Object Segmentation Based Disparity Map Determination for Stereo Images-A Review

Komal D. Bhavsar[1] Prof. Virendrasingh[2]
[1,2]Dept. of Electronics & Communication Engineering
[1,2]Sagar Institute of Research & Technology, Bhopal, Madhya Pradesh, India

Abstract--- Aiming at establishing stereo correspondence for the extraction of the 3-D structure of a scene, major recent developments are reviewed. Wide categories of stereo algorithms are identified based upon differences in image geometry, matching primitives, and the computational structure used. The matching approach to solve the supposed correspondence problem in static, binocular stereo vision has its limitations. Specifically, matching is of no use in occluded areas because there is nothing to match in those regions. Other kinds of problems, like large regions of the image with a very uniform surface result in erroneous matching in almost every case. Disparity in such regions can be determined with a different approach, based on well-known details and principles of stereo vision. Here Performance of these Disparity map on various classes of test images is reviewed and the possible direction of future research is indicated.

I. INTRODUCTION:

Stereo matching is one of the most active areas in computer vision because it is the basis for the accurate acquiring of image depth, which is important to applications in vision systems, such as object tracking, recognition, and path planning, etc. Although lots of algorithms have been proposed for stereo matching, especially for disparity map calculation, there are still many challenging works needed to be done caused by texture less, occlusion, etc. Bela Julesz was using computer synthesized stereoscopic pairs to explain binocular depth perception in the 1960's [11]. Some of the first computer algorithms to find depth from an arbitrary stereoscopic pair were devised in the 1970's [12], [13], [14], when researchers developed cooperative algorithms to investigate stereopsis.

The correspondence problem (stereo matching), has had a more or less continuous evolution with its ups and downs. From the beginning, the difficulty of the matching problem was recognized and a set of constraints and rules were proposed to limit the number of possible matchings [14]. Since good quality matching's occur only sparsely along a stereo pair many algorithms have concentrated on producing a *sparse* disparity map, [9]. Also, many algorithms have been devised to produce a *dense* disparity map. A review of the vast literature that has been published on stereo matching will not be attempted in this document but readers may refer to [15], and [16]. Here we provide a brief review of the state of the art in stereo vision. An exhaustive survey of the literature is beyond the scope of this document. Some books cover the basics of the subject: [5], [6], [7], [8], [9], [10].

II. RECTIFIED IMAGES

Fig. 1 shows a flow chart of the disparity map. The input Contains two rectified images. A transformation which makes pairs of conjugate epipolar lines become collinear and parallel to the horizontal axis (i.e., baseline).Searching for corresponding points becomes much simpler for the case of rectified images.

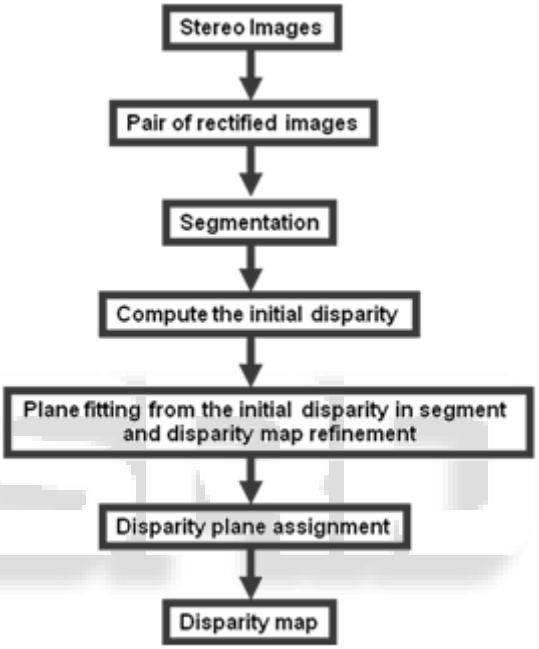

Fig. 1: The Flowchart of general Disparity Map

III. REFERENCE IMAGE SEGMENTATION

The first step in the workflow is to decompose the reference image into regions of homogeneous color or grayscale. The algorithm assumes that disparity values vary smoothly in those regions and that depth discontinuities only occur on region boundaries. Over-segmentation is preferred, since it helps to meet this assumption in practice. Therefore mean shift color segmentation recently successfully applied to image segmentation by Comaniciu and Meer [2] is used. The Mean-shift analysis approach is essentially defined as a gradient ascent search for maxima in a density function defined over a high dimensional feature space. The feature space includes a combination of the spatial coordinates and all its associated attributes that are considered during the analysis. The main advantage of the mean-shift approach is based on the fact that edge information is incorporated as well.

