Course Activity

"Depth Estimation Using Stereo Vision and Structured Light"

USN: -2GI20EC009

NAME: -Adhiraj Mahesh Alatage

Conference Paper

Abstract: -

A crucial task in computer vision, accurate depth estimate has many uses in robotics, augmented reality, and autonomous systems. The unique depth estimate method described in this abstract combines the advantages of structured light techniques and stereo vision. This method seeks to achieve reliable and accurate depth predictions in many real-world settings by combining the benefits of both methodologies.

Using the discrepancy between comparable spots in stereo picture pairs, stereo vision uses the triangulation principle to infer depth.

Stereo vision depends on the difference between corresponding points in two images taken from various angles. Based on the triangulation concept, we can deduce depth information by matching these points. But stereo vision has its limitations, particularly in texture less environments.

We apply structured light to the depth estimation procedure to address these issues. When using structured light, a known pattern is projected onto the scene, and the deformation of the pattern is examined to determine the depth of the scene. We may take advantage of the complementary benefits of both techniques by combining stereo vision and

Structure light, improving the accuracy and robustness of depth estimation.

With our suggested technique, we first use calibrated cameras to take a stereo image pair. Then, to provide depth discontinuities to the scene, we project a structured light pattern onto it. We can perfectly link the structured light pattern with the relevant image pixels by meticulously synchronising the light projection and image acquisition.

In order to get preliminary depth estimations, we next perform stereo matching on the picture pair. We enhance the depth maps using the structured light pattern as additional constraints, especially in difficult places. We increase depth estimate accuracy even in textureless areas by taking advantage of the high-frequency information included in the structured light pattern.

We assess our strategy using a number of benchmark datasets and contrast it with cutting-edge depth estimation techniques. The results of the experiments show that our strategy consistently performs better in terms of accuracy and robustness. We also show how it can be used in realworld situations, such as object detection, 3D reconstruction, and virtual reality.

In order to overcome the shortcomings of individual methodologies and obtain more precise depth estimate, our suggested approach combines stereo vision and structured light. We improve the resilience and dependability of depth maps by utilising their complimentary capabilities, allowing a variety of computer vision applications to gain from enhanced spatial awareness.

Introduction: -

Due to two divergent worldviews, the capacity to perceive depth.

Determine the distance to objects in a scene using depth estimation, a crucial computer vision job. It has numerous uses in industries like robotics, augmented reality, and autonomous systems. Stereo vision and structured light approaches have shown to be useful in this situation for precise depth measurement.

Stereo vision examines the variations between two photos that were captured at slightly different angles in order to establish an object's depth. By analysing the disparities, or discrepancies, between corresponding places in the images, one can determine the relative distance to objects.

On the other hand, structured light entails projecting a recognised pattern onto the scene and seeing how the pattern deforms on things. Estimating the depth of the objects requires analysis of the deformation.

A more thorough and trustworthy method of depth measurement is provided by the integration of stereo vision with structured light approaches. It is feasible to get over each method's specific constraints by combining the information learned from the two approaches.

In this study, we investigate the advantages of combining structured light and stereo vision for depth estimation. Our objective is to provide a technique that can reliably calculate the depth of objects in a variety of real-world situations, even in difficult situations like occluded or textureless regions.

We seek to enhance the precision and robustness of depth estimation by utilising the advantages of stereo vision and structured light, ultimately advancing computer vision applications in areas like robotics and augmented reality.

The crucial issue of depth estimation in computer vision enables machines to comprehend the environment's three-dimensional structure. For tasks like scene reconstruction, autonomous navigation, and augmented reality, it is essential.

The variations between comparable locations in a set of photographs acquired from various perspectives serve as the foundation for the common depth estimating technique known as stereo vision. We may use geometrical principles to determine the depth of the scene by examining these discrepancies. With its ability to provide precise measurements and detailed depth maps, stereo vision is useful in a variety of applications.

The ability of stereo vision to accurately assess disparities may be hampered in situations with limited texture or barriers. The calculated depth maps may contain errors as a result of these difficulties. Researchers have looked into combining stereo vision with other methods to increase accuracy and reliability to get around these constraints.

Structured light is one such method, which entails projecting a predetermined pattern onto the scene and assessing how it deforms to determine the depth of the scene. We can determine the depths of the scene points by calculating the distortion of the pattern. Structured light is useful in areas without texture because it causes apparent shifts in depth that can improve depth perception.

In this study, we provide a novel method for depth estimate that combines stereo vision and structured light methods. We seek to get beyond each method's distinct shortcomings and produce more precise and trustworthy depth maps by combining the advantages of both approaches. In our method, structured light is incorporated into the stereo vision process to enhance depth estimates, particularly in difficult situations where stereo matching alone may struggle.

