introduce 3D human sensing algorithms based on 3D video. Since 3D video capturing does not require the object to attach special markers, we can capture the original information such as body motion or viewing directions without any disturbance caused by the sensing system itself.

The first strategy, known as passive vision, attempts to analyze the structure of the scene under ambient light. In contrast, the second, known as active vision, attempts to reduce the way in which images are formed. Sensors that capitalize on active vision can resolve most of the ambiguities found with 2-D imaging systems. For 3D human sensing, model based approaches are suitable since image-based methods do not produce 3D information.

The 3D shape estimation done in the model-based approaches is a classic but open problem in computer vision. The development of 3D video in recent years realizes 3D surface capturing of human in motion as is. In this seminar, introduce 3D human sensing algorithms based on 3D video. Since 3D video capturing does not require the object to attach special markers, can capture the original information such as body motion or viewing directions without any disturbance caused by the sensing system itself.

## CONTENTS

**Chapter No. TITLE Page no.**

1. INTRODUCTION 1
2. HISTORY 4
3. WHAT IS A 3D HUMAN SENSING 5
   1. [Definition 5](#_TOC_250002)
   2. [Key Attributes of 3D Human Sensing 6](#_TOC_250001)
   3. [System Implementation 8](#_TOC_250000)
4. WORKING 9
5. FEATURES 11
6. APPLICATION 12

7 ADVANTAGES & DIS ADVANTAGES 13

1. FUTURE WORK 14
2. CONCLUSION 15

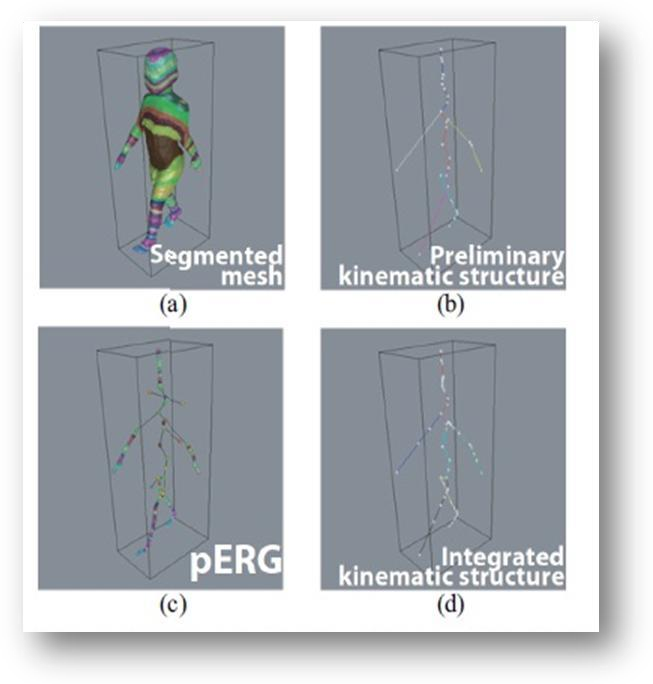
BIBLIOGRAPHY 16

### LIST OF FIGURES

**FIG NO. NAME OF THE FIGURE PAGE NO.**

|  |  |  |
| --- | --- | --- |
| Fig1 | Human Sensing | 5 |
| Fig2 | Depth Sensing | 6 |
| Fig3 | Gesture Recognition | 6 |
| Fig4 | Multi-person Sensing | 7 |

