International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, Volume 2, Issue 1, January 2012)

Area Based Image Matching Methods – A Survey

Jyoti Joglekar[1], Shirish S. Gedam[2]

[1]Research Scholar, Centre of Studies in Resources Engineering
IIT Bombay Mumbai, India

[2]Associate Professor, Centre of Studies in Resources Engineering
IIT Bombay, Mumbai, India

jyotij@iitb.ac.in
shirish@iitb.ac.in

Abstract— **This paper aims to present a review of recent as well as classic area based image matching methods. Determination of three dimensional data from images is very important task in the machine vision field. Stereo vision has wide range of potential application areas including three dimensional reconstruction of a view. Matching algorithms play a key role in deciding correspondences between two image scenes. Mainly, the matching algorithms are distinguished as area based matching and feature based matching. Area based methods sometimes called correlation like methods or template matching, merge the feature detection step with the matching part. These methods deal with the images without attempting to detect salient objects. Windows of predefined size is used for the estimation of correspondence. The major goal of this paper is to provide comprehensive reference for the researchers involved in, area based image matching. Main contribution, advantages and drawbacks of the area based methods are mentioned in this paper.**

Keywords—**Image matching, Correlation, Disparity, Fourier, Local algorithms**

I. INTRODUCTION

In the field of machine vision, three dimensional data determination is of prime importance. Stereo vision is one of the direct ways of achieving three dimensional data. Stereo vision has a wide range of application areas including robot vision, three dimensional map building. Correspondence solution is achieved using variety of constraints depending upon properties of the data. This is reflected in the broad range of algorithms that have been developed. In case of availability of camera calibration, epi-polar constraints can be used.

The absence of transparent objects allows the use of disparity gradient limits.Direct pixel matching using photo-metric properties is possible if the images are generated under constrained lightning conditions. The constraint factors have a strong influence on the quality and reliability of the three dimensional data recovery by an algorithm.

Consequently one algorithm is not capable of retrieving the best of an arbitrary set of data. In 1988 a survey of 15 institutions [1] found that, many researchers were using combinations of more than one constraint for solving correspondence problem in stereo vision. The survey concludes that each of the different approaches had its relative merits and disadvantages dependant on nature of correspondence problem. Hannah [2] suggested that a more accurate stereo solution would need to combine more than one approaches in a cooperative fashion. A practical approach of the combined approach is demonstrated by Baker [3], who successfully combined edge and photometric based stereo techniques. Bearing this in mind the following discussion should only be seen as analysis of area of applicability of each approach and not an attempt to decide a particular method best or worst.

A basic knowledge of the common approaches used by stereo matching algorithm is assumed for the next sections. For basic knowledge the reader may refer to following work [4, 5, and 2]. The purpose of this paper is to analyse the work presented for area of stereo vision. There are two main categories, area based stereo and feature based stereo. Area based stereo is used to classify algorithms where image domain similarity metrics are used for dense point to point correspondence.

Area based algorithms can be further divided as following categories.

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, Volume 2, Issue 1, January 2012)

- Cross-correlation based
- Fourier based
- Mutual information based
- Optimization methods (simulated annealing)

In addition to the different categories listed above other approaches are namely

- Human interaction
- Hierarchical processing
- Interpolation of a partial solution

II. THE PRINCIPLES OF AREA BASED ALGORITHMS

A degree of similarity exists between two views of the same scene at some image scale. If the scale is coarser the views become more similar.

If a view is considered as smaller sub regions the number of features for a given sub region decreases and the given sub region will look more similar to its corresponding sub region in the other view. An area based similarity metrics can be applied to sub regions or blocks to define the most likely correspondence between the same sub regions from two different views. If similarity is defined by dot product score of two vectors a = (x_1,y_1) and B = (x_2,y_2) where all bins are non negative then their dot product score will always increase if the vectors are coarsened by integrating bin contents [6]. The dot product is given by

$$AB = x_1x_2 + y_1y_2 \qquad (1)$$

Coarsening A and B by bin integration now give two more terms

$$C = x_1 + y_1 \qquad D = x_2 + y_2$$

$$C.D = x_1y_1 + y_1y_2 + x_1y_2 + y_1x_2$$

$$C.D = AB + x_1y_2 + y_1x_2 \qquad (2)$$

Since the coarser dot product contains extra terms it will be always be larger than the original. Appropriate similarity metric is to be used for area based stereo methods for enforcing consistence constraints and constraints like surface smoothness. The concept of cross-correlation function and a search space is also introduced with the use of area based methods. Similarity metrics derived from probability density functions theoretically offer best solution although in practice approximations such as Euclidean distance and dot product metrics have been used.

