Stereo Vision: An Introductory Approach

Binsan Khadka
Kathmandu University
Dhulikhel, Kavre.
00977 1 4482747
binsan_khadka@yahoo.com

Digdarshan Lal Dhonju
Kathmandu University
Dhulikhel, Kavre.
00977 1 4271384
dhonjudigdarshan@hotmail.com

Amol Shrestha
Kathmandu University
Dhulikhel, Kavre.
00977 21 524509
amolshrestha@yahoo.com

ABSTRACT

A camera creates images that are projections of some limited part of our 3-Dimensional world onto a 2-Dimensional plane. If we can somehow reverse this projective transformation, we can recover information about the 3-Dimensional world from images. Unfortunately, a single image does not contain sufficient information to invert the projection. There is ambiguity because a given feature in the image could correspond to a large distant object or a small nearby object. This ambiguity can be resolved using multiple views of the scene. We will describe here one specific technique in which only a pair of image is taken for the creation of three-dimensional model, therefore called Stereo Vision.

In this paper, we are going to give a brief account of contemporary research works, the developments done so far and about what can be expected of Stereo Vision in the future.

Keyword(s)

Stereo Vision.

1. BACKGROUND

Human beings generally come equipped with two eyes located side-by-side in the front of their heads. The close side-by-side positioning of each eye makes it possible to take a view of the same area from a slightly different angle. The two eye views have plenty in common, but each eye picks up visual information the other doesn't. [19]

Each eye captures its own view and the two separate images are sent on to the brain for processing. When the two images arrive simultaneously in the back of the brain, they are united into one picture. The mind combines the two images by matching up the similarities and adding in the small differences. The small differences between the two images add up to a big difference in the final picture. The combined image is more than the sum of its parts. It is a three-dimensional stereo picture.

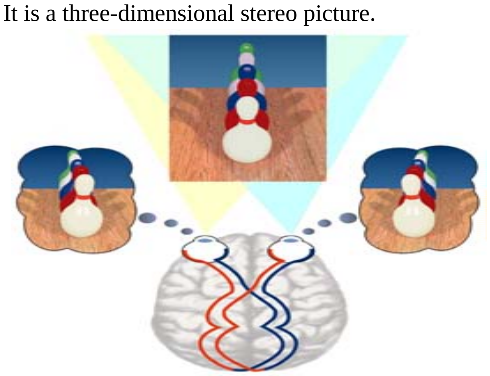

Figure 1: Human eye using a pair of images to interpret correct vision. (Ref: http://www.vision3d.com/stereo.html)

The word "stereo" comes from the Greek word "stereos" which means firm or solid. With Stereo Vision we see an object as solid in three spatial dimensions: width, height and depth or x, y and z. It is the added perception of the depth dimension that makes Stereo Vision so rich and special.

Stereo Vision or stereoscopic vision probably evolved as a means of survival. With Stereo Vision, we can see where objects are in relation to our own bodies with much greater precision, especially when those objects are moving toward or away from us in the depth dimension. We can see a little bit around solid objects without moving our heads and we can even perceive and measure "empty" space with our eyes and brains.

Here are a few examples of occupations that depend heavily on Stereo Vision: Baseball player, Waitress, Driver, Architect, Surgeon, Dentist. Here are just a few examples of general actions that depend heavily on Stereo Vision:

- Throwing, catching or hitting a ball.
- Driving and parking a car.
- Planning and building a three-dimensional object.
- Threading a needle and sewing.
- Reaching out to shake someone's hand.
- Pouring into a container.
- Stepping off a step.

The different perspectives of our two eyes lead to slight relative displacements of objects (disparities[1]) in the two monocular views of scene. The human visual system is able to

- Use these disparities for depth-estimation.
- Merge both monocular views into a fused cyclopean view of the scene. [Figure 1]

Stereo Vision is one of the interesting concepts under the vast topic of Computer Vision. The term Computer Vision has been previously identified as Digital Image processing. Digital image processing is not a new phenomenon; techniques for the manipulation, correction and enhancement of digital images have been in practical use for over forty years – an early application being the removal of defects from images obtained by NASA's unmanned lunar probes. We don't have to look very far these days to see an example of image processing at work. It has insinuated itself into many different areas of human endeavor, ranging from small-scale activities such as desktop publishing and healthcare, through to activity on the largest scales imaginable: the search for natural resources on Earth, or the study of other planets, stars and galaxies in our universe.

[1] The word disparity is explained in section 5.2.

This paper provides a theoretical introduction to Stereo Vision, avoiding mathematical detail and focusing more on the introductory approach and research direction of the subject so that it is aimed at a broader audience, but is likely to appeal most strongly to the computer enthusiast with some programming experience, or to those in an undergraduate-computing course. In Nepal, the concept of Stereo Vision is still in the dark phase. Our intent is that the paper should both introduce the basic concepts and provide the computer-literate reader with the minimum available resources as the means of initial step to research or experiment with those concepts in Nepal.