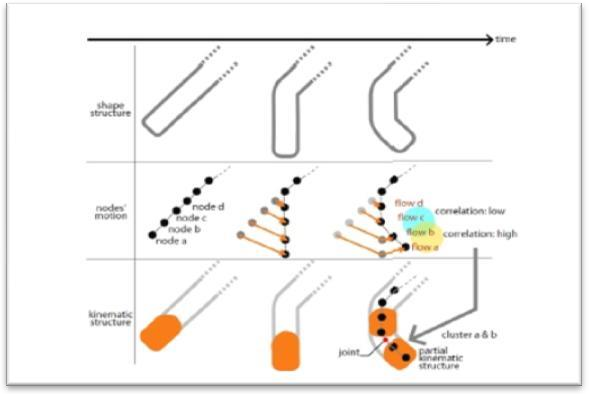
## CHAPTER-1 INTRODUCTION



The expansion of 3D video technology in the recent years has been grasped 3D shape apprehending of the object in motion as is [3] [1] [2] [4]. Subsequently 3D video is captured by conformist 2D cameras, the object is not required to attach special markers or to wear a special clothing. This is a vibrantbenefitin contradiction with other motion capturing technologies, and so 3D video is apposite for 3D digital archiving the motion of humans including immaterial cultural resources. Nevertheless, 3D video itse lf is purely a non-structured 3D outward data as same as pixel creeks of conventional 2D video. This paper shows how we can intelligence the human motion from raw 3D video.This is an illposed problem to guesstimate the original 3D shape from its 2D predictions. Many papers have anticipated practical algorithms which assimilate conventional stereo matching and shape-fromsilhouette technique to produce full 3D shape as photographexterior. We are imagining that we have the optimal photo exterior of the object and use it as the 3D real surface of the object [8][5]. Figure 1 shows our 3D video capturing scheme. The top and second rows show an example of multi-viewpoint input images and object regions in them respectively. The visual hull of the object is then calculated using multi-viewpoint silhouettes as shown in the third row, and we refine it through pictureconstancy optimization and gain the optimal 3D surface of the object. Lastly, we map the touches on the 3D superficial. The lowestrumpus shows sample version of the concluding 3D surface estimated from the multi-viewpoint images shown by the top row.

###### Kinematic Structure Estimation from 3D Video

In this section we familiarize an algorithm to estimate the kinematic edifice of an uttered object caught as 3D video. The idea is a time-series of 3D surfaces, and we build up the kinematic structure chastely from the effort data. Let Mt mean the say 3D exterior at time t (Figure 2(a)). We first figure the Reebgraph [6] of Mt as shown in Figure 2(b). Reeb graph is added based on the essential of geodesic reserves on Mt and gives a graph structure like to the kinematic construction. Figure 2(a) shows the surface portioncreated on the essential of geodesic objectivities.

Though, the sense of Reeb graph does not assure to the whole graph limits pass privileged of Mt and aboutlimits can go outdoors. So we adapt such shares of the Reeb graph to make unquestionable that it will be encaged by Mt . Figure 2(c) shows the adapted graph which we call We start from building pERGs at eachborder, then we choice “seed” pERGs which have no collapse of their body parts. Here we use a humblesupposition that a kernel ERG should have five twigs since we focus on the human behaviour.

Then we do pERG-to-pERGappropriate from kerneledges to their neighbours. We distort the kernel edges so as to adequate to its neighbours, and recurrence it till the appropriatefaultoverdoes a definiteverge. This procedurestretches topologically isomorphic interlude for each kernel edges as shown in the top of Figure 3. In each intermission, we put ona knotbunching to discovervoicedconstruction (Figure 4). Lastly, we assimilate enunciated constructions probable at all intermissions into acombined kinematic building as shown in the bottom of Figure 3. Figure 2(d) and 5 show the concludingunited kinematic constructionprobablechastely from the effort 3D surface arrangement.

###### Visibility of the Model Surface

Initially we introduced our discernibility definition on the model M(p) using the collision detection between the parts of the body. Since crashedareas cannot be experimented from any cameras in common, we notice such areas as shown in Figure 8. The colour indicates the distance between the points to its

closest surface of additional parts. Using this distance and discernibility, we describe the reliability of M(p) as

Where v denotes a vertex in M (p), and d (M(p),v) denotes the distance from v to the closest point of other parts. Visibility of the Experiential Surface Next we introduce the discernibility of the experimental surface Mt. Since Mt is estimated from the multi-viewpoint images, the vertices on Mt can be considered by the number of the cameras which can detect it. If one or less camera can observe a vertex v, v cannot be photo-consistent and the position of v is interpolated by its neighbours. Meanwhile, if two or more cameras can detect v, v should be photo-consistent and its 3D location is projected explicitly by the stereo-matching. So we can conclude that the number of noticeable cameras of v expresses the dependability on its 3D location.

## CHAPTER-2 HISTORY

**Human sensing** (also called **human detection** or [human presence detection](https://en.wikipedia.org/wiki/Human_presence_detection)) encompasses a range of

technologies for detecting the presence of a human body in an area of space, typically without the intentional participation of the detected person. Common applications include [search and rescue](https://en.wikipedia.org/wiki/Search_and_rescue), [surveillance](https://en.wikipedia.org/wiki/Surveillance), and [customer analytics](https://en.wikipedia.org/wiki/Customer_analytics) (for example, [people counters](https://en.wikipedia.org/wiki/People_counter)).

