Segment Controlled Window Shape to Compute Disparity Map from Stereo Images

Rachna
Computer Science & Engineering Department

H.S. Singh
Physics Department

A. K. Verma
Production and Industrial Engineering Department

J.N.V. University, Jodhpur, Rajasthan, India

ABSTRACT
Stereo correspondence mapping is the fundamental problem to achieve human like vision capabilities to machines and robots. Many local and global algorithms have been reported in literature in the last decade. Window-based cost aggregation methods for solving the correspondence problem have attracted researches as it can be implemented in real time using parallel processors. In this paper a new window-based stereo matching algorithm with segment controlled window at each pixel to compute disparity map has been proposed. The proposed method uses sum of square difference correlation function on the window. In the proposed algorithm, pixels of square window which lie on the same segment to which the center pixel belongs are only considered while creating the window. Further, left-right consistency check is applied to generate disparity map taking full advantage of speed and simplicity of window based method.

Keywords-- stereo vision, correspondence, disparity, correlation, segmentation

1. INTRODUCTION
It has been established that the depth can be recovered from two (or more) images of the scene taken from slightly different known viewpoints using stereo vision based on principle of triangulation. However, in terms of speed and performance no conclusive solution has been developed yet. The fundamental basis for stereo vision is the fact that a single visible three dimensional physical location of a scene is projected to a unique pair of image locations in two observing cameras. As a result, given two images, if it is possible to locate the image locations that correspond to the same physical point in the scene, then it is possible to determine its three dimensional location [2]. The recovery of image depth permits the creation of a three dimensional model of the scene from 2-D images. This is called 3-D object reconstruction.

The depth of various scene points may be recovered if the disparity of the corresponding points can be calculated from rectified stereo images. Rectification means to transform and rotate the images so that the epipolar lines are aligned horizontally. This reduces the search of corresponding pixel in 2D image to 1D i.e. in rectified images the corresponding points will always lay on the same horizontal scan line. The disparities of all image points of stereo images is called disparity map which can be stored and displayed as an image. The depth *z* of a point in the scene is calculated using equation (1) that uses principal of triangulation.

$$z = \frac{bf}{(x'_l - x'_r)} \tag{1}$$

Where,

b = the baseline distance (difference between the position of two cameras),

f = focal length of the camera,

z = depth to be computed and

$d = x'_l - x'_r$ is the calculated disparity

One of the problems in the Stereo Vision to be solved is the Correspondence problem i.e. to locate the image locations in the stereo images that correspond to the same physical point in the scene. The complexity of the matching problem can be reduced if the images are rectified. In the rectified images the corresponding points will always lie on the same scan-line. Stereo correspondence algorithms developed during the last three decades can be broadly divided into local methods and global methods. In Local methods pixels in a small window surrounding pixel of interest (support window) are used whereas global methods use complete scan-lines or the entire image. Global methods offer high quality disparity maps but are computationally expensive and do not meet real time requirements. Most recent work focuses on global methods because of high performance but local methods attract researchers because they can be implemented in parallel and produces results in real time.

Local Stereo matching typically operates on a window that is shifted on the corresponding scanline in the right (target) view to find the point of maximum correspondence. The choice of an appropriate window size and correlation functions are important for disparity determination. Small window do not capture enough intensity variation to give correct results in less-textured regions. On the other hand, large window tend to blur the depth boundaries and do not capture well small details and thin objects.

In window-based local method the elements to match are image windows of fixed size, and the similarity criterion is a measure of the correlation between windows in the two images. The corresponding element is given by the window that maximizes the similarity criterion within a search region. The left image is taken as a reference image and the window is shifted over the disparity range in the right image as a target image. The aim is to find out the best match for the reference image coordinate from the target image over the disparity range.

One of the major problem of window based local method is that computed disparity map fattens at depth boundaries. This paper focuses on improving the disparity map at object boundaries and thus reduces the fattening effect typically generated by window based methods. Assuming that with image segmentation depth discontinuity coincides with colour boundaries, the proposed method uses image segmentation during disparity computation. Any pixel which lies outside the segmented region is considered as an outlier and is not used in cost aggregation step. Therefore the square window only contains image intensities of the pixels from the segment to which the centre pixel belongs. The experimental result shows that this approach produces good disparity map at depth discontinuity compared to the state of the art of local stereo algorithms.

2. RELATED WORKS

The last two decades have seen a lot of research in the field of stereo vision and many algorithms have been developed. Scharstein and Szeliski [1] and Brown et al. [2] are excellent review of stereo vision research. Recently window based correspondence matching has attracted researchers due to its real time capabilities. In this approach, a rectangular window is used in place of pixel of interest in an effort to reduce the effect of noise and other radiometry variations and researchers proposed various cost aggregation approaches to match corresponding pixels in left and right images. Comparative performance study of various cost aggregation approaches implemented on the GPU for real-time stereo matching is presented by [7]. It has been observed that at depth discontinuity boundaries, color discontinuity boundaries also occur. Based on this observation many methods employed segmentation in their approach and select the shape and size for support window. Geerits and Bekaert[4] proposed a support aggregation method based on segmentation of the reference image only (I_r). Any pixel that lied outside the image segment belonging to the central pixel is weighted by a small value. The major drawback of their method was that the support aggregation relies only on the segmentation of the reference image. To get accurate results at depth discontinuity as well as in homogeneous regions, Yoon and Kweon [5] proposed that the larger weights must be assigned to closer pixels and to pixels which are similar in color. They adjusted the support weight of the pixel in a given support window based on color similarity and proximity to the center pixel. These weights regulate a pixel’s influence in the matching process. A combination of [4] and [5] was proposed by [6] which applied segmentation on both reference and target images. The classification and comparison of stereo matching methods based on a variable support window has been reported by [3] in terms of accuracy and computational requirements.

