A Review Paper on Stereo Vision Based Depth Estimation

Radhika Raval[1] Mahasweta Joshi[2] Bhavesh Tanawala[3]
[1]M.E. Student [2,3]Assistant Professor
[1,2,3]Department of Computer Engineering
[1,2,3]B.V.M Engineering College, Vallabh Vidhyanagar, Anand (Gujarat), India

Abstract— Stereo vision is a challenging problem and it is a wide research topic in computer vision. It has got a lot of attraction because it is a cost efficient way in place of using costly sensors. Stereo vision has found a great importance in many fields and applications in today's world. Some of the applications include robotics, 3-D scanning, 3-D reconstruction, driver assistance systems, forensics, 3-D tracking etc. The main challenge of stereo vision is to generate accurate disparity map. Stereo vision algorithms usually perform four steps: first, matching cost computation; second, cost aggregation; third, disparity computation or optimization; and fourth, disparity refinement. Stereo matching problems are also discussed. A large number of algorithms have been developed for stereo vision. But characterization of their performance has achieved less attraction. This paper gives a brief overview of the existing stereo vision algorithms. After evaluating the papers we can say that focus has been on cost aggregation and multi-step refinement process. Segment-based methods have also attracted attention due to their good performance. Also, using improved filter for cost aggregation in stereo matching achieves better results.

Key words: Stereo Vision, Disparity Map, Matching Cost Computation, Cost Aggregation, Disparity Computation, Disparity Optimization, Disparity Refinement, Segment-based Method, Stereo Matching

I. Introduction

Stereo vision is one of the most researched topics in computer vision. It has found a great importance in many fields and applications in today's world. Some of the applications include robotics, 3-D scanning, 3-D reconstruction, driver assistance systems, forensics, 3-D tracking etc. Stereo vision is used to infer depth from two images acquired from different viewpoint. Stereo vision determines the position of a point in space by finding the intersection of two lines passing through the center of projection and the projection of the point in each image [14].

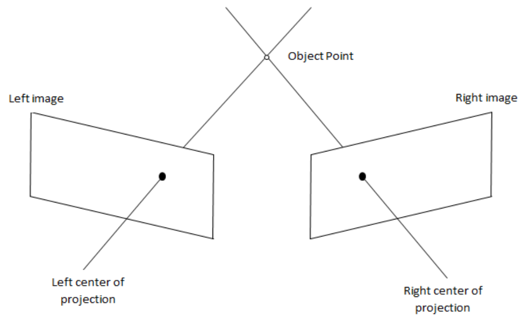

Fig. 1: The principle underlying stereo vision[14]

The depth is an important cue of a scene, which is lost in standard image acquisition systems. Several strategies need to be proposed to extract the depth. Computing the distance of an object from the camera system is called depth estimation [12]. And this is calculated from the pixel difference of its left and right image on the left and right camera respectively [12]. Disparity calculation based on stereo vision is an important problem in computer vision research [4].

A. Stereo Matching Problems

The problem of matching the same points or areas in stereo image pairs is called correspondence problem. If the camera geometry, the two dimensional search for corresponding points could be simplified to one dimensional search based on epipolar rectification [2].

There is also a problem of uniquely matching two points due to the fact that large regions with constant luminance exist, and in such regions more than one corresponding point could be identified [2].

Sometimes it happens that for some pixels in left image the corresponding pixel in right image does not even exist.

The remaining of the paper is structured as follows: Methodology of stereo vision algorithms in section 2. Section 3 gives a survey on some of the stereo vision algorithms with emphasis on their characteristics. And section 4 gives our conclusion.

II. Methodology

Stereo algorithms generally perform the following four steps [1]:

1) Matching Cost Computation
2) Cost Aggregation
3) Disparity Computation/Optimization
4) Disparity Refinement

The actual sequence of steps taken depends on the specific algorithm [1].

For example, local methods utilize colour or intensity values within a finite window to determine the disparity for each pixel [1]. On the other hand, global algorithms make explicit smoothness assumptions and then solve an optimization problem [1].

Each of the four steps is explained below briefly and different methods used by several authors have been reviewed with each step.