Modern technologies proposed or deployed for human sensing include:[[1]](https://en.wikipedia.org/wiki/Human_sensing#cite_note-1)

* [Acoustic sensors](https://en.wikipedia.org/wiki/Audio_signal_processing)
* [Image recognition](https://en.wikipedia.org/wiki/Computer_vision#Bababoui) of human shapes
* [Infrared detectors](https://en.wikipedia.org/wiki/Thermographic_camera)
* [Pressure-sensitive floor tiles](https://en.wikipedia.org/wiki/Sensing_floor)
* [Radar](https://en.wikipedia.org/wiki/Radar)
* [Chemical sensors](https://en.wikipedia.org/wiki/Sensor#Chemical_sensor)
* Detection of the [mobile phone,](https://en.wikipedia.org/wiki/Mobile_phone) [computer](https://en.wikipedia.org/wiki/Computer), [Bluetooth](https://en.wikipedia.org/wiki/Bluetooth), or [Wi-Fi](https://en.wikipedia.org/wiki/Wi-Fi) signals of a device assumed to be in the possession of a person.
* [WiFi Sensing](https://en.wikipedia.org/wiki/WiFi_Sensing)

In 1997 [CAPTCHA](https://en.wikipedia.org/wiki/CAPTCHA) ("Completely Automated Public Turing test to tell Computers and Humans Apart") was invented. A test is presented to detect that a computer is being used by a human operator, preventing access to a protected resource by programs such as spam robots.

Various commercial heartbeat detection systems employ a set of vibration or seismic sensors to detect the presence of a person inside a vehicle or container by sensing vibrations caused by the human [heartbeat](https://en.wikipedia.org/wiki/Cardiac_cycle).

Another commercial product uses infrared light to detect the level of [carbon dioxide](https://en.wikipedia.org/wiki/Carbon_dioxide) in an enclosed space, from which it infers the presence of humans or other living creatures.

In September 2013, the [United States Department of Homeland Security](https://en.wikipedia.org/wiki/United_States_Department_of_Homeland_Security)'s [Science and Technology](https://en.wikipedia.org/wiki/DHS_Science_and_Technology_Directorate) [Directorate](https://en.wikipedia.org/wiki/DHS_Science_and_Technology_Directorate) demonstrated a prototype of the FINDER radar technology device, which it developed in conjunction with NASA's [Jet Propulsion Laboratory](https://en.wikipedia.org/wiki/Jet_Propulsion_Laboratory). FINDER uses microwave radar to detect the unique signature of a human's breathing pattern and heartbeat, through 20 feet of solid concrete, 30 feet of a crushed mixture of concrete and [rebar](https://en.wikipedia.org/wiki/Rebar), and 100 feet of open space.

In September 2014, the DHS promoted the technology to [SWAT teams](https://en.wikipedia.org/wiki/SWAT) at the [Urban Shield](https://en.wikipedia.org/wiki/Urban_Shield) trade show.

## CHAPTER 3

**WHAT IS 3D HUMAN SENSING?**

#### Definition:

It is a depth-sensing technology that augments camera capabilities for facial and object recognition. The process of capturing a real-world object's length, width, and height with more clarity and in-depth detail than can be achieved using a number of different technologies.

Three dimensional (3D) technology is a momentous scientific breakthrough. It is a depth-sensing technology that augments camera capabilities for facial and object recognition. The process of capturing a real-world object’s length, width, and height with more clarity and in-depth detail than can be achieved using a number of different technologies. 3D technology delivers unique advancements in the way day-to-day activities are perceived and approached.

3D is a real game-changer as manufacturers scramble to incorporate these new advancements into consumer products such as mobile phones. 3D sensing technology mimics the human visual system using optical technology, which facilitates the emergence and integration of augmented reality, AI (Artificial Intelligence), and the Internet of Things (IoT). This creates unique opportunities in consumer applications.

