International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 05 | May 2020 www.irjet.net p-ISSN: 2395-0072

Advances in Depth Sensing Technologies for Computer Vision

Keerthan Bhat H[1], Prapulla S B[2]

[1]Student, Dept. of Computer Science & Engineering, RV College of Engineering
[2]Assistant Professor, Dept. of Computer Science & Engineering, RV College of Engineering

---***---

Abstract – *The modality of depth is very important in Computer Vision applications since it plays a huge role to helping machines see the world like humans do. Traditionally, only the 2D image of the 3D world was used but not anymore. Today, depth is being captured using a variety of technologies and each of it has its own benefits and disadvantages. This paper addresses that. The importance of depth, various depth sensing technologies and their comparison along with its applications has been discussed in this paper.*

Key Words: depth, computer vision, 3D world, stereo vision, structured light, time-of-flight, imaging

1.INTRODUCTION

This A typical camera covert the 3D world into a 2D image. The 3rd dimension of depth had been lost and felt as unimportant for a long time now. But now, with skyrocketing improvement of Computer Vision (CV) combined with Deep Learning (DL), many ambitious researchers have tried various methods to help machines see our world so that human tasks can be performed better with added machine vision capabilities. The current state-of-the-art Computer Vision techniques include handwriting recognition, object classification, semantic segmentation, face detection and a number of other things. In many of those tasks where 2D information is sufficient, CV algorithms have shown a lot of promise. However, when they deal with the real 3D world, researchers find that the 2D information is not sufficient. The third dimension of depth is required. Human beings have two eyes that enable us to sense depth naturally. However, most CV applications depend on one camera to capture and interpret the world around it. The lost third dimension of depth has to be captured to make machines see the 3D world more accurately.

Depth maps play an important role in perceiving the 3D world. In 3D computer graphics and computer vision, a depth map is formally defined as an image channel that contains information pertaining to the distance of the objects in the scene from a particular viewpoint. Each pixel of the depth map encodes the distance to the corresponding point in the scene. These images which gives us the estimation of the third dimension of the world around us, are very essential in understanding it and operating in it. This modality has many practical applications such as robotic navigation, 3D scene reconstruction, AR/VR, human-machine interaction and special imaging effects in modern smartphones to name a few.

Depth imaging can be done in many ways. The traditional approaches for sensing depth include stereo camera technology, structured light approaches and triangulation methods. Recent methods, which saw an increase in the adoption due to its promising results and robustness, is the Time-of-Flight (ToF) imaging and LiDAR methods. These cameras have seen huge usages in domains such as automotive safety, human-machine interfaces, gaming, smartphone cameras, robotics and autonomous vehicles. This paper provides a review of the various depth sensing techniques along with a comparison between them and also explains the applications of the depth modality in CV.

2. DEPTH SENSING TECHNIQUES

This section describes the prevalent 3D imaging technologies briefly.

1. *Stereo Vision*

The Stereo Vision uses multiple cameras placed side by side to capture multiple images of the same target and using these, the depth map is calculated based on the geometry. This setup essentially mimics the two eyes of humans (binocular vision), so generally two cameras are placed few millimeters apart. This is also known as Stereoview or Stereoscopy. The distance between the cameras creates a disparity between each of the positions in the images captured. Based on geometric calculations and matching techniques, the depth image is then obtained.

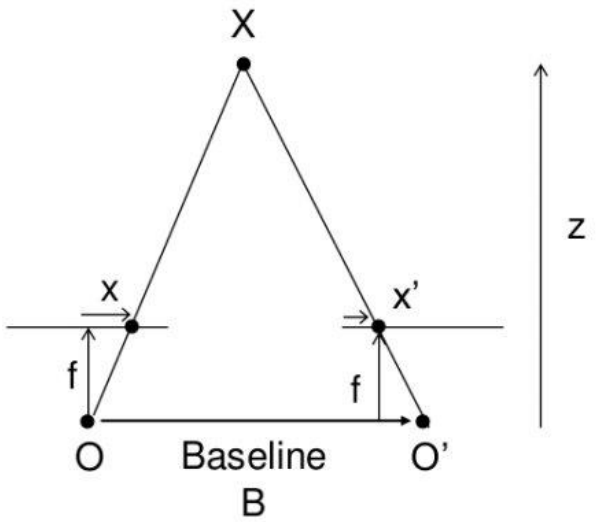

Fig 1: Disparity of X's image points (X and X') on left and right image planes

The main challenge is to find matching points in multiple images. The matching point searching involves complex CV algorithms. Deep Learning can find matching points with good accuracy but its computation cost is high and robustness is quite low. This is an example of passive depth sensing.