III. Area Based Stereo Algorithms – A Review

Work by Levine m.d.et.al 1973 forms the basis for a robot control system capable of exploring its environment [4]. Correspondence search is restricted using epipolar constraint to a search along a single epipolar line. Classical correlation measure as shown in equation 3 is used by the algorithm which is applied within a variable window size.

$$C_{X,Y,D} = \frac{\sum_{y}^{N}\sum_{x}^{N}(L_{x,y}R_{x+D,y} - \mu_L\mu_R)}{\sqrt{\sum_{y}^{N}\sum_{x}^{N}(L_{x,y}^2 - \mu_L^2)}\sqrt{\sum_{y}^{N}\sum_{x}^{N}(R_{x+D,y}^2 - \mu_R^2)}} \qquad (3)$$

Where X and Y define a block location in the left image and D is disparity parameter. The model of the world assumes that the surface consists of approximately horizontal plane which extends to infinity. With this well constrained model the algorithm seeks to build up horizontal contours which describe the ground plane using points called "tie points"; A coarse search of the correlation space is performed to identify candidate areas for the solution to achieve computational efficiency. A fine search is then performed at the most likely candidate areas. Assuming that a data model of the correlation function surface is known a robust quantization step to be used for the coarse search.

Most of the earliest work in stereo vision is concerned with the processing of aerial photographs for estimation of terrain depth. The work of Mori k .et.al [5] involves the development for such problems that incorporate some basic principles. Registration of images is done by normalized dot product score with the addition of Gaussian weighing. In many applications the dot product cross correlation function is used as similarity metric where maximum score is taken to represent the best match.

A 2D Gaussian kernel used for defining the weighing term w(x, y) that gives priority to components in the centre of calculation. In his work, dot product score is normalized against the autocorrelation score for the respective sub region in the left and right images.

The Mori.et.al algorithm also demonstrates the use of high confidence estimates, variable window sizes and iterative refinement of an interpolated disparity map. The algorithm assumes that edges within the images represent surface detail and not depth discontinuities. Due to this assumption regions containing edges are to be treated as high confidence estimates and from these starting points a solution are propagated and estimation of depth is used to reconstruct right image. The new right image is then recorrelated with original right image and the process of iteration is continued to refine the solution.

Hannah's work [2, 7] involving SRI's stereo system suggests the application of a number of different algorithms and incorporates the ideas of interest operators, hierarchical processing, area correlation and left right consistency. The Morovec interest operator is proposed for locating points within the image for which a high level of confidence about their matchability is assumed. Stereo matching by cross correlation is applied to a subset of regions centered on peaks in interest function.

High confidence matches are evaluated and used to steer further matching. Hannah's interest operator calculated the product of local intensity variance and heuristic directional variance quantity. Hierarchical processing is used at coarser resolutions to "set the context" for matches at higher image resolution.

This work shows how different methods can be applied together in a co-operative luminance cross-correlation. The hierarchical processing constraint assumes that the surface model is smooth both globally and locally. This algorithm is more suitable for finding depth information from aerial photographs.

Based on the findings of Gulech[1], Inria's algorithm obtain dense depth information from region based similarity metrics and some other constraints [8]. Main aim is to obtain dense depth map which is possible because of correlation test. Afterwards interpolation is used to obtain full depth map which is dense. Parallel hardware is used by this algorithm to simplify control flow.

Instead of locating regions of the image which are considered worth for and matching an attempt is made to match whole image and invalid matches are rejected in the validation phase. This work uses two correlation functions; a Euclidean distance based metric and normalized Euclidean distance metric by mean value of the region. It is suggested that due to normalization, invariance is obtained for linear transformations of the image grey level. In this work different methods of validation for the correlation are used to reject wrong matches and not using threshold for normalized correlation score to examine the peak in the correlation surface. Even the left right consistency constraint is used to reinforce the hypotheses obtained by matching from each image independently.

Okutomi et.al.[9] work addresses the issues of window size selection and presents a statistically sound technique which minimizes uncertainty in the estimation of disparity at each pixel of the depth map. The author suggests to select the window size in a dynamic fashion as the matching process proceeds as longer matching windows provide better disambiguation ability and less accuracy.

IV. AREA BASED METHODS

These methods are sometimes called as correlation methods or template matching. These methods merge the feature matching step with the matching part. In this method matching is done without detecting salient object. Windows of predefined size or sometimes the entire images are used for estimation of correspondence [7].

There are certain limitations in area based methods. Firstly the rectangular window suits the registration of images which locally differ only by a translation. If images are deformed by complex transformations then rectangular window can not able to cover same part of the scene in the reference image and sensed images.