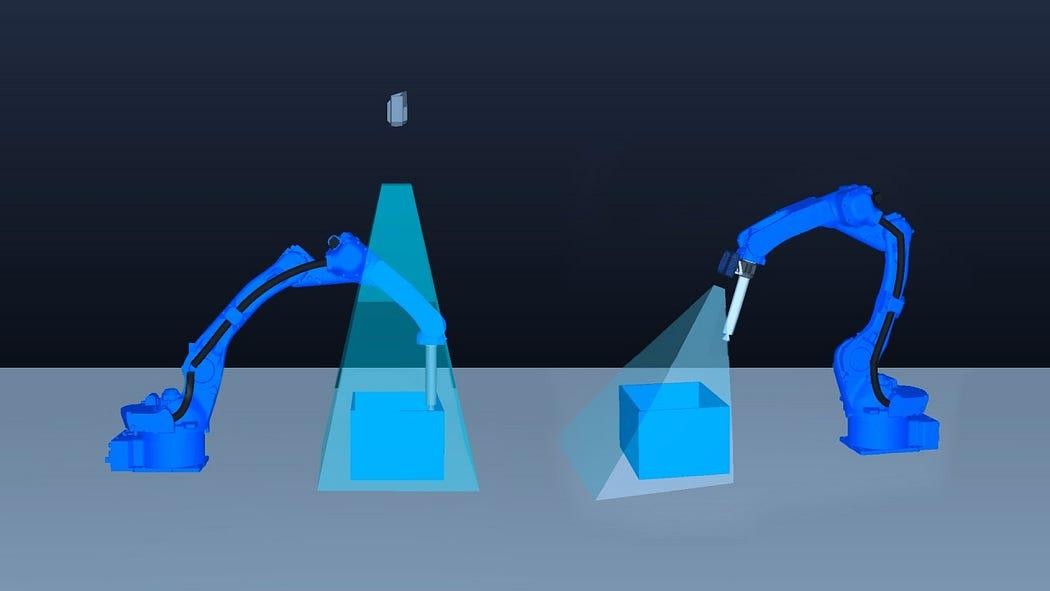


**Fig1:** 3D Human Sensing

#### Key Attributes of 3D Human Sensing:

###### Depth Sensing:

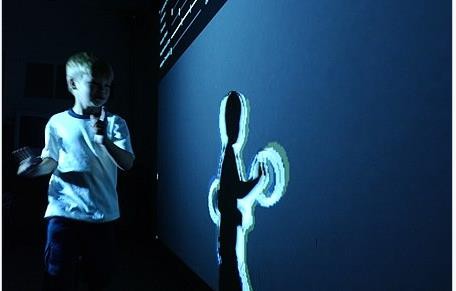
Depth-sensing is the measuring of distance from a device to an object or the distance between two objects. A 3D depth-sensing camera is used for this purpose, which automatically detects the presence of any object nearby and measures the distance to it on the go.



**Fig2:** Depth Sensing

###### Gesture recognition:

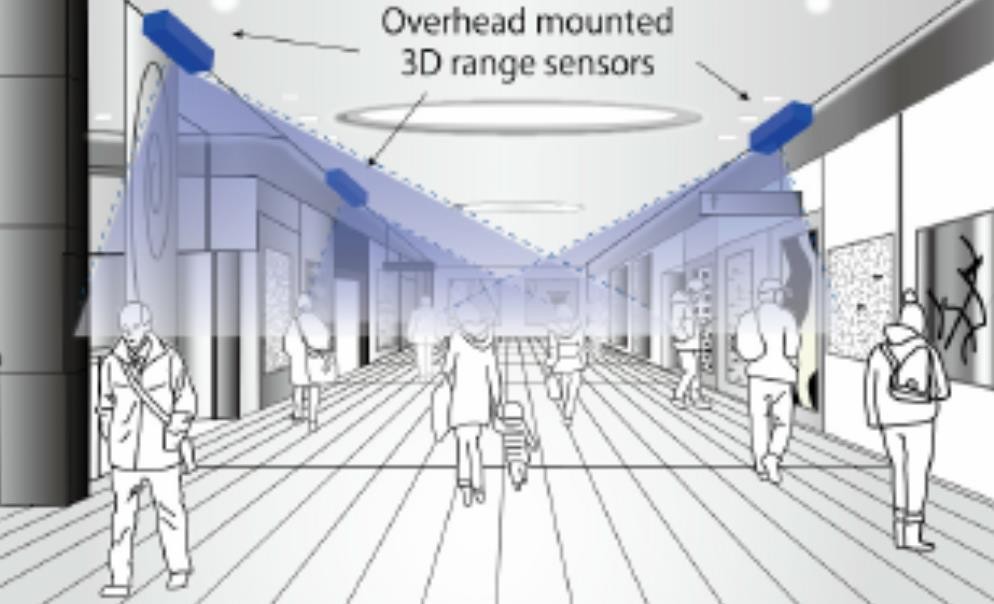
Gesture recognition is an area of research and development in computer science and language technology concerned with the recognition and interpretation of human gestures. A subdiscipline of computer vision, it employs mathematical algorithms to interpret gestures.