2. INTRODUCTION

As the name suggests, Stereo Vision refers to the ability to infer information on the 3D Structures and the distances of a scene from at least two images (Left and right), taken from different viewpoints. Stereo Vision utilizes the slightly different views of a scene, projected on to the right and left images, to recover depth information, which is known as Binocular Stereo Fusion. From a generic point of view, in both humans and machines, the problem reduces to a matching of the two views, in order to find the displacement (disparity) of corresponding patterns of the projected images. Thus, A Stereo system must solve two essential problems – correspondence and reconstruction [3] explained later in this paper.

Stereo Vision has long been researched, and a number of computational approaches, which include feature-based, area-based and phase-based methods, have been proposed [4][5][6]. One early research done in the field of Stereo Vision was by Marr and Poggio [1]. All these methods have their intrinsic problems, caused by the very assumptions inherent in these approaches.

The use of automated machines in the study of other planets has emphasized the use of Stereo Vision where a precise depth perception is needed to detect and avoid obstacles to guide the route itself further, collect samples to be examined etc. Even the invention of automated robots, toys need the use of Stereo Vision to perform all the task as mentioned above like, to reach out and shake hands, pour into the container, grab a cup and serve, take a step, and every human like activities.

2.1 STEREO VISION: A USUAL PARADIGM

The usual Paradigm for Stereo Algorithms includes the following steps: [7]

- **i.** Features[2] are located in each of the two images independently.
- **ii.** Features from one image are matched with features from the other image. That is, for every feature in the left image corresponding to a certain point in the scene, a feature must be found in the right image such that it corresponds to the projection of the same scene point[3].
- **iii.** The disparity between features is used, together with the parameters of the imaging geometry (i.e. relative separation and orientation of the cameras), to determine the distance to the corresponding point in the scene.

The resulting depth points are often sparse whereas depth must be computed at every point in the scene. Thus, the depth points are interpolated to obtain a surface, or a complete depth map.

Stereo Vision has been intensively investigated, and a number of computational models have been proposed [4][5][6] in human vision– to explain psychophysical and neurophysiological data, & in machine vision– to solve problems such as reconstructing 3D shapes, understand motion and moving shapes. One of the Early Research done in the field of Stereo Vision was by Marr and Poggio [1][16]. To extract disparity information from a stereo image pair, Marr and Poggio worked to develop a cooperative algorithm. These types of algorithms are a popular means of solving problems that seem computationally difficult, but our brains can cope to handle it with ease.

According to Marr & Poggio, three steps are required to solve the problem of measuring stereo disparity. These steps are the one listed above. This solution is based on two constraints (C), which are each mapped to two rules that relate to how the information of an image pair is combined:

C1. At a point in time, a point in space has a unique position.

Rule 1: Each item from each image may be assigned at most one disparity value.

C2. Matter is cohesive, and it is separated into different objects. The surfaces of objects are generally smooth relative to their distance from the viewer

Rule 2: Disparity varies smoothly almost everywhere.

3. APPLICATION OF STEREO VISION

Stereo Vision offers a wide range of application areas as in:

3.1 Autonomous Robots, Vehicles

3D object detection, location and depth perception calculated from the stereoscopic view of objects by the pair of cameras located on the robotic body can be used for the mobilization of the robot or robotic vehicle. The calculated results from the stereoscopic views can be used to detect and avoid the obstacles and take a safe path of action.

3.2 Industrial Inspection and Quality Control

Stereoscopic vision can be used to detect the cracks, damage, unusuality in goods, raw materials or products, during manufacturing process.

3.3 3D Object Location and Detection

The two stereoscopic views of an object can be reconstructed to obtain a 3D object which can be recognized or detected from a known database and depth calculation can determine its location in the 3D space.

3.4 Virtual Reality

3D illusion can be created by matching & fusing two different stereoscopic views of objects, scenes, etc. 3D Animations, game play station, interactive 3D virtual world are some of explorations in virtual reality.

3.5 Road Monitoring/Traffic Lights

Road monitoring and traffic light management can be really dynamic rather than the existing static approach. The frequency and duration of traffic light alteration can be dynamically determined by the amount of load in the street using Stereo Vision.

---

[2] Feature refers to the intensity value of a pixel, edge, subset of an image or anything that defines an image.

[3] This leads to correspondence problem, explained in section 5.1.