Some authors proposed circular shape on window for mutually rotated images. But when complex geometric transformations are present even such simple shape windows is violated. Another disadvantage of area based method is that windows containing a smooth area without prominent details will be matched incorrectly with other smooth areas in the reference image due to non-saliency. Classical area based methods like cross-correlation(CC), where matching is done using image intensities and structural analysis is not used.

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, Volume 2, Issue 1, January 2012)

Consequently they are sensitive to the intensity changes introduced by noise, different viewpoints or by different sensor types.

A. Correlation Methods:

Normalized Cross Correlation and its modifications are classical area based methods [10]. The similarity measure is computed for window pairs from sensed image and reference image and its maximum is searched. The window pair for which similarity measure is maximum are set as the corresponding one.

For sub-pixel accuracy interpolation of the CC measure values is used. Although mutually translated images can be aligned with CC based measures it can also be applied when slight rotation and scaling is present. Generalized versions of CC are there for geometrically more deformed image. In this method CC is computed for each assumed geometric transformation of the window of sensed image. Thus this method can handle more complex geometric transformations than only translation.

In the method by Berthilsson [11] even affinely deformed images are matched. In this paper a method is proposed for maximizing affine correlation between images. The method is based on coordinates change at certain positions in the images and use Fast Fourier Transformation (FFT). Simper [12] proposed a method where he used divide and conquer system and CC technique for registering images differing by perspective changes as well as changes due to lens imperfections.

However if the transformation complexity increases then computational complexity also increases. Extended CC methods based on increment sign correlation is used in case of images with partially occluded objects [13].

The Sequential similarity detection algorithm (SSDA)[14] uses sequential search and computationally simple distance measure than the CC. Sum of absolute difference of intensity values is used for candidate pair of windows from sensed and reference images and a threshold criteria is used for rejection of unmatched windows. This method is likely to be less accurate but faster than CC method. Sum of squared differences similarity measure was used in ref [15].

Recently correlation ratio based methods are used in the area of multimodal registration. This similarity measure can handle intensity differences between sensors comparison of this approach to other algorithms which are used for multimodal images is in ref [16].

A method is proposed by Huttenlocher et al [17] which used similarity measure as Hausdorff distance [HD]. The sensed and reference images are registered binary images obtained by edge detectors are used for transforming translation and rotation. They compared HD algorithm with CC algorithm.

There are drawbacks in correlation methods as flatness of similarity measure and high computational complexity. By using edge or vector correlation the maximum can be sharpened. In the method proposed by Pratt [18] to improve CC performance on noisy or highly correlated images, image filtering is applied prior to matching stage.

Van Wie [19] used the edge based correlation which is computed on the edges extracted from the images rather than the original images. This method is not sensitive to intensity difference between images because of different viewpoints. Correlation methods are used often because their hardware implementation is easy and is useful for real time applications.

B. Fourier Methods

To increase computational speed if the images are acquired under varying condition or they are corrupted by noise, which is frequency dependant the Fourier method are preferred over correlation methods. Frequency domain representation of images are exploited for this. The phase correlation method is based on Fourier shift theorem [20] and proposed for registering translated images. Cross-over spectrum of the image pair is computed and location of peak is searched in the inverse. This method is robust against highly correlated and frequency dependant noise disturbances. If the images which are to be registered are large this method provides computational time savings.

C. Local Methods

Local methods are usually fast and can produce better results. Several new methods have been presented. A method proposed by Munhlmann et al [21] uses SAD correlation measure for color images. It gives results with high speed and quality. Left right consistency and uniqueness constraints are used for validations in this method.

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, Volume 2, Issue 1, January 2012)

A speed of 20 fps is achieved for an image size of 160x120 pixels, making this method suitable for real time applications. A Fast area-based stereo matching algorithm is presented in [22] and referred as SMP (Single Matching Phase). In this method as you proceed in matching process previous matches are rejected if better matches are detected further based on uniqueness constraint.

These methods perform in one matching phase in contrast to bidirectional matching and have similar results. Similarity measure used in this method is Sum of Absolute Differences (SAD). Results are tested for reliability and sub-pixel refinement. A dense disparity map is produced by this method and it is useful for real time applications. It achieves 39.59 fps speed for an image with size 320x240 pixels and 16 disparity levels and the root mean square error for the standard Tsukuba pair is 5.77.

One more advanced method is found in [23]. This method uses zero mean normalized cross correlation (ZNCC) as similarity measure and this matching cost is integrated with neural network model which uses least mean square delta rule as a supervised learning and training rule. Proper shape and size for window is decided for each region using the neural network. Results obtained are satisfactory but the running time needed for some standard image sets show that it is not suitable for real time applications.