**Fig3:** Gesture recognition

###### Multi-person tracking:

Multi-person tracking in 3D human sensing refers to the capability of a system to simultaneously detect, identify, and track the movements of multiple individuals within a given environment.



**Fig4:**Multi-person Sensing

#### System Implementation:

Implementing a 3D human sensing system involves combining various technologies to accurately detect and track human bodies and movements in three-dimensional space. Here's an overview of the typical components and approaches involved in such a system:

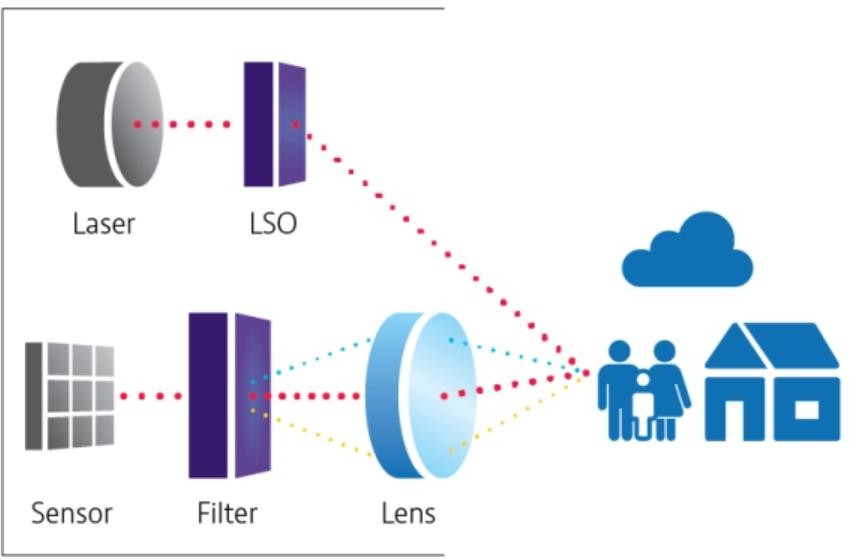
1. **Depth Sensing Technology**: Depth sensors, such as those based on Time-of-Flight (ToF) or structured light technology, are essential for capturing depth information. These sensors emit infrared light patterns and measure the time it takes for the light to bounce back, creating a depth map of the scene.
2. **Camera Systems**: RGB cameras are often used in conjunction with depth sensors to capture color information. Combining RGB data with depth information enables more detailed analysis of the scene.
3. **Skeleton Tracking Algorithms**: Once the depth data is captured, skeleton tracking algorithms are applied to identify and track human body joints. These algorithms analyze the depth map to detect key

## CHAPTER 4

**WORKING**

The technology behind 3D sensing is based on the projection of a light source toward an object and the collection of the same light waves after reflection to determine the shape and position of the object. Monitoring of a specific projected wavelength known as the “desired signal” throughout the process is a key factor.

Optical components shape and direct the outgoing light waves so that the illumination pattern of the desired signal is optimized. On the return path, focusing optics funnel all reflected light into a filter which separates the desired signal from erroneous light signals. This light is then directed to a sensor so that distance, shape, and depth information can be calculated.



Types of 3D Sensing

Two primary categories of 3D sensing technology are known as Time of Flight (ToF) and Structured Light (SL). Each sensing method lends itself well to multiple applications, depending on the specific requirements of the user.

###### Direct Time of Flight

As the name suggests, Direct Time of Flight (DToF) 3D sensing uses short, timed pulses of light followed by a direct measurement of the reflected light’s return time to determine the position of

an object. Much like ultrasound or sonar, the resolution of this method depends on the volume and