3.6 Military Application

With and intension to target with precision heavy loss in enemy & few loss in offending side is key to military actions. Use of stereoscopic views to precisely calculated depth, detected & located enemy targets, auto pilot or remote controlled vehicles, air-ships can have wide use of Stereo Vision

3.7 Bio-Medical

The Stereo Fusion of stereoscopic views can help explain the psychophysical and neurophysiological data. In the field like hair transplant, the operation procedure requires great care and accuracy. There is a widely used product “Mantis Stereo viewing system” produced by vision engineering [17] is used for this purpose. Another product of same group is “Lynx Eyepieceless Stereo Microscope” which is providing highly sophisticated product to make angioplasty heart surgery a safer and less complicated procedure.

4. ISUES AND CHALLENGES

Some of the important issues that can be raised during the study and implementation of Stereo Vision are discussed below.

- Digital Image Processing and use of complex mathematical formulae in the algorithms requires great deal of computation, thus consuming more time. Use of hardware with high capacity is recommended for better performance.
- Identification of suitable algorithms to implement is a challenging job, where issues of performance, correctness [2], and restrictions have to be considered.
- If two (or more) images are available, then the three-dimensional point can be obtained as the mapping of the corresponding points of the two images. A number of things are needed for this
    - Corresponding image points
    - Relative pose of the camera for the different views
    - Relation between the image points and the corresponding line of sight

  Note, however that the calculation of disparity is not an easy task to attain.
- The correct and fast estimation of disparities is a difficult problem. Besides disparities, various additional image variations occur between the left and right view of a scene. Differences might be caused by occlusions of objects, specular reflections, which move independently of the surfaces of objects, sensor noise, and various other causes.
- The relation between an image point and its line of sight is given by the camera model (e.g. pinhole camera) and the calibration parameters. These parameters are often called the intrinsic camera parameters while the position and orientation of the camera are in general called extrinsic parameters. How can all these elements can be retrieved from the images. The key for this are the relations between multiple views which tell us that corresponding sets of points must contain some structure and that this structure is related to the poses and the calibration of the camera. [15]
- Another type of problems is caused when the imaging process does not satisfy the camera model that is used. Some times radial distortion is present in the image. This means that the assumption of a pinhole camera is not satisfied. It is however possible to extend the model to take the distortion into account. However, sometimes image is much harder to use when important part of the scene is not in focus. And the problem becomes more prominent when blooming is present (i.e. overflow of CCD-pixel to the whole column). Most of these problems can however be avoided under normal imaging circumstance.
- Implementation of stereo system in the real time system must consider different affecting factors like motion of the scene, varying light intensity, appearance changes of models, complex natural objects, presence of noise (unwanted feature that hinders the real one) etc.
- Note that different viewpoints are not the only depth cues that are available in images. Shading, Shadows, Symmetry, Texture and focus also give some hints about depth or local geometry, but considering those cues add an extra overhead.

5. IMPORTANT TERMINOLOGIES

5.1 Correspondence Problem

In Stereo Algorithm, feature of two or more images are extracted and a comparison between the two is done. The difficult part is to match the particular features in the two images, which is known as Correspondence Problem.

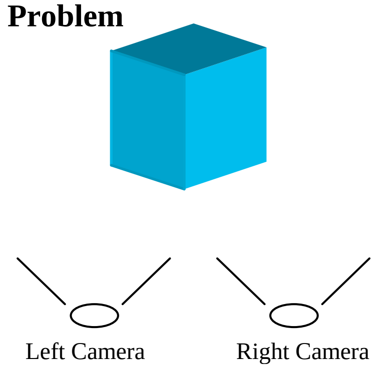

Figure 2: Camera orientation of Stereo Vision.

The Edges of a cube as seen by two camera [Figure 2] looks like the images shown in Figure 3. Looking at the image and

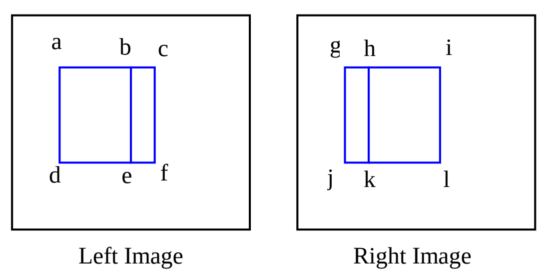

Figure 3: Edges of Cube seen from the two cameras.

orientation of camera, for human it is easy to say which point of left image corresponds on the right image. But for computer, it is difficult; for example to correspond points of edge ‘b-e’ with points of edge ‘h-k’ is difficult. Heiko has proposed one way to improve the solution for this problem. [2]

5.2 Disparity

Disparity refers to the separation of image points of a scene taken from two cameras located at horizontal separation. There is an inverse relationship between disparity and depth in the scene; disparity will be relatively large for points in the scene that are near to camera and relatively small for the points that are far away. For example in Figure 3, the edge 'b-e' of left image is near to camera as compared to edges 'a-d' and 'c-f'. Hence, the displacement of the corresponding edge of 'b-e', i.e. 'h-k' in right image is greater with respect to left image.