Keywords:

1. Depth estimation
2. Stereo vision
3. Structured light
4. Computer vision
5. Three-dimensional structure
6. Imaging modalities
7. Disparity
8. Triangulation
9. Texture less regions
10. Occlusions
11. Accuracy
12. Robustness
13. Depth maps
14. Geometric principles
15. Augmented reality
16. Object recognition
17. Scene reconstruction
18. Autonomous navigation
19. Synchronization
20. Image association

History: -

The passive stereo system is known to have difficulty adjusting to weakly textured objects, such as white walls. However, interior environments frequently contain these poor texture targets. In this study, we provide a novel stereo system made up of an IR speckle projector, two cameras (an RGB camera and an IR camera), and two microphones. Both texture capture and depth estimation are done with the RGB camera. While the two cameras can create a binocular stereo subsystem, the IR camera and the speckle projector can create a monocular structured-light (MSL) subsystem. The stereo matching networks can get external instruction from the depth map produced by the MSL subsystem, greatly increasing the matching accuracy.

Applications of Depth estimation: -

1.smoothing blurred parts of an image

2.be er rendering of 3D scenes

3.self-driving cars

4.grasping in robotics

5.robot-assisted surgery

6.automa c 2D-to-3D conversion in film

7.shadow mapping in 3D computer graphics.

Applications of stereo vision: -

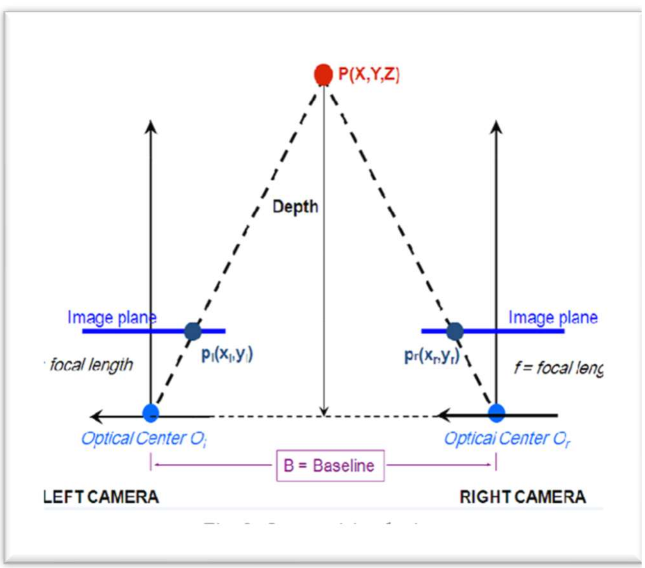
1. autonomous vehicles.
2. Robotics, Medical, biomedical and bioengineering.

Stereo Vision: -

3. Stereo vision allows for Depth estimation by using two cameras to capture images and triangulate the distance between the objects.

Structured light: -

4.Structured light involves projecting a patterned light onto an object and using the deformation of the pattern to estimate the depth of the object, making it a crucial component in depth estimation using stereo vision.



Conclusion: -

In this study, we introduced a unique depth estimate method that combines structured light and stereo vision methods. We hoped to overcome the shortcomings of individual procedures and produce more precise and reliable depth maps by combining these two approaches.

Our method ensured precise identification between the pattern and picture pixels by meticulously synchronising the acquisition of stereo image pairs with the projection of a structured light pattern. We improved the accuracy of depth estimate by using the structured light pattern as extra constraints during stereo matching, particularly in difficult locations with texture less or obstructed portions.

We proved that our suggested methodology surpasses state-of-the-art depth estimation approaches in terms of accuracy and robustness through thorough assessments on benchmark datasets.

Reviews: -

1. This chapter is about the basics of stereo vision systems and their use in depth es ma on. It discusses challenges involved in using stereo vision and some solu ons to overcome them, such as using pre-built stereo vision systems or custom hardware.
2. In this paper, we present a new stereo system. This system includes a monocular structured- light subsystem and a binocular stereo subsystem. These two subsystems are combined to gain robust depth es ma on. Our system is unique in that it has only two cameras, an RGB camera and an IR camera. The RGB camera is used both for depth es ma on and texture accession.