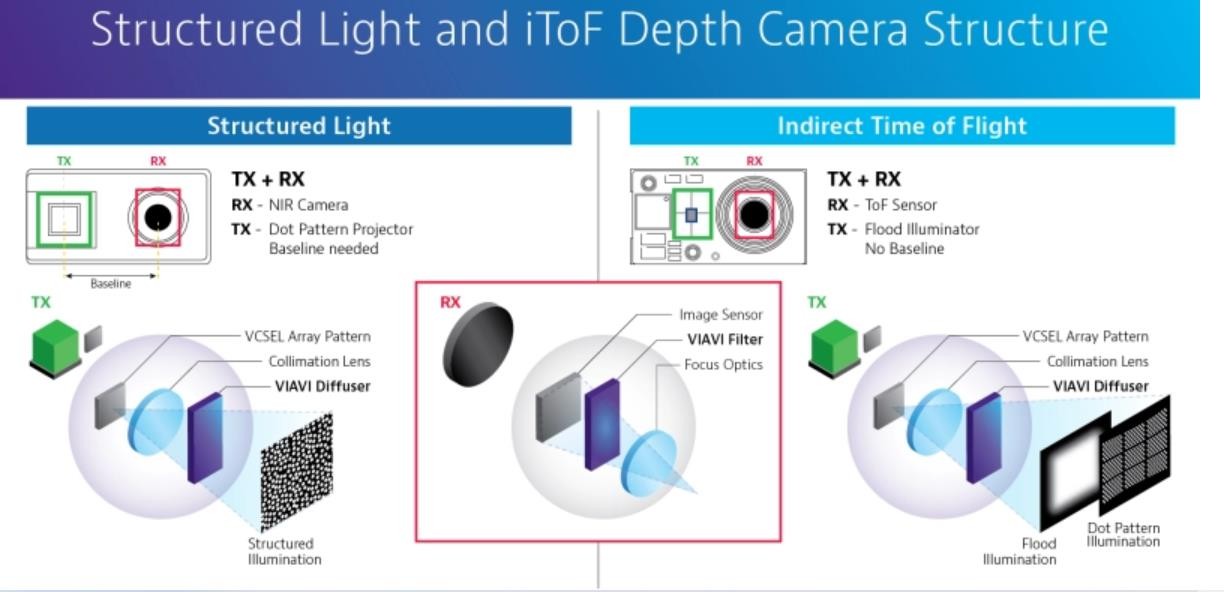
speed of data capture. DToF is used for Light Detection and Ranging ([LiDAR](https://www.viavisolutions.com/en-us/osp/products/lidar-filters)) systems in consumer electronics, autonomous vehicles, and other industrial applications.

###### Indirect Time of Flight

Indirect Time of Flight (IToF) technology uses the phase, rather than time signature, of the reflected light to measure distance from individual points on an object. Indirect ToF technology includes a modulated light source at a set frequency. Distances are determined based on the phase difference between incoming and outgoing light. High accuracy IToF sensors perform best in short range conditions of 30 meters or less.

###### Structured Light

Structured light systems utilize a calibrated pattern of infrared light projected onto the object. A 3D sensing camera then detects the distortion and intensity of the pattern to determine the relative distance and shape of the object. A computer algorithm is used to reconstruct the 3D surface. Some structured light systems use [phase shifted patterns](https://www.roboticstomorrow.com/article/2018/04/what-is-structured-light-imaging/11821) or IR dots rather than a square grid. Structured light 3D scanning provides high accuracy levels with lower power consumption than time of flight sensors.



## CHAPTER 5 FEATURES

The features of 3D human sensing encompass a wide range of capabilities that enable the detection, tracking, and interpretation of human movements and gestures in three-dimensional space. Some key features include:

1. **Depth Perception**: 3D human sensing systems can accurately perceive depth, allowing them to understand the spatial relationships between objects and individuals within a scene.
2. **Skeleton Tracking**: These systems can track the movements and positions of key points on the human body, often referred to as joints or keypoints, to reconstruct the skeletal structure and pose of a person.
3. **Gesture Recognition**: They can recognize and interpret various hand and body gestures, allowing users to interact with devices and applications using natural movements.
4. **Real-time Tracking**: 3D human sensing systems operate in real-time, providing instantaneous feedback and response to the user's movements.
5. **Multi-person Tracking**: Some systems have the capability to track multiple individuals simultaneously, distinguishing between different people and tracking their movements independently.

## CHAPTER 6 APPLICATIONS

Certainly, the most visible application of 3D sensing is in today's smartphones. Through facial

recognition, user-facing 3D scanning increases security, whereas world-facing 3D sensing opens up new possibilities for augmented reality and high-performance depth sensing photography. As the demand for 3D camera technology expands, so will the manufacturing volumes of illumination sources, LAS filters, and specialized diffusers. Here are some amazing applications of 3D sensors in various industrial verticals.