A. Matching Cost Computation

In this step, each pixel is initialized with matching costs at all disparity levels. The most common pixel-based matching costs include squared intensity differences (SD) and absolute intensity differences (AD) [1]. Klaus, Sormann and Kamer (2006) combines sum of absolute intensity differences (SAD) and a gradient based measure. Mukherjee and Reddy (2014) uses K-Means clustering and identifies cluster to which the pixel has been assigned. Mei et al. (2011) proposes AD-Census cost measure which effectively combines the absolute differences (AD) and the census transform. This measure gives more accurate matching results than common individual measures [10]. The work in [5] proposes a novel combined matching cost measurement that consists of the image colour absolute difference, the census transform and the new double-RGB gradient absolute difference [5]. Jiao et al. (2014) proposes a new cost measure by merging truncated absolute difference of colour, gradients and a modified colour census transform which improves the initial matching performance [6].

B. Cost Aggregation

Cost aggregation is an important step in stereo matching [1]. Most aggregation functions can be roughly classified into windows-based method [5, 6], filter-based method [9] and segment-tree-based method [7, 8, 11] Recently, full image support based cost aggregation methods [7] have gained considerable attention due to their high accuracy and low computational complexity. Algorithm in [11] use colour segmentation combined with an initial disparity estimate to compute an initial set of plane equations, and then refine these equations. Muninder, Soumik and Krishna (2014) proposed a novel method to compute sub-pixel precision disparity maps using minimum spanning tree (MST) based cost aggregation framework. Although based on segments, this method is highly robust to segmentation errors and parameter variations [8]. The work in [5] proposed improved exponential step aggregation function. This function also works as a novel postprocessing in the refinement step [5]. Jiao et al. (2014) uses a symmetric guided filter for cost aggregation. Huang et al. (2014) proposed a full-image guided filtering based on eight-connected weight propagation. This method outperforms local cost aggregation methods for low textured regions [9]. Mei et al. (2011) proposes enhanced cross-based aggregation method.

C. Disparity Computation/Optimization

Papers [5, 6, 7, 8, 10] take the 'Winner-Takes-All' strategy [1] to compute the raw disparity map. In [4] the disparities of only the boundary pixels are determined using sum of absolute differences (SAD) and then the disparity map is reconstructed from boundaries. Mei et al. (2011) uses WTA strategy to compute the raw disparity map and further scanline optimization is used to compute the intermediate disparity results. The authors in [11] determine disparities based on belief propagation.

D. Disparity Refinement

Raw disparity maps computed by correspondence algorithms contain outliers that must be identified and corrected. Several approaches aimed at improving the raw disparity maps have been proposed. Mukherjee and Reddy (2014) reconstructs the entire disparity map of the scene from the boundaries' disparity through disparity propagation along scan lines and disparity prediction of regions of uncertainty by considering disparities of the neighbouring pixels [4]. The authors in [7] propose to enhance the tree structure with a second segmentation process, which employs both colour and the estimated depth information [7]. In [8], the authors repeat the iterative framework of plane estimation and assignment to enhance the results. In [11] the optimized solution is found using belief propagation.

Sometimes it is difficult to remove all of the errors with only one method. The algorithms in [5, 6, 10] propose multi-step disparity refinement process. In [5] the authors performed outlier classification, simplified cross-skeleton support region, four-direction propagation, leftmost propagation and exponential step filter. Jiao et al. (2014) proposed Remaining Artifacts Detection and Refinement (RADAR) which included small hole filling, inconsistent region detection and modified occweight. Mei et al. (2011) performed outlier detection, iterative region voting, proper interpolation, depth discontinuity adjustment and sub-pixel enhancement. The multi-step refinement process gives better results than single one.

III. RELATED WORK

In [4], a dense disparity map is generated by using only 18% pixels of either left or right image of a stereo image pair. First, it segments the lightness values of left image pixels using K-means clustering. Then, a boundary map is generated which contains false identification of pixels at the segmented boundaries. So, it refines these segmented boundaries using morphological filtering and connected components analysis. SAD (Sum of Absolute Differences) cost function is used to determine the disparities of boundary pixels. The disparity map is reconstructed using disparity propagation and then pixels whose disparity has not yet been determined the values of the neighbouring pixels are used to estimate the disparity. The main advantage of this algorithm is that it estimates the disparity of refined boundary pixels only, thus reducing the number of computations needed.

The work in [5], improves the accuracy of local stereo matching method using combined matching cost and multi-step disparity refinement. This algorithm uses a guidance image for the whole system instead of directly using the raw image. The combined matching cost here achieves better performance than single. It consists of novel double RGB gradient, image colour and improved light weight census transform. Next, the cost aggregation step reduces noise but consumes most of the time. It uses improved exponential step aggregation function. Disparity is computed using the Winner-Takes-All (WTA) strategy. A multi-step refinement process is used which consists of outliers detection, four-direction propagation, leftmost propagation and exponential step filter. This step removes outliers from raw disparity map. This algorithm gives high accuracy performance for both indoor and outdoor images.