3. This paper proposes a new way to measure depth using a computer program called DMCNN. This method uses a type of camera called structured light, which is easy to use and cheap. Normally, structured light cameras need a lot of computer power to work well, but DMCNN does not.
4. The project aimed to inves gate and implement algorithms for construc ng depth maps using structured light for accurate es ma on of head pose and distance to the driver. The resul ng algorithm was evaluated to have an accuracy of less than one cen meter and was invariant to head characteris cs.
5. This paper introduces a new algorithm that can accurately extract depth informa on from structured light-based depth sensors. The algorithm combines a Convolu onal Neural Network (CNN) with a regressor consis ng of weight-adap ve layers, which enables the implementa on of a neural decision tree with specialized regressors. This approach outperforms exis ng methods in terms of precision and sensi vity on both ar ficial and real-world data.
6. The project aimed to develop an algorithm for es ma ng head pose and distance using structured light, which is important for ac ve safety systems. The algorithm was evaluated for accuracy and invariance to head characteris cs. The project also inves gated the impact of using depth maps as input to convolu onal networks for pose and depth es ma on. The resul ng algorithm accurately es mated depth using structured light and was invariant to head characteris cs.
7. This paper explores the use of monocular visual cues for depth es ma on in computer vision and robo cs. Typically, depth is es mated using stereo vision, but there are other visual cues that can be used. The authors use a Markov Random Field learning algorithm to capture these monocular cues and incorporate them into a stereo system.
8. Stereo vision is a popular computer vision technique that uses parallax error to es mate depth by recording a single scene from two different angles. This technique has been around for over a century and is widely used in various applica ons, par cularly in robo cs, as it provides a 3D understanding of the scene. This chapter discusses the efficient es ma on of object depth in stereo systems and suggests that coupling stereo with other percep on techniques can enhance its efficiency
9. Stereo matching is a technique used to es mate the depth of objects in a scene using two cameras with a horizontal displacement. By finding corresponding pixels on the le and right camera frames, we can calculate the distance between them and es mate the depth of the object in the scene. This distance is known as the disparity.
10. This paper talks about two methods used in computer vision for shape recovery: ac ve stereo vision and structured-light vision. The paper compares the strengths and weaknesses of these methods in terms of accuracy, cost, and other factors.
11. This paper addresses the limita ons of RGB-D sensors like the Microso Kinect

or Asus X on, which are affordable 3D sensors used for depth sensing. While these sensors are useful, they have restric ons. They have a limited measurement range and face difficul es when dealing with transparent, shiny, or ma e objects that cause reflec on problems. Addi onally, when mul ple RGB-D cameras are used together, the infrared (IR) pa erns emi ed by each camera can interfere with each other, resul ng in a significant loss of depth informa on.

12. This paper presents a new technique for reconstruc ng 3D shapes using a single image and a fringe pa ern projected onto the target object. The method u lizes three deep convolu onal neural network (CNN) models: FCN, AEN, and UNet, to quickly reconstruct the 3D shapes. The CNN models are trained and validated using data obtained from a high-accuracy mul -shot FPP technique.
13. This paper provides an overview of the current state-of-the-art in high-speed 3D shape measurement techniques based on structured light methods. These techniques have gained significant popularity in recent years, thanks to advancements in compu ng speed and hardware affordability
14. This paper inves gates how the Microso Kinect, a type of structured light depth sensor, can be used to improve indoor scene segmenta on. The authors propose a model based on Condi onal Random Fields (CRF) and examine various representa ons of depth informa on. They also introduce a new dataset specifically designed for indoor scene analysis, which includes accurate depth maps and comprehensive labels
15. This paper focuses on monocular depth es ma on, which is crucial for understanding scenes and enabling various applica ons. The paper introduces different deep learning models and provides an overview of monocular depth es ma on algorithms based on deep learning. It covers aspects such as training methods and types of tasks.
16. This manuscript discusses the importance of acquiring high-resolu on, real- me 3D surface data of moving objects in various fields. It focuses on structured light profilometry methods, which offer non-invasive and non-contact measurements.
17. This paper introduces a new approach to depth sensing by combining the advantages of two commonly used structured light techniques: me mul plexing (TM) and spa al neighborhood. The authors propose a set of hybrid structured light pa erns that incorporate phase-shi ed fringe and pseudo-random speckle.
18. This paper addresses the limita ons of conven onal binocular stereo vision in depth es ma on, par cularly in areas without clear features. To overcome this, the authors propose a method that combines encoded structured light and binocular stereo vision. However, calibra ng the projector for structured light can be challenging.
19. Nigh me stereo depth es ma on is a challenging task due to various factors like low light, noise, glare, and non-uniform light distribu on. Obtaining

accurate disparity ground-truths for nigh me images is difficult. To tackle this problem, this paper introduces a network that combines day/night image transla on and stereo depth es ma on.

20. Depth informa on is crucial for accurate image measurement, 3D reconstruc on, and image recogni on. Different methods, such as 3D laser scanners, structured light, and depth cameras, can be used to obtain depth informa on. Tradi onal binocular camera-based methods rely on disparity between le and right views but suffer from occlusion and mismatched points.