IV. STEREO MATCHING FOR INITIAL DISPARITY

The many different approaches that have been developed differ in the kind of features used for matching, or in their conception of matching space, or in the nature of matching

algorithms, or in the metrics used to judge similarity, Salvador et al.[17] gives a detailed descriptions. A variety of features have been used for matching by a number of authors including:

- Pixel to pixel stereo: Regardless of any interpolation procedure or any mode of interaction with neighboring pixels, or any support aggregation scheme, algorithms use intensity values of individual pixels to estimate disparity [18].
- Window based (fixed 2D window): The basis for comparison of positions on different images is the result of a computation on the elements of a neighborhood of fixed size. Windows have been very popular and are traditional within the correlation approaches [9]. This approach has been made more robust by methods that work on a ranking of intensities of the window elements and use special metrics to compare candidate matching. An approach using more than one fixed window for each position is described by [19]. Other window-based features can involve the output of filters or edge detectors [20]. • Variable 2D window: Some approaches adaptively increase the size of an initial window, depending on a threshold on a variance measure [21], being more robust in large homogeneous areas of stereoscopic pairs. An advanced variable window method was proposed by [22] that find the affine transformation that deforms the window in one of the images in such a way that a correlation measure is optimized.
- Adaptive-window Selecting Using HOG: In this method, we extract three channels of reference image and mark the three images as IR, IG, and IB, then divide IR, IG, and IB into an image window of fixed size, say, 32x32 pixels. Next, ori (h, w) is calculated in each window [1]. Computing the variance in each image window of IR, IG, and IB and comparing the variances in corresponding window, if the largest value is greater than a given threshold, then we Consider the orientation in this window is obvious in one channel and therefore we divide the window into four equally sized sub-windows, then calculate HOG within the sub window in the same way. The procedure will be stopped when the window cannot be divided further. In this way we can obtain the optional window for each pixel in the reference image.
- Arbitrary feature vector: A feature vector for each position is constructed with results of computations such as the output (magnitude and phase) of a bank of Gabor filters that sample all possible orientations, frequencies and scales, (as reported in [23]. Another example is a feature vector with three components: grey-level intensity in the first component, and derivatives along the *x* and *y* directions in the second and third components.

All of the above choices may use color information for matching purposes increasing reliability significantly. The matching space is the geometrical disposition of information useful for matching. It may be imagined as a continuous space, but algorithms work with sampled, discrete versions of it. For most matching procedures working with epipolar lines it is a 2D space with its axes corresponding to epipolar lines from the left and right images. For 3D approaches it usually is a 3D space where two of its coordinate axes are just the horizontal and vertical axes of one of the images, the third axis representing disparity or depth. Some approaches define and use a more sophisticated matching space where it is a projective 3D space, allowing conversion between disparity and depth, and transformation of information between different points of view. Sometimes adding extra dimensions on top of this, which aids in decision making for disparity, color and transparency retrieval. The nature of the matching algorithm can be very different from one approach to another:

- Dynamic programming which minimizes some sort of cost function is a popular choice because it allows natural statements of some constraints such as occlusions, continuity and monotonicity [19], ordering, exclusion of double occlusions etc. See also [18].
- Graph theoretical algorithms play a role in approaches that state the matching problem as a problem in a graph. When stating it particularly as a maximum flow problem there is no need for explicit use of epipolar geometry, allowing use of multiple cameras with arbitrary geometries. The solution gives a minimum-cut that corresponds to disparity. •The Bayesian approach allows a probabilistic statement of the matching problem, involving an imaging model that takes into account a priori information necessary to add constraints to possible solutions, and a prior model that reflects statistical properties of scenes where the theory is supposed to work. Bayes' Theorem combines these models giving a posterior distribution. Minimization of the expected value of a cost function computed with respect to the posterior distribution gives the MAP (Maximum a Posteriori) or the MPM (Maximizer of Posterior Marginals) estimator, depending on the cost function definition. The optimization problem may be solved using dynamic programming [19] or when working with paradigms like Gauss-Markov-Measure-Fields the problem will be solvable using some other standard optimization techniques.
- Phase-Based Methods: Images are convolved with quadrature filters (v.g.; Gabor filters) and disparity is computed from the measured phase difference. The simplicity of these approaches is appealing and they automatically provide sub pixel precision, however, the disparity range in which these methods are reliable is usually small (about one half the filter's wavelength) and it is difficult to obtain precise disparity edges. For these reasons they were not implemented for comparison and discussion in this document. This approach can be combined with motion cues to improve performance.
- A geometric approach using a partial differential equations (PDE). It defines a variational principle that must be satisfied by the surfaces of the objects in the scene and their images (more than two). The derived Euler-Lagrange equations provide a set of PDE's which govern evolution of an initial surface towards the observed scene objects. When implemented with level sets surface evolution it can manage multiple objects. It assumes that scene objects are graphs of smooth

functions and that they are perfectly lambertian. It can handle multiple views. It has been applied to simple synthetically objects.

- Cooperative algorithms were developed which operate on many "input" elements and reach global organization through local interaction constraints [14]. Two constraints were identified: CI, where ach point has a unique position in space at any time; and C2, where matter is cohesive. These constraints lead to two rules: R1. On uniqueness (each point from each image can be assigned at most one disparity value); R2, on continuity (disparity varies smoothly almost everywhere). These constraints and rules have been applied to random dot stereograms. Recently, Zitnick and Kanade have proposed a cooperative algorithm that works with 3D support to enforce or inhibit match values in a 3D disparity space.
- Multi-frame procedures use more than two images to strengthen the certainty of matches or simply provide a natural way to include information from more than two images. • Multi-resolution approaches estimate disparity on a hierarchy of scales, processing large scales first and using these estimates to initialize matching procedures on smaller scales.
- The stereo matching problem has been stated as a nearest-neighbor problem through the use of intrinsic curves, which are paths that a set of image descriptors trace as an image scan line is traversed from left to right (reminding space phase trajectories in dynamical systems). Metrics are those procedures used by matching algorithms to judge similarity between features. If features are point like, such as single pixel grey-scale values they may be compared using the absolute value of the difference of candidate points. Alternatively, squared differences can also be used. If ID, 2D or 3D features are used, the L\, L_2 or L_x norms may be used, or alternatively a correlation measure. Some probabilistic approaches using Bayesian estimation employ likelihoods as metrics, where greater likelihood values correspond to greater similarity.

The scope of most matching algorithms extends to a disparity map, though there has been some recent interest in reconstructing realistic 3D scenes mapping textures on a depth map (with applications to virtual reality in mind). These algorithms require interaction with graphics which poses new problems, since the quality of the output of most matching algorithms is not enough to meet the demands of these new applications. Some matching algorithms are now designed to retrieve disparity, color and transparency simultaneously [24] Matching algorithms can be found in software or hardware implementations, sequential or parallel. Some general purpose stereo systems include two, three or more cameras, and allow video rate computation of disparities.

V. PLANE FITTING AND DISPARITY MAP REFINEMENT

The reliable correspondences are used to derive a set of disparity planes that are adequate to represent the scene structure. This is achieved by applying plane fitting method and a successive refinement step.

A. Segmentation-based plane fitting:

Depth in each segment is then represented as a 3D planar surface model according to Tao et al. [39]: 1/Z = Ax +By+ C Since depth of a pixel is inversely proportional to its disparity: Z= λ Bf/d, we can model disparities in each segment as a 3D planar surface: d=ax+by+c.----(1) Our goal is to find out the least square solution (a,b,c) of this linear system. Considering that the least square method is very sensitive to the effects of outliers, only those points with reliable disparities are adopted to fit the plane.

