INTERNATIONAL JOURNAL OF RESEARCH – GRANTHAALAYAH
A knowledge Repository

RACSIT - 17

LITERATURE REVIEW ON VARIOUS DEPTH ESTIMATION METHODS FOR AN IMAGE

Saloni Bahadur[1], Rashmee Shrestha[2], Yalapi Sumaharshini[3], Gnv Ravi Teja[4], Kalpitha.N [5]
[*1, 2, 3, 4] UG Scholar, Department of Computer Science, School of Engineering and Technology, Jain University, India
[5]Assistant Professor, School of Engineering and Technology, Jain University, India

DOI: https://doi.org/10.29121/granthaalayah.v5.i4RACSIT.2017.3342

Abstract

In this survey paper, different depth estimation techniques using cues from two images are observed. In this paper, different methods for depth estimation like Vergence, Stereo Disparity, Stereo Matching, Familiar Size, Defocus Cue, Convex Optimization, and Sum of Absolute Differences Algorithm are reviewed. Depth is determined by using maximum and minimum disparity.

Keywords: Stereovision; Disparity; Depth Estimation; Camera Calibration.

Cite This Article: Saloni Bahadur, Rashmee Shrestha, Yalapi Sumaharshini, Gnv Ravi Teja, and Kalpitha.N. (2017). "LITERATURE REVIEW ON VARIOUS DEPTH ESTIMATION METHODS FOR AN IMAGE." *International Journal of Research - Granthaalayah*, 5(4) RACSIT, 8-13. https://doi.org/10.29121/granthaalayah.v5.i4RACSIT.2017.3342.

1. Introduction

3D reconstruction is the method through which shape and appearances of real objects are captured from a set of 2D images. It is widely used in fields such as computer vision, computer graphics, 3D reconstruction and robotics.

Recovering 3d depth from images has always been a problem in computer vision. Most of the work based on visual 3d reconstruction has focused on binocular vision (stereopsis) and on other algorithms which require multiple images, such as structure from motion and depth from defocus. Geometric differences are mainly considered when these algorithms are used. Various monocular cues such as texture variations and gradients, defocus, color/haze, etc also contain some important depth information. Most of the work on depth estimation is based on stereovision.

Estimating depth from a single image is considered a challenging task. This is mainly because depth generally remains uncertain given only local image features. The main objective of our work is to build a model which generates second image from first view. It is done by using the distance by which camera has been moved. The camera has to be moved at the same height which in turn may involve errors and the second image that is generated by manual movement of camera may contain some errors.

In our view depth estimation is very small but crucial step towards image understanding with large goals. This helps in some tasks such as understanding the spatial layout of a specific scene, finding usable areas in a scene, detecting different objects, etc. Major work on depth estimation is focused on different methods that require multiple images, such as stereovision. In these algorithms they consider only the stereo (triangulation) cues and do not apply when a single image is available. Beyond stereo/triangulation cues, there are also numerous monocular cues, such as texture, variations and gradients, defocus, color/haze, etc. This can be beneficial to obtain rich 3D information. This paper thus, summarizes some concluding remarks on various depth estimation methods.

2. Literature Survey

Ashutosh Saxena et al. proposed a paper regarding 3D depth estimation from a single still image. Here, a training set of monocular images is collected. Unlike triangulation-based algorithms which are widely used, such as stereopsis and structure from motion, an algorithm has been developed that makes use of a large orthogonal set of monocular cues. The proposed model uses a hierarchical, multiscale Markov Random Field (MRF) that uses such cues to estimate depth from a single still image. It also estimates the depth and the relation between depths at different points in the image. It has been shown that even on unstructured scenes, this algorithm is repeatedly used to recover reasonably accurate depth maps.[1]

Osamu Ikeda considered the problem of estimating the depth of each pixel in a scene from a single monocular image. First, a semantic segmentation of the scene is performed and the semantic labels are used to show the 3D reconstruction. The photometric stereo with color segmentation and the binocular stereopsis have been combined to reconstruct accurate shapes from two color images and also to reconstruct the scene. Finally, the existing approach depends on accurate ground truth data for learning the parameters of the linear regression model. It aims at extending the model in order to learn from richer data sources together with synthetic data. [2]