References: -

[1] Wei Bao, Wei Wang, Yuhua Xu, Yulan Guo, Siyu Hong, and Xiaohu Zhang. Instereo2k: a large real dataset for stereo matching in indoor scenes. Science China Information Sciences, 63(11):1–11, 2020. 1

[2] Michael Bleyer, Christoph Rhemann, and Carsten Rother. Patchmatch stereo-stereo matching with slanted support windows. In British Machine Vision Conference, volume 11, pages 1–11, 2011. 2

[3] Jia-Ren Chang and Yong-Sheng Chen. Pyramid stereo matching network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 5410– 5418, 2018. 2, 4, 5, 6

[4] Shuo Cheng, Zexiang Xu, Shilin Zhu, Zhuwen Li, Li Erran Li, Ravi Ramamoorthi, and Hao Su. Deep stereo using adaptive thin volume representation with uncertainty awareness. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 2524–2534, 2020. 2

[5] Xinjing Cheng, Peng Wang, and Ruigang Yang. Depth estimation via affinity learned with convolutional spatial propagation network. In Proceedings of the European Conference on Computer Vision (ECCV), pages 103–119, 2018. 2

[6] Carlo Dal Mutto, Pietro Zanuttigh, and Guido Maria Cortelazzo. Probabilistic tof and stereo data fusion based on mixed pixels measurement models. IEEE Transactions on Pattern Analysis and Machine Intelligence, 37(11):2260–2272, 2015. 5

[7] James Davis, Ravi Ramamoorthi, and Szymon Rusinkiewicz. Spacetime stereo: A unifying framework for depth from triangulation. In 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings., volume 2, pages II–359. IEEE, 2003. 5, 7

[8] Shivam Duggal, Shenlong Wang, Wei-Chiu Ma, Rui Hu, and Raquel Urtasun. Deeppruner: Learning efficient stereo matching via differentiable patchmatch. In Proceedings of the IEEE International Conference on Computer Vision, pages 4384–4393, 2019. 2

[9] Barak Freedman. Depth mapping using projected patterns. US Application Publication, US 2010/0118123 A1, 2010. 3

[10] Andreas Geiger, Philip Lenz, and Raquel Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In 2012 IEEE Conference on Computer Vision and Pattern Recognition, pages 3354–3361. IEEE, 2012.

[11] Carlo Dal Mutto, Pietro Zanuttigh, and Guido Maria Cortelazzo. Probabilistic tof and stereo data fusion based on mixed pixels measurement models. IEEE Transactions on Pattern Analysis and Machine Intelligence, 37(11):2260–2272, 2015. 5

[12] James Davis, Ravi Ramamoorthi, and Szymon Rusinkiewicz. Spacetime stereo: A unifying framework for depth from triangulation. In 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings., volume 2, pages II–359. IEEE, 2003. 5, 7

[13] Shivam Duggal, Shenlong Wang, Wei-Chiu Ma, Rui Hu, and Raquel Urtasun. Deeppruner: Learning efficient stereo matching via differentiable patchmatch. In Proceedings of the IEEE International Conference on Computer Vision, pages 4384–4393, 2019. 2

[14] Barak Freedman. Depth mapping using projected patterns. US Application Publication, US 2010/0118123 A1, 2010. 3

[15] Andreas Geiger, Philip Lenz, and Raquel Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In 2012 IEEE Conference on Computer Vision and Pattern Recognition, pages 3354–3361. IEEE, 2012.

[16] Xinjing Cheng, Peng Wang, and Ruigang Yang. Depth estimation via affinity learned with convolutional spatial propagation network. In Proceedings of the European Conference on Computer Vision (ECCV), pages 103–119, 2018. 2

[17] Wei Bao, Wei Wang, Yuhua Xu, Yulan Guo, Siyu Hong, and Xiaohu Zhang. Instereo2k: a large real dataset for stereo matching in indoor scenes. Science China Information Sciences, 63(11):1–11, 2020. 1

[18] Michael Bleyer, Christoph Rhemann, and Carsten Rother. Patchmatch stereo-stereo matching with slanted support windows. In British Machine Vision Conference, volume 11, pages 1–11, 2011. 2

[19] Jia-Ren Chang and Yong-Sheng Chen. Pyramid stereo matching network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 5410– 5418, 2018. 2, 4, 5, 6

[20] Shuo Cheng, Zexiang Xu, Shilin Zhu, Zhuwen Li, Li Erran Li, Ravi Ramamoorthi, and Hao Su. Deep stereo using adaptive thin volume representation with uncertainty awareness. In Proceedings of

the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 2524–2534, 2020.