A robust fitting process proposed is used to obtain plane parameters. In the first iteration, reliable points in a segment are selected to fit the plane. Then, if the disparity of a point is not within a given range α of the fitted plane, we assign it with a new disparity which can achieve Lowest matching cost in the given range. The renewed disparity map should be used to fit a new plane. The process iterates until the change of plane parameters is below a threshold β. If some segments, which are too small or lying on occlusion areas, don't have enough reliable points for solving a linear system, they should not be fitted to a plane model.

B. Disparity map refinement:

The purpose of this step is to increase the accuracy of the disparity plane set by repeating the plane fitting for grouped regions that are dedicated to the same disparity plane. With fitted-plane parameters of each segment, we can fill the occlusion regions in disparity map. Let m(Xm,Ym) denote an occluded pixel, we compute its disparity according to(1). dm =aXm +bYm +c. (2)

Similar as in (1), to refine the disparity map for the segment that hasn't been fitted, all the plane models computed in the previous step are used to compute a matching cost. Then the model which gives the minimal matching cost is selected. The matching cost of a segment is defined as the sum of pixel-to-pixel matching costs inside it:

Cseg (A,P)= Σ C(t,d) , t ϵ A

Where A is the segment, P is the plane model; t d denotes the computed disparity for pixel t under model P.

VI. DISPARITY PLANE ASSIGNMENT

In the final step an optimal solution for the segment-to disparity plane assignment is searched. Therefore the stereo matching is formulated as an energy minimization problem for the labeling f that assigns each segment s ϵ R a corresponding plane f(s) ϵ D. The energy for a labeling f is given by:

$$E(f) = E_{data}(f) + E_{smooth}(f),$$

where

$$E_{data}(f) = \sum_{s \in R} C_{SEG}(s, f(s))$$

and

$$E_{smooth}(f) = \sum_{(\forall (s_i, s_j) \in S_N \mid f(s_i) \neq f(s_j))} \lambda_{disc}(s_i, s_j).$$

SN represents a set of all adjacent segments and λdisc(si, sj) is a discontinuity penalty that incorporates the common border lengths and the mean color similarity as proposed in [29]. An optimal labeling with minimum energy is approximated using Loopy Belief Propagation [30] where the message passing takes place between adjacent segments.

VII. RESULTS

The stereo image pairs downloaded from Middlebury website http://cat.middlebury. edu/stereo/. Table 1 compares the stereo matching performance quantitatively with data from Middlebury evaluation engine. The number in the table is the error rate of each method. ADSW, BP, CDS, OR, GC, Double BP, Adaptive BP. The error rates of the methods without from Middlebury's evaluation website and [26]. The program of ADSW was written by Yoon [25], but the performance is different from what he claimed due to the lack of restoration phase. The error rate of the proposed CDS is better than BP, ADSW, GC. Note that the performance of these methods may also be enhanced by combining with segment constraint or occlusion constraint. OR's stereo matching performance is very similar to CDS's performance. In tsukuba and venus, OR achieves error rates lower than the error rates achieved by CDS. However, in more complex stereo pairs such as teddy and cones, OR's error rates are higher than CDS's. Among the compared methods, adapting BP [27] can achieve a remarkable low error rate because it combines segment constraint, ADSW, plane-fitting, and belief propagation all together. It is by far the method with the best performance, but may also be the slowest method of all.

Method	Tsukuba	Venus	Teddy	Cones
Adapting BP	1.37	0.21	7.06	7.92
Double BP	1.29	0.60	8.71	9.24
Cds	3.81	1.83	14.8	13.2
OR	2.27	1.22	19.4	17.4
ADSW	4.18	3.41	20.6	16
GC	4.12	3.44	25.0	18.2
BP	5.14	5.32	23.9	19.5