D. A Composite Area Based Image Matching Method

The method proposed in [24] is a composite technique where first the similarity measure between template window and search window is found by normalized cross correlation technique. Few best matches are selected for the template window from the search sub-windows, considering the largest normalized cross correlation coefficient. Further edge map is obtained for stereo image pair using canny edge detector. The matches for the template window are filtered using Hausdorff distance technique. Further texture analysis of the same template window and selected search windows is the third measure to decide the accurate match. Texture analysis is done with the co-occurrence matrices which is a two dimensional histogram of the occurrence of pair of intensity value in a given spatial relationship. With this composite method dense point to point correspondence can be achieved with greater accuracy.

This method is tolerant to radiometric distortions and parallel processing of the three techniques will improve the speed.

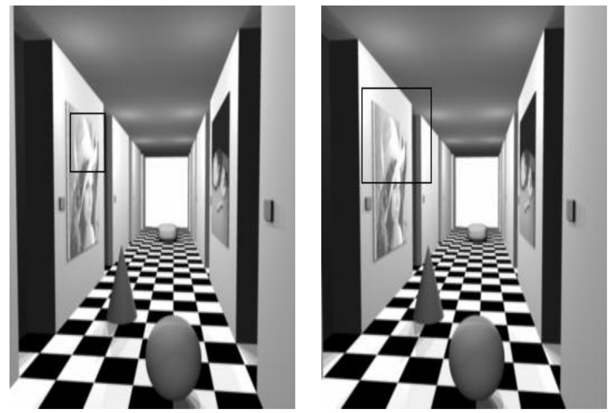

Fig. 1 : Stereo image pair with left image with template window and right image with search window

The proposed method in [24] is executed on the test stereo image pair shown in Fig. 1. As shown in the figure the left image shows the template window and the right image shows the larger search window. The size of the search window is decided by considering horizontal and vertical disparity measured by visual inspection of stereo image pair in Fig. 1.

For the proposed algorithm in [24], the three methodologies used are zero mean normal cross correlation, Hausdorff distance and texture analysis. Parallel processing of the three methodologies can be incorporated for improving speed. Using zero mean normal cross correlation and considering largest five coefficient for the best matches for the given template window as shown in Fig. 2a.

The matches found are shown in Fig. 2b. The minimum threshold of 0.8 is used for selecting the search sub-window as a match. If the normalized cross correlation coefficient is less than 0.8 for all search sub-window means that there is no match in the right image for the template window under consideration in the left image.

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, Volume 2, Issue 1, January 2012)

Fig. 2a: Template window from left image from stereo image pair

Fig. 2b: Selected sub widows as probable matches

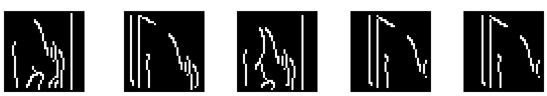

Fig. 2c: Binary edge images of sub windows of Fig 2b

Fig. 2d: Selected sub windows as matches for template window atfter applying Hausdorff distance criteria

Fig. 2e: Best match selected after texture analysis

Fig 2: Results of Composite Area based Image matching method [24]

Further on the selected best matches and template window canny edge detector is applied and a set of binary images are obtained as shown in Fig. 2c. Computation of Hausdorff distance and ranking the matches based on Hausdorff distance is done. After the third step of the algorithm the best matches are filtered to the output as shown in Fig. 2d. The third refinement is done with texture analysis. Texture analysis is done on template window and selected search sub-window of the Fig. 2d. The fourteen texture features are defined by (Haralick, R. M., Shanmugam, K. S., and Dinstein, I., 1973) in [25]. Five texture features are computed from the co-occurrence matrices of the template window and the search sub-window and the difference between the feature coefficients is compared with a threshold which is empirically designed out for each feature.

If the difference of the feature coefficients of three features are within the predefined range then the match is accepted. Sometimes cross-correlation technique fails to compare areas of smooth variance of intensities. Hence the accuracy of the match can be improved by texture analysis technique.

V. CONCLUSION

The stereo correspondence problem is an active area of research. Accuracy in real time operations is demanded by more and more applications. Both area based and feature based algorithms walk towards this objective. But when dense disparity map and simple algorithms for less computational complexity are needed area based algorithms outperform feature based algorithms. Area-based methods may be comparatively more accurate, because they take into account a whole neighborhood around the points being analyzed to establish correspondences. The survey concludes that each of the different approaches had its relative merits and disadvantages, dependent on nature of correspondence problem. A more accurate matching solution can be obtained by combining more than one approaches. Hence a composite ara based method would be the solution for more accurate results.