Min Sun and Andrew Y. Ng et al. said that the problem of depth perception is essential to computer vision. For any image, depth and shape perception are mapped as they are important elements for many applications like object recognition, grasping, navigation, image composing and video retrieval. Monocular depth perception has an ability to improve all of these applications, provided only a single image scene is available. In a still image we have many monocular cues like texture variations, gradients, defocus, color which are used for depth perception.[3]

Hansung Kim and Kwanghoon Sohn noted that 3D construction algorithm is a two stage algorithm which contains disparity estimation and regularization. It is mainly used for process of estimation and also to increase efficiency. This method is proposed to solve occlusion and interaction problems that exist between real and virtual objects in MR system. The proposed algorithm has not yet been included in tracking system because it doesn't work in real time. Its goal is to improve computational efficiency and the possible solutions are implementing algorithm to FPGA (Field Programmable Gate Array) or DSP (digital signal processor). [4]

Beyang Liu et al. proposed an approach where, semantic segmentation is performed and used to show 3d reconstruction. Several advantages can be achieved using this approach. Depth and geometry constraints can easily imposed by knowing semantic class of a pixel or region. In this approach firstly, the foreground class is divided into subdivisions which allows addition of some more modeling constraints. Second, geometry currently makes strong assumptions about position of supporting pixels. Finally, it depends on accurate ground data for learning parameters of our linear regression model. [5]

N.Otsu presented a non-parametric method of automatic threshold selection for picture segmentation. A discriminant criterion is used to select an optimal threshold, in order to make best use of the separability of the resultant classes in gray level. In order to support the validity of the method, several experimental results are also presented. The method is rather common as it covers a wide scope of unsupervised decision procedure. [6]

Zhengyou Zhang introduced a new technique to easily calibrate a camera. It only requires the camera to examine a planar pattern that is shown at a few different orientations. If pixels are square, the minimum number of orientations is two. However, it is suggested that four or five different orientations should be used to improve quality. Both computer simulation and real data have been used to test the proposed technique and the results obtained have been quite accurate. In comparison to the traditional techniques which use expensive equipment, the proposed technique is easy to use and flexible. [7]

3. Methods of Depth Estimation

Various methods can be used for depth estimation as found in the literature, out of which a few are discussed in this section.

1) Vergence

When we position both eyes in such a way that the optical axes overlap on the surface of an object, it allows the projection of the object to fall on the foveae of both retinae. This allows us to obtain stereo fixation on the object. This movement of the eye is called edvergence. It gives a significant source of information about depth in the human visual system. Here, we make use of the vergence control algorithm to achieve fixation on the object surface.

2) Stereo Disparity (SD)

The difference between the locations at which an object is not at the stereo fixation and it projects to different locations on the left and right retina according to the horizontal baseline and scene depth that separates the eyes is called the stereo disparity (SD). It is a commonly used depth cue in artificial vision systems. Here stereo disparity is used to compute depth. Absolute depth information is obtained by a rectification process. The block matching algorithm is used to compute the disparity maps. This algorithm is used to refine disparities by means of post. The disparities of the object in the disparity maps are obtained using color based segmentation methods and the average of these disparities was used for depth estimation. The depth z is computed as:

$$z = (bf/d) + r + f$$

Where, d is the disparity, d=xVL−xVR(where xVL and xV Rare the projections of the object on the virtual left and right image planes).
f is the focal length of the camera, and
r is the distance from the center of rotation of the cameras to the image planes.

3) Stereo matching

The stereo matching problem takes two images that are captured by close by cameras and tries to map every pixel in one image to a position in the other image. The objective is to find the mapping which matches the corresponding pixels. Various constraints that are normally satisfied by true matches helps simplify the depth estimation algorithm, such as similarity, smoothness, ordering and uniqueness. The matching process is a conceptual approach in order to discover similar characteristics in different images. It can also be subjected to errors. Therefore, comparators are used to implement matching which in turn allows different identification strategies such as minimum square errors (MSE), sum of absolute differences (SAD) or sum of squared differences (SSD). The attribute that is compared through the matching process can be anything quantifiable. Thus, algorithms matching points, edges, regions or other image cues are used.