Table 1: Performance on Middlebury

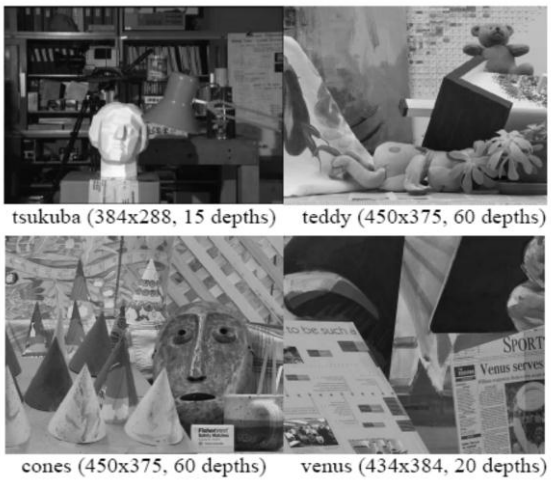

Fig. 2: The stereo image pair

Fig. 3 shows the qualitative result of CDS method. The disparity map is similar to the ground truth. The disparity boundaries of the teddy bear in teddy and the lamp in tsukuba are almost perfectly preserved. However, in tsukuba, the repeating pattern region on the left of the camera failed to correspond to correct disparities. The pixels in this repeating pattern region are isolated within the same segment. Therefore, their disparity could not be restored by exchanging the matching cost with the neighboring segments. The same problem also happened for the occluded area on the bottom-left of the house roof in teddy. In venus, incorrect disparities are shown near the boundary of object due to wrong segment constraint. This implies the CDS method is dependent on the quality of the segment constraint. In summary, the CDS is a simpler solution which achieves acceptable stereo matching performance slightly lower than the state-of-the-art method, but is much lower in complexity and requires much less runtime.

VIII. CONCLUSION

In this paper we have presented a review of the major techniques developed in the recent past for disparity map from the 3-D structure of a scene from stereo images. It has been shown that the matching approach to solve the so called correspondence problem in stereo vision has intrinsic limitations.Specifically, matching is of no use in occluded areas because there is nothing to match in those regions. Other kinds of problems, like large regions of the image with a very homogeneous texture will result in erroneous matching in almost every case. A method was proposed to compute disparity in such regions using a different approach, based on well known facts and principles of stereo vision, and its performance was compared to state of the art stereo algorithms. Algorithms need to be improved to give a lower percentage of false matches as well as better accuracy of depth estimates. Performance of algorithms needs to be evaluated over a broad range of image types in order to test their robustness. Most of the stereo work done so far has been limited to developing basic stereo matching capabilities for working with simplistic images. A great deal of research in stereo is needed in order to not only overcome the abovementioned difficulties but also to apply stereo techniques to solve more real-world problems.

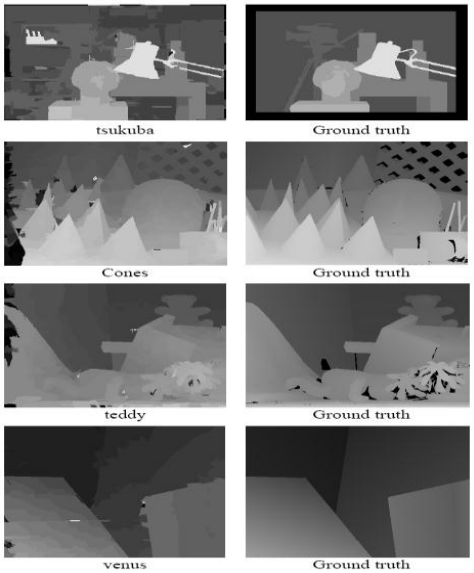

Fig. 3: Result of CDS method.