The work in [6], presents two strategies to improve the performance of local stereo matching method. First, combined matching cost is performed by modified colour census transform (MCCT), truncated absolute difference of colour and gradient and symmetric guided filter aggregation. Second, a secondary refinement approach called Remaining Artifacts Detection and Refinement (RADAR) further refines the result. The algorithm also works on real-world sequences and depth-based applications. The disadvantage of this algorithm is that the aggregation step and disparity refinement steps are time-consuming.

In [7], a tree based cost aggregation for stereo matching has been presented. For this, a tree structure, Segment-Tree (ST) is proposed. The ST is constructed in three steps. First, the pixels of the image are grouped in a set of segments. Second, a subtree for each segment is built. Lastly, all the subtrees are connected to build final ST. Two-pass cost aggregation on the tree structure is performed.

Further, the ST is enhanced using a second segmentation process. This algorithm performs better than the traditional MST method. But the performance of the algorithm is based on segmentation is correct or not.

In [8], the disparity estimation is done on per-pixel basis instead of general methods that estimate disparities on per-segment basis. First of all the initial disparity map is generated using any local or global algorithm. Then colour segments are generated using mean-shift segmentation. And then initial set of planes is determined. Pixel-wise cost volume is computed and minimum spanning tree (MST) is used to compute aggregated cost. To generate more accurate disparity map plane filtering is done followed by re-labeling. The result can be enhanced by repeating the plane estimation and assignment steps. The accuracy improves with increase in iterations. The main advantage of this algorithm is that even if the initial disparity map is poor, good results can be acquired.

The work in [9] presents an improved filter for stereo matching. This improved filter is based on eight-connected weight propagation. The filter is applied to cost aggregation in stereo matching methods. Comparison with earlier four-connected weight propagation suggests that the proposed eight-connected weight propagation is more approximate. The main advantage of this algorithm is that we can parallelize the filter on hardware platform.

The work in [10] presents a GPU friendly stereo matching system. The aim of the paper is to provide an accurate system with near real-time performance. Here, initial matching cost-volume is computed using AD-Census. Cross-based cost aggregation on each pixel's matching cost is computed and using scanline optimization the intermediate disparity results are produced. Disparity refinement step handles disparity errors in multi-step. But the multi-step mechanism brings a large set of parameters.

The work in [11] presents a segment-based stereo matching system. First, the reference image is decomposed into homogenous regions using mean-shift colour segmentation. The previous step generates a large number of disparity planes so a requirement arises to extract the planes that are sufficient enough to represent the scene. So for that local matching in pixel domain followed by disparity plane estimation step is performed. Lastly, approximate optimal disparity plane assignment is carried out. Here, the mean-shift segmentation consumes more time.

Advantages and limitations of the above analyzed papers are summarized in Table 1.

Work	Methods	Advantages	Limitations
Mukherjee and Reddy [4]	CIE Lab colour space, K-Means clustering, SAD, FILL, PEEK	Less number of computations is needed, scalable to high resolution stereo image.	Reduces performance in presence of shadows, may end up altering the shape of the objects, no occlusion handling.
Zhan, Gu, Huang, Zhang, and Hu. [5]	Mask filter, AD, double RGB gradient filter, census transform, exponential step filter, WTA, LRC, cross skeleton, leftmost propagation	Multi-step refinement reduces the error percentage, lowest average error, high accuracy performance in both indoor and outdoor environments.	Mask filter fails to work with a relatively small kernel size, improves average error except for nonocc and disc regions.
Work	**Methods**	**Advantages**	**Limitations**
Jiao et al. [6]	MCCT, AD, symmetric guided filter, WTA, LRC, RADAR, MOW	Applicable for high-resolution images as well as video sequences.	Aggregation and multi-step refinement process is time consuming.
Mei, Sun, Dong, Wang, and Zhang [7]	ST, MST, WTA	Performs better with more accurate depth borders and less noise, leading disparity accuracy and processing speed.	Performance is based on segmentation is correct or not, works for indoor dataset only.
Muninder, Soumik and Krishna [8]	Mean-shift segmentation, MST, WTA, occlusion filling	Computes cost volume at pixel level, gives good results even with poor initial disparity map.	The selection of window size is not trivial in MST.
Huang, Cui, Zhang [9]	Eight-connected weight propagation	Outperforms local cost aggregation methods for low textured regions, achieves competitive results with low complexity.	Works for limited amount of dataset.
Mei et al. [10]	AD census, cross-based support regions, WTA, scanline optimization, interpolation, sub-pixel enhancement	GPU friendly system design, works on images as well as video sequences.	Multi-step mechanism brings large set of parameters, artifacts are visible around depth borders and occlusion regions in video.
Klaus, Sormann and Kamer [11]	Over segmentation, mean-shift colour segmentation, SAD, gradients, belief propagation	SAD is one of the simplest similarity measure, mean-shift approach incorporates edge information.	Mean-shift segmentation is time consuming.