###### Automotive

Initially, the automotive sector appeared to be an odd consumer of 3D sensing technology. Since the advent of autonomous vehicles and advanced driver assistance systems (ADAS), facilitated by 5G and the IoT, 3D sensing has evolved into a critical component of road safety. LiDAR systems provide short and long-range 3D sensing, allowing vehicles to assess their surroundings automatically in real time.

###### Consumer Electronics

Biometric scanning and other innovative 3D sensing technologies are reshaping the consumer electronics industry. Laptops and tablets also used 3D sensors to enable motion sensing, augmented reality, virtual reality, and security functions. In addition, [Microsoft Kinect's gesture recognition capabilities](https://www.researchgate.net/publication/220774493_Gesture_recognition_using_Microsoft_KinectR), which transformed the home gaming industry, have now evolved to multi-player 3D position tracking, facial expression identification, and touchless heart rate monitoring.

###### Industrial

ToF and structured light technologies are used by agile robots operating in highly dynamic situations to quickly recognize work objects and barriers. With 3D sensing, automated warehouse order fulfilment can be accomplished on a gigantic scale while monitoring product damage and other non-conformances more effectively.

Humans, robots, and machinery working in confined areas can pose safety and security problems. 3D sensing technology is used in potentially sensitive environments such as power plants, sawmills, and data centers to identify harmful situations and distinguish persons based on facial traits. To avoid accidents, injuries, or unauthorized entrance, the facilities and equipment are automatically shut off.

## CHAPTER 7 ADVANTAGES AND DISADVANTAGES

### ADVANTAGES

* + - 3D Human Sensing is Fun Gaming and Virtual Reality.
    - 3D Human Sensing is Healthcare Help.
    - 3D Human Sensing is Enhanced Security and Surveillance.
    - 3D Human Sensing is 3D human sensing enables precise recognition and interpretation of gestures and movements.

### DISADVANTAGES

* + - 3d scheme is Costly Equipment.
    - Complex Setup and Calibration.
    - Limited Accuracy in Certain Conditions.

## CHAPTER 8 FUTURE WORK

The future of 3D human sensing holds promising avenues for research, development, and application. Some potential directions for future work include:

1. **Improved Accuracy and Precision**: Advancements in sensor technology, machine learning algorithms, and computer vision techniques can lead to even higher accuracy and precision in 3D human sensing. This includes better tracking of complex movements, finer details in skeletal estimation, and reduced errors in challenging environments.
2. **Enhanced Real-time Performance**: Future work may focus on optimizing algorithms and hardware to achieve even faster real-time performance in 3D human sensing applications. This could enable more responsive interactions in virtual environments, gaming, and other real-time applications.
3. **Multi-modal Sensing**: Integrating multiple sensing modalities, such as RGB cameras, depth sensors, inertial sensors, and microphone arrays, can provide richer information about human activities and interactions. Future research may explore methods for fusing data from these different modalities to improve robustness and accuracy in 3D human sensing.
4. **Semantic Understanding of Human Activities**: Beyond basic pose estimation and gesture recognition, future work may aim to develop systems that can understand the semantic meaning behind human activities and interactions. This could involve higher-level scene understanding, context-aware action recognition, and intention prediction based on observed behaviors.

## CHAPTER 9 CONCLUSION

We introduced human activity sensing algorithms from 3D video. Our algorithms cover (1) global kinematic structure, (2) complex motion estimation, and (3) detailed face and eye direction estimation. These are all non-contact sensing and do not require the object to use neither a special marker nor a costume. This is a clear advantage of our 3D video based sensing.

## CHAPTER 10 BIBLIOGRAPHY

1. T. Kanade, P. Rander, and P. J. Narayanan. Virtualized reality: Constructing virtual worlds from real scenes.

IEEE Multimedia, pages 34–47, 1997.

1. T. Matsuyama, X. Wu, T. Takai, and S. Nobuhara. Real-time 3d shape reconstruction, dynamic 3d mesh deformation and high fidelity visualization for 3d video. CVIU, 96:393–434, Dec. 2004.
2. S. Moezzi, L.-C. Tai, and P. Gerard. Virtual view generation for 3d digital video. IEEE Multimedia, pages 18–26, 1997.
3. J. Starck and A. Hilton. Surface capture for performance based animation. IEEE Computer Graphics and Applications, 27(3):21–31, 2007.