4) Familiar Size (FS)

If the object's real size is known, we can estimate the depth of an object from the size of its projection on the images that can be captured by the camera. This method consists of various approaches. The depth z can be obtained as:

$$Z= ((fW/w)+r+f)\cos\theta$$

Where, θ is the camera angle and $\cos\theta \approx 1$.
W is the size of the object.
w is the retinal size.

5) Using defocus cue

This method is used for the simultaneous depth estimation and image restoration from defocused observations. It makes use of two defocused observations of a scene that are captured using different camera parameters for the depth estimation. This method involves two steps. In the first step, the focused image is used to obtain the depth estimate. In the second step, the solution is refined by using fast optimization.

6) Convex Optimization Approach

This approach is used for robust depth estimation from stereo pair under different illumination conditions. A spatially varying multiplicative model is developed to show the change in brightness between the left and right views. Depth estimation is obtained as a constraint optimization problem in which convex object function is decreased.

7) Using object placement relation

One of the vision cues normally used to identify 3D positions efficiently is known as object placement. Extraction of such information is not so important. This method presents an algorithm in which placement information is defined as a constraint and it is used to estimate depth from a single scene image which has many random objects.

4. Comparative Study

METHOD	TYPE OF IMAGES USED	USE OF THE METHOD
Vergence	Optical axes overlap through eyes	Achieve fixation on object surface
Stereo Disparity (SD)	Image from different locations	refine disparities by means of post
Stereo matching	Dual images	discover similar characteristics in different images
Familiar Size (FS)	Camera captured image	estimate the depth of an object from the size of its projection
Using defocus cue	Different camera parameters	simultaneous depth estimation and image restoration
Convex Optimization	Different views	robust depth estimation from stereo pair
Object placement relation	Single scene image	estimate depth from a single scene image

5. Conclusion

The depth is an important aspect of a scene. However, it is usually lost in standard image acquisition systems. Due to this reason, various strategies have been developed to extract the depth. We have gone through the active methods, which can project some energy onto the scene to process the reflections, along with the passive methods that deal only with the natural received energy from the scene. Several strategies have also been discussed to solve the matching problem, or to identify the same physical points in two or more images. Through this analysis, we have come to know about the advantages and disadvantages in every system, in terms of energy needs, computational load and, therefore, speed, complexity, accuracy, range, hardware implementation or price, among others. Thus, there is no method that can be considered a winner among all the analyzed solutions. We have presented an efficient and robust algorithm in depth estimation which is appropriate for mobile robot navigation and obstacle detection in real time applications.

References

[1] Ashutosh Saxena, Min Sun and Andrew Y. Ng, "Learning 3-D Scene Structure from a Single Still Image", IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), pp.1-8, 2008.

[2] Osamu Ikeda, "Shape Reconstruction from Two Color Images using Photometric Stereo combined with Segmentation and Stereopsis", IEEE - AV SS, pp. 434-438, 2005.

[3] Make3D: Depth Perception from a Single Still Image Ashutosh Saxena, Min Sun and Andrew Y. Ng.

[4] Hansung Kim, Kwanghoon Sohn, "3D reconstruction from stereo Images for interactions between real and virtual objects", Elsevier journal on Signal Processing: Image Communication, vo1.20,pp.61-75, 2005.

[5] Beyang Liu, Stephen Gould, Daphne Koller, "Single Image Depth Estimation From Predicted Semantic Labels", IEEE conference on Computer Vision and Pattern Recognition: pp 1253-1260,2010.

[6] N.Otsu, "A threshold selection method from gray level histograms", IEEE Trans. Syst. Man Cybern. SMC-9, 62-66 (1979).

[7] Zhengyou Zhang, "A flexible new technique for camera calibration", IEEE Transactions on Pattern Analysis and Machine Intelligence 22(11) : 1330-1334 (2000).

[8] Dineesh Mohan, Dr. A. Ranjith Ram “A Review on Depth Estimation for Computer Vision Applications” International Journal of Engineering and Innovative Technology (IJEIT) Volume 4, Issue 11, May 2015.