2. *Structured Light*

Structured Light method is an active depth sensing method where it uses a laser light source to project a known pattern and a receiver to detect the distortion of the reflected pattern and hence, calculates the depth map based on the geometry of the scene. The whole plane must be scanned to get the depth map which is time consuming yet highly accurate. However, this method is sensitive to environment brightness, so it's usually implemented in dark or indoor conditions.

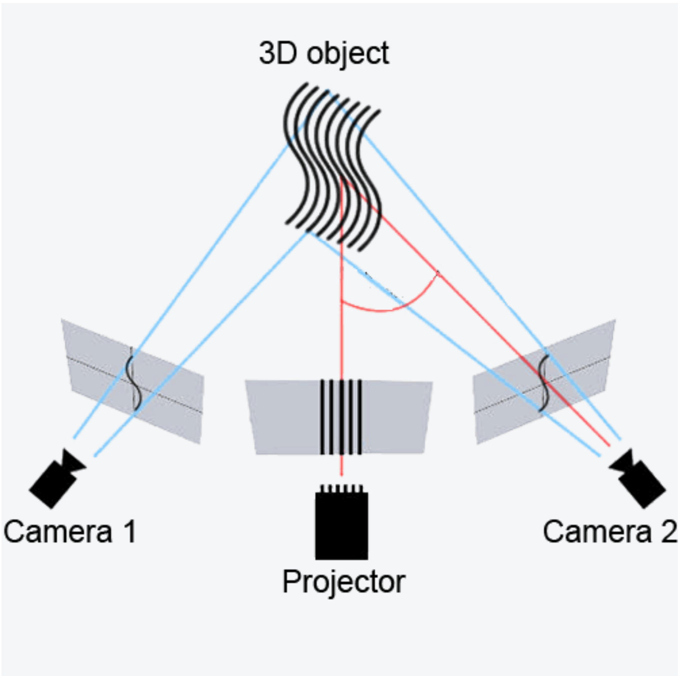

Fig 2: Structured Light method

3. *Triangulation Method*

Also called as Photogrammetry, it captures 3D measurements by pairing a laser illumination source with a camera. The laser beam and the camera are both aimed at the inspection target, however by adopting a known angular offset between the laser source and the camera sensor, it is possible to measure depth differences using trigonometry. It essentially determines the position of the point in 3D space given its projection on two images.

International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 05 | May 2020 www.irjet.net p-ISSN: 2395-0072

4. *LiDAR*

LiDARs use active sensors which emit light waves to illuminate the target object. They emit laser which is invisible to human eye and receive the reflected wave. Then, they derive the distance using the return time and wavelength of the laser. By rapidly scanning the target area point by point with laser pulses, a depth map of the scene can be determined. LiDARs can generate depth maps with high precision and high resolution. They are mainly used for applications like terrestrial mapping because they are generally very expensive and bulky.

5. *Time-of-Flight (ToF) Imaging*

ToF cameras work by measuring the phase-delay of the reflected infrared (IR) light from the scene objects. The IR wave is sent by an emitter to the scene and once it hits the object, the reflect wave is captured by the sensor. Using the phase delay between the emitter IR & reflected IR, the depth map is calculated. The working principle of this technique is similar to that of LiDAR; however, it should be noted that the phase delay is being used for depth calculation.

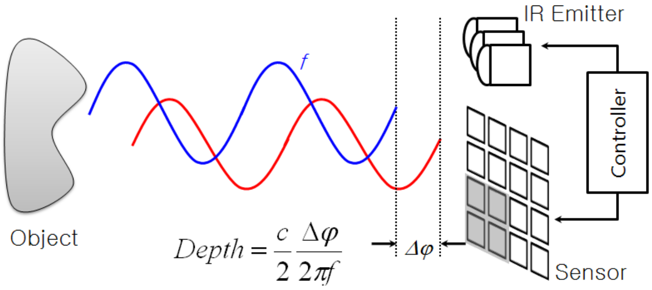

Fig 3: Time-of-Flight Imaging

3. COMPARISON

This section gives out a comparison between only the main depth sensing technologies. In the current day, we don't get to see triangulation methods used a lot as compared to the other techniques and LiDAR is quite synonymous with ToF in terms of principle of depth sensing with the only difference that LiDAR is bulkier and more expensive than ToF. Hence, we give a comparison between stereo vision, structured light and ToF.