Table 1: Comparison of analyzed papers

IV. Conclusion

Estimation of disparity maps from stereo image pairs is a challenging task. Stereo vision is one of the most extensively researched topic in computer vision. We have seen the general methodology for stereo vision algorithms and important features of some of the existing stereo vision algorithms. The analysis has revealed advantages and disadvantages of every system. We cannot conclude to a single winner from the analyzed algorithms. We can say that there is still space for improving the accuracy of disparity maps by providing new ways of generating more accurate disparity maps.

REFERENCES

[1] Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondance algorithms" IJCV, vol. 47, no.1-3, pp.7–42, 2002.

[2] M. Gosta and M. Grgic, "Accomplishments and challenges of computer stereo vision" ELMAR, 2010 Proceedings, pp. 57-64, Sept. 2010.

[3] Pablo Revuelta Sanz, Belen Ruiz Mezcua and José M. Sánchez Pena (2012). Depth Estimation - An Introduction, Current Advancements in Stereo Vision, Dr. Asim Bhatti (Ed.), ISBN: 978-953-51-0660-9, InTech, DOI: 10.5772/45904. Available from: http://www.intechopen.com/books/current-advancements-in-stereo-vision/depth-estimation-an-introduction.

[4] S. Mukherjee, R. M. G. Reddy., "A Hybrid Algorithm for Disparity Calculation from Sparse Disparity Estimates based on Stereo Vision" IEEE Trans. SPCOM, pp. 1-6, July -2014.

[5] Y. Zhan, Y. Gu, K. Huang, C. Zhang, and K. Hu., "Accurate image-guided stereo matching with efficient matching cost and disparity refinement" IEEE Trans. Circuits Syst. Video Technol., vol. PP, no. 99, pp. 1., 2015.

[6] J. Jiao, R. Wang, W. Wang, S. Dong, Z. Wang, and W. Gao., "Local stereo matching with improved matching cost and disparity refinement" IEEE Trans. Multimedia, vol. 21, no. 4, pp. 16-27, 2014.

[7] X. Mei, X. Sun, W. Dong, H. Wang, and X. Zhang., "Segment-tree based cost aggregation for stereo matching" IEEE Trans. CVPR, pp. 313-320, 2013.

[8] V. Muninder, U. Soumik, and A. Krishna., "Robust segment-based Stereo using Cost Aggregation" Proceedings of the British Machine Vision Conference. BMVA Press, September 2014.

[9] X. Huang, G. Cui, Y. Zhang., "An Improved Filtering for Fast Stereo Matching" IEEE Trans. ICPR., pp. 2448-2452, Aug 2014.

[10] X. Mei, X. Sun, M. Zhou, S. Jiao, H. Wang, and X. Zhang., "On building an accurate stereo matching system on graphics hardware" IEEE Trans. ICCV Workshops, pp. 467-474, Nov 2011.

[11] A. Klaus, M. Sormann and K. Karner., "Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure" IEEE Trans. ICPR, vol. 3, pp. 15-18, 2006.

[12] Depth Estimation from Disparity and Defocusing. (2015, October 1). Retrieved from https://sites.google.com/a/eng.ucsd.edu/ece191-engineering-group-design-project-fall-2015/Projects/f

[13] Stefano Mattoccia. (2014, May). Stereo Vision: Algorithms and Applications [PowerPoint slides]. Retrievedfromvision.deis.unibo.it/~smatt/Seminars/StereoVision.pdf

[14] Stereo Vision. Retrieved from www.cse.unr.edu/~bebis/CS791E/Notes/StereoCamera.pdf