References:

[1] Guelch E., Results of Test on Image Matching. *ISPRS WG III/4, Institute of Photogrammetry*, University of Stuttgart, 1988.

[2] Hannah M.J., A System for Digital Stereo Image Matching. *Photogrammatric Engineering and Remote Sensing* 1765-1770, 1989.

[3] Baker H.H., Depth from Edge and Intensity Based Stereo, *Stanford AI Lab.*, Stanford University, 1982.

[4] Levine M.D., O'Handley D.A., Yagi G.M., Computer Determination of Depth Maps. *Computer Graphic and Images Processing* vol. 2, 131-150, 1973.

[5] Mori K., Kidode M. and Asada H., An Iterative Prediction and Correction Method for Automatic Stereocomparison. *Computer Graphics and Image Processing* 2, 393-401, 1973.

[6] Ashbrook A.P., Pairwise Geometric Histograms, *ESG MPhil. to PhD. Transfer Report*, Department of EEE, University of Sheffield, Sheffield, 1993.

[7] Hannah M.J., SRI's Baseline Stereo System, *Proc. of DARPA Image Understanding Workshop*, 149-155, 1985.

[8] Inria Research Labs. France., A Parallel Stereo Algorithm that Produces Dense Depth Maps and Preserves Image Features. *Research Report* No. 1369, 1991.

[9] Okutomi M. and Kanade T., A Locally Adaptive Window for Signal Matching. *Intl. Journal of Computer Vision* 7:2, 143-162, 1992.

[10] W.K. Pratt, Digital Image Processing, *2nd ed., Wiley, New York*, 1991.

[11] R. Berthilsson, Affine correlation. *Proceedings of the International Conference on Pattern Recognition ICPR'98*, Brisbane, Australia, 1998, p. 1458–1461.

[12] A. Simper, Correcting general band-to-band misregistrations, *Proceedings of the IEEE International Conference on Image Processing ICIP'96*, Lausanne, Switzerland, 1996, 2, pp. 597–600.

[13] S. Kaneko, I. Murase, S. Igarashi, Robust image registration by increment sign correlation, *Pattern Recognition* 35 (2002) 2223–2234.

[14] D.I. Barnea, H.F. Silverman, A class of algorithms for fast digital image registration, *IEEE Transactions on Computing* 21 (1972)179–186.

[15] G. Wolberg, S. Zokai, Image registration for perspective deformation recovery, *SPIE Conference on Automatic Target Recognition* X, Orlando, Florida, USA, April 2000, p. 12.

[16] A. Roche, G. Malandain, X. Pennec, N. Ayache, The correlation ratio as a new similarity measure for multimodal image registration, *Proceedings of the First International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI'98)*, Lecture Notes in Computer Science, Cambridge, USA, vol. 1496, 1998, pp. 1115–1124.

[17] D.P. Huttenlocher, G.A. Klanderman, W.J. Rucklidge, Comparing images using the Hausdorff distance, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 15 (1993) 850–863.

[18] W.K. Pratt, Correlation techniques of image registration, *IEEE Transactions on Aerospace and Electronic Systems* 10 (1974) 353–358.

[19] P. Van Wie, M. Stein, A landsat digital image rectification system, *IEEE Transactions on Geoscience Electronics* 15 (1977) 130–136.

[20] R.N. Bracewell, The Fourier Transform and Its Applications, *McGraw-Hill*, New York, 1965.

[21] K. Muhlmann, D. Maier, J. Hesser and R. Manner, Calculating Dense Disparity Maps from Color Stereo Images, an Efficient Implementation, *IJCV 47*, 1/2/3, pp. 79-88, 2002.

[22] Luigi Di Stefano, Massimiliano Marchionni, Stefano Mattoccia, A fast area-based stereo matching algorithm, *Image and Vision Computing* 22,pp. 983-1005, 2004.

[23] P.H.s. Torra, A. Criminisi, Dense stereo using pivoted dynamic programming, *Image and Vision Computing* 22, pp. 795-806, 2004.

[24] Joglekar J. V, Gedam S. S., Area Based Image Matching Technique using Haudorff distance and Texture analysis, *ISPRS WG III/4 International conference PIA11* , Technical University of Munchen, Germany, 2011.

[25] Haralick, R. M., Shanmugam, K. S., and Dinstein, I., Textural Features for Image Classification, *IEEE International Conference on Systems, Man, and Cybernetics*, Vol. SMC-3, No.6, pp 610-621, 1973