REFERENCES

[1] A New Segment-Based Algorithm for Stereo Matching,Zhihua Liu, Zhenjun Han, Qixiang Ye, Jianbin Jiao,Graduate University of Chinese Academy of Sciences, Beijing, China,{liuzhihua-b07 & hanzhenjun06} @mails.gucas.ac.cn
[2] D. Comanicu, P. Meer: "Mean shift: "A robust approach toward feature,space analysis", IEEE Trans. Pattern Anal. Machine Intell., May 2002.
[3] A. Klaus, M. Sormann, and K. Karner, "Segment-Based Stereo Matching Using Belief Propagation and a Self-Adapting Dissimilarity Measure", ICPR 2006, Vol. 3, pp.15 – 18, 2006.
[4] K.-J.Yoon, I.-S.Kweon, "Locally Adaptive Support-Weight Approach for Visual Correspondence Search," IEEE Conf. on Computer Vision and Pattern Recognition,924-931, 2005. [5]B. Julesz, Foundations of Cyclopean Perception. Chicago and London: The University of Chicago Press, 1971.
[6] W. E. L. Grimson, From Images to Surfaces. Cambridge, Massachusetts: MIT Press, 1981.
[7] H. L. F. V. Helmholtz, Treatise on Physiological Optics. New York: Dover, 1925.
[8] B. K. P. Horn, Robot Vision. Cambridge, Massachusetts: MIT Press, 1986.
[9] O. Faugeras, Three-Dimensional Computer Vision: A Geometric Viewpoint. MIT Press, 1993. [10]R. N. Klaus Voss and M. Schubert, Monokulare Rekonstruktion Fur Robotvision. Verlag Shaker, 1996. [11]B. Julesz, "Binocular depth perception of computer-generated patterns," Bell System Tech., vol. 39, pp. 1125-1161, September 1960. [12]J. I. Nelson Journal of Theoretical Biology, vol. 49, pp. 1-xx, 1975. [13]P. Dev International Journal of Man-Machine Studies, vol. 7, pp. 420-xxx, 1975.
[14] D. Marr and T. Poggio, "Cooperative computation of stereo disparity," SCIENCE, pp. 283-287, 1976.
[15] S. T. Barnard and M. A. Fischler, "Computational stereo," Computing Surveys, vol. 14, no. 4, pp. 553-572, 1982.
[16] U. R. Dhond and J. K. Aggarwal, "Structure from stereo - a review," IEEE Transactions on Systems, Man, and Cybernetics, vol. 19, no. 6, pp. 1489-1510, 1989.
[17] Disparity estimation and reconstruction in stereo vision,Salvador Gutiérrez and José Luis Marroquín,Comunicación Técnica No I-03-07/7-04-2003(CC/CIMAT). [18]I. J. Cox, S. L. Hingorani, and S. B. Rao, "A maximum likelihood stereo algorithm," Computer Vision and Image Understanding, vol. 63, pp. 542-567, May 1996.
[19] D. Geiger, B. Ladendorf, and A. Yuille, "Occlusions and binocular stereo," IJCV, vol. 14, pp. 211-226, April 1995.
[20] H. H. Baker, Edge Based Stereo Correlation, pp. 168-175. L.S. Baumann (Ed.), 1980.
[21] R. D. Arnold, "Automated stereo perception," Tech. Rep. AIM-351, Artificial Intelligence Laboratory, Stanford University, 1983.
[22] M. A. V. Robert Maas, Bart M. Ter Haar Romeny, "Area-based computation of stereo disparity with model-based window size selection," in Proceedings of the IEEE Computer Society Conferenceon Computer Vision and Pattern Recognition (CVPR'99), pp. 106-112, IEEE, 1999. [23]S. Gutierrez, Robust Methods for Disparity Estimation in Stereo Vision. Ph.D. thesis, Centro de Investigacion enMatematicas (CIMAT), Apdo. Postal 402, Guanajuato, Guanajuato, Mexico, CP. 36000, Mar 2001.
[24] R. Szeliski and G. Hinton, "Solving random-dot stereograms using the heat equation," (San Francisco, California), pp. 284288, IEEE Computer Society Press, 1985.
[25] K.J. Yoon and I.S. Kweon, "Adaptive Support-weight Approach for Correspondence search," IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006.
[26] M. Gerrits and P. Bekaert, "Local stereo matching with segmentation-based outlier rejection," in Proc. 3rd Canadian Conf. on Computer and Robot Vision (CRV'06), pp. 66-66, 2006.
[27] A. Klaus, M. Sormann and K. Karner. "Segment-based Stereo Matching Using Belief Propagation and a Self-Adapting Dissimilarity Measure," ICPR, 2006.
[28] Census diffusion with segment constraint for disparity estimation in stereo vision,Tusng-Hsien Tsai,Yeng-Chung, Nelson, Chang,E-mail: {thtsai, ycchang, tyucheng and tschang} @ twins .ee. nctu.edu.tw
[29] M. Bleyer and M. Gelautz. Graph-based surface reconstruction from stereo pairs using image segmentation. In SPIE, pages vol. 5665: 288–299, January 2005. [30]P. F. Felzenszwalb and D. P. Huttenlocher. Efficient belief propagation for early vision. In CVPR, pages I: 261–268,2004.