5.3 Occlusion Detection

For each point on the left image, there may not be corresponding match with the right image. Occlusion detection involves in finding such unmatched points. Consider for example, the leftmost side of a left image is not seen in the right image, an attempt to match the corresponding points of such image gives an ambiguous result. Thus, these regions have to be explicitly detected. There are several methods to explicitly detect occlusions, like methods using intensity edge [9], multiple cameras with camera masking [10] and bi-directional matching [8]. Recently, several stereo algorithms, Zitnic & Kanade [6], Geiger, Ledendrof & Yuille [11], Intille & Bobick [12] and Belhumeur and Mumford[13] have proposed finding occlusions & matches simultaneously to help in identifying disparity discontinuities.

5.4 Reconstruction

Reconstruction recovers the depth of point P, using the estimated disparity and a model of the stereo rig specifying the pose of each camera and its optical parameters. The problem of reconstruction is preceded by the camera calibration, which is the measurement of camera model parameters in the problem on its own.[14]

6. CONCLUSION

Researchers have been investigating methods to acquire 3D information from objects and scenes for many years. In the past the main applications were visual inspection and robot guidance. Nowadays however the emphasis is shifting. There is more and more demand for 3D models in computer graphics, virtual reality and communication. This results in a change in emphasis for the requirements. The visual quality becomes one of the main points of attention. Therefore not only the position of a small number of points have to be measured with high accuracy, but the geometry and appearance of all points of the surface have to be measured. This requires an extensive use of Stereo Vision.

The sole intention of the paper is to enrich individuals, students, researchers, faculties, etc with information regarding the introduction to Stereo Vision. This paper can be used for study, information, research, or implement the achievements in further projects.

REFERENCES

1. D. Marr & T. Poggio (November 1997), "A Theory Of Human Stereo Vision", MIT AI Lab, AI Memo No. 451.
2. Heiko Hirschmuller, 'Improvements in Real-Time Correlation-Based Stereo Vision", De Montfort University, Leicester.
3. E. Trucco, A. Verri, "Introductory Techniques for 3D computer Vision", Prentice Hall, 1998.
4. William Hoff, Narendra Ahuja, "Depth From Stereo", University of Illinois, USA.
5. William Hoff, Narendra Ahuja, " Surfaces From Stereo: Integrating Feature Matching, Disparity Estimation, and Contour Detection", IEEE Transaction on pattern analysis and machine intelligence, Vol 11, No. 2, Feb 1989, (pp 121 – 128)
6. C. Lawrence Zitnic, Takeo Kanade, " A cooperative Algorithm for Stereo Matching and Occlusion Detection", The Robotics Institute, Carnegie Mellon University, Pittsburgh, Pennsylvania October 1999.
7. William Hoff, Narendra Ahuja, "Extracting Surfaces from Stereo Images: An Integrated Approach", Coordinated Science Laboratory, University of Illinois, USA.
8. P. Fua, "A parallel stereo algorithm that produces dense depth maps and preserves image features," Machine Vision and Applications, vol. 6, pp. 35-49, 1993.
9. S. Intille and A. Bobick, "Incorporating intensity edges in the recovery of occlusion regions," International Conference on Pattern Recognition, Jerusalem, Israel, Vol. I, pp. 674-677, October 1994.
10. Y. Nakamura, T. Matsuura, K. Satoh and Y. Ohta, "Occlusion detectable stereo --- Occlusion patterns in camera matrix,", IEEE Conf. on Computer Vision and Pattern Recognition, 1996.
11. D. Geiger, B. Ladendorf and A. Yuille, "Occlusions and Binocular Stereo," International Journal of Computer Vision (IJCV), Vol. 14, 1995, pp. 211-226.
12. S. Intille and A. Bobick, "Disparity-space images and large occlusion stereo," European Conference on Computer Vision, J-O. Eklundh (ed.), Stockholm, Sweden, Vol. 801, pp. 179-186, May 1994.
13. P. N. Belhumeur and D. A. Mumford, "A bayesian treatment of the stereo correspondence problem using half-occluded regions," In Proc. IEEE Conf. On Computer Vision and Pattern Recognition, 1992.
14. O. Faugeras, "Three-Dimensional Computer Vision: A Geometric viewpoint." The MIT press, Cambridge, 1993.
15. Forsyth and Ponce, " Computer Vision -A Modern Approach", Pearson Education. Chapter 1.
16. D. Marr, T. Poggio, "Cooperative Computation of Stereo Disparity". Science, New Series, Volume 194, Issue 4262 (oct. 15, 1976), pp 283-287
17. http://www.visioneng.com/
18. http://axon.physik.uni-bremen.de/research/stereo/principle.html
19. http://www.vision3d.com/stereo.html