Technology/ Criteria	STEREO VISION	STRUCTURED LIGHT	TIME-OF-FLIGHT
Principle of Operation	Two 2D cameras which emulates human eyes	Pattern illumination and distortion detection	Emits IR wave and measures phase delay
Active illumination	No	Yes	Yes
Operating distance	Less than 2m	0.2m - 3m	0.4m – 5m
Resolution	Medium	High	Low
Power consumption	Low	Medium	High
Frame rate	High	~30fps	Variable
Cost	Low	High	Medium
Post processing requirement	High	Medium	Low
Latency	Medium	Medium	Low
Low light performance	Weak	Good	Good
Bright light performance	Good	Weak	Medium Weak
Applications	Ranging, 3D reconstruction	Face recognition	AR/VR, autonomous vehicles

Table 1: Pros & cons of depth sensors

4. APPLICATIONS

AR/VR: A fundamental problem is Augmented Reality (AR) and Virtual Reality (VR) is to place an object in the 3D space such that its orientation, scale and perspective are properly calibrated. For this, depth is very essential. Pokémon Go is one of the most popular AR games. Major companies like Facebook and Microsoft sell VR headsets and applications. Today, there are many AR mobile apps such as measuring physical distances, AR emoji, virtual wearing of dresses & jewellery and visualizing home products inside the house before purchasing.

Autonomous Vehicles: Autonomous Vehicles (AV) are growing in demand right now and being used extensively in automating industries. It includes all kinds of robots, drones and self -driving cars. For navigation, localization, mapping and avoiding collision, it requires to see the 3D world and for this to happen, depth information is essential. With this, the AVs can estimate their trajectory of motion in the 3D world. Every major automobile company is investing huge money into these self-driving cars. Logistics companies use robots in their warehouses thereby eliminating the need for manpower. Robots find extensive applications in almost all domains such as education, healthcare, agriculture etc. We might see drone delivery too in the near future.

3D scene reconstruction: 3D imaging and scene reconstruction are important applications of the depth camera. Every pixel has to capture accurate depth along with higher frame rate and spatial resolution for precise 3D scene acquisition. For outdoor scenes, the daylight condition has to be measured. Therefore, range ambiguity matters in this application. Depth motion blur which might distort the 3D scene edges has to be removed before reconstruction.

Security cameras: Computer Vision techniques are increasingly popular today in security cameras which does facial recognition and action detection among many other things. Traditional face recognition algorithms only use 2D images of faces which might not be highly accurate since bad actors can fool the system easily. Whereas a face recognition system which uses depth data also would capture more features of the face and detect it accurately. Same goes for action detection and gait analysis applications.

Gaming: Gesture and proximity detection are being used in high end gaming consoles which require humans to actually move and play the game. For example, tennis games require players to actually swing the hand and mimic playing real tennis. This gesture should be captured by the system. 3D information plays a huge role in actively performing these applications and giving us great gaming experience. Microsoft's Kinect is the best example for these types of gaming console.

5. CONCLUSION

Depth sensing systems combined with current Computer Vision applications have enhanced performance and meets all requirement for real-life deployment. We discussed the various depth sensing technologies and compared the pros and cons of it along with its major applications. Depth sensing is a huge market today and many researchers, companies are investing a lot in it because there is still a lot of room for improvement and making the systems more accurate. Along with these technologies, depth estimation using only the RGB data is also performed in many systems however this is a software solution for depth. Structure from Motion (SfM) and Simultaneous Localization and Mapping (SLAM) are methods also widely used in robotics.

REFERENCES

[1] https://blog.cometlabs.io/depth-sensors-are-the-key-to-unlocking-next-level-computer-vision-applications-3499533d3246
[2] https://www.revopoint3d.com/comparing-three-prevalent-3d-imaging-technologies-tof-structured-light-and-binocular-stereo-vision/
[3] https://www.iotforall.com/choosing-3d-vision-camera/
[4] S. Lee, "Depth camera image processing and applications," 2012 19th IEEE International Conference on Image Processing, Orlando, FL, 2012, pp. 545-548, doi: 10.1109/ICIP.2012.6466917.
[5] https://docs.opencv.org/3.4/dd/d53/tutorial_py_depth map.html