3. PROPOSED METHOD

This paper proposes an efficient segmentation based local window based disparity map calculation for stereo vision. In local stereo matching the aim is to find out the best match for the left image coordinate from the right image over the disparity range. The proposed method is modified version of method presented by [4] and deviate from it in terms of (a) use of simple SSD function in place of computationally expensive Geman-McClure function and (b) assignment of zero weight to pixels not lying in the segment of the pixel of interest. The proposed method is composed of following three steps:

(1) Segment controlled window creation
(2) Dissimilarity computation based on sum of square difference correlation function
(3) Disparity selection.

The description of each step is given in the following subsection.

2.1 Segment Controlled Window Creation

The basic idea of the proposed method is to employ segmentation information in the reference image. Since it is observed that depth varies smoothly inside a segment, hence all pixels outside the segment are treated as outliers. Therefore all the pixels of the square window that lie outside the segment of the central pixel under consideration are given zero weight value for cost aggregation in the window. The algorithm for window creation is as follows

Input:
Image intensities I_l (left image), I_r (right image) and segmented image S
Algorithm

1. *For each pixel (x,y) in the reference image*
2. *Create a window w of size m x m centered at (x,y).*
3. *Get the segment value z of the central pixel.*
4. *Fill window values according to the following*

```
For i=1 to m.
    For j=1 to m
        If S(i,j) is equal to z
            w(i,j)=I_l(x,y)
        else
            w(i,j)=0
        end
    end
end
```

The shape of the window is not fixed but varies at each pixel because it contains either zero or image intensity values in a fixed size window. This window is shifted over the disparity range in the right image to find out the best match. For image segmentation, the software EDISION is used. This is available at http://coewww.rutgers.edu/riul/research/code/EDISON/ which uses mean shift segmentation [8] to get homogeneous regions in images.

2.2 Dissimilarity Computation

The computation of the disparity using segment controlled window depends only on image intensity values of corresponding segement only. For the matching procedure, the sum of squared differences (SSD) correlation function is used as a similarity measure. For each displacement d in the disparity search range the cost at each pixel of interest (x,y) is calculated using SSD is given in equation 1.

$$c(d) = \sum_{n}\sum_{m}(I_l(x+m, y+m) - I_r(x+m, y+m+d))^2$$

Where the size of the window is m x m and I_l and I_r are the image intensities which is set to zero for those pixels which do not lie on same segment of the center pixel.

2.3 Disparity Selection

Finally the disparity map is obtained for each pixel p of the reference image using the following equation

$$D_p = \min_{d \in S} c(d)$$

Where S represents the set of all possible disparities which varies from d_{min} to d_{max} and c(d) is the matching cost for disparity d. The values of d_{min} and d_{max} are known.

4. DISPARITY COMPUTATION ALGORITHM

The input to the algorithm is a rectified stereo pair image I_l and I_r, window size m x m and disparity range d_{min} to d_{max}. The algorithm performs the following steps similar to the window based approaches [1].

a. Matching cost computation-The matching cost is the squared difference of intensity values at a given disparity.
b. Cost aggregation-aggregation is done by summing matching cost over window with constant disparity.
c. Disparity computation-disparities are computed by selecting the minimal aggregated value at each pixel.

The disparity obtained for all points of the left image can be displayed as an image. All the pixels in the disparity image which appear brighter have higher disparity and are nearer to the camera, dark points have lower disparity and are farther from the camera.

To improve the accuracy of the proposed method, disparity map for both the reference image and the target image is obtained. Both disparity maps are combined to eliminate errors occurred at depth discontinuity and consistent disparity map is obtained. The left and right disparity map generated is used to filter out the mismatched disparities and consistent disparity map is obtained by applying left-right consistency check [2]. This is done by only retaining the disparity values that are same in both left and right disparity maps and for mismatched disparity the minimum of both the disparity is considered. For example, let d_l is the disparity at (x, y) in the left disparity map. Let d_r is the disparity at (x, y- d_l) in the right disparity map. If d_l is same as d_r, d_l is assigned to consistent disparity map else, the minimum of both values is assigned.

5. IMPLEMENTATION AND RESULTS

The proposed algorithm is implemented in Matlab software. The algorithm is tested for its performance on benchmark rectified stereo pair available at http://www.middlebury.edu/stereo. The result presented in the table (1) are for window size 21 x 21 (m=10). Segmentation parameters are spatial=7, color=6.5, minimum region =25. The evaluated results of the proposed method are shown in the figure 1. The table (1) shows the performance of the proposed algorithm with respect to other algorithms of various researchers. Its performance is comparable with the other algorithms.

6. CONCLUSIONS

In this paper an efficient local window-based method with segment controlled window is used to compute disparity map. Left and right consistency check is applied for computing final disparity map. The result of the proposed algorithm is comparable with the results of the other researches of local based methods. However, further research is required to reduce the disparity errors for the robust and quality real time implementation of stereo vision.

7. ACKNOWLEDGEMENT

The first author is thankful to the UGC, New Delhi, India for funding this research vide their San. No. F. 37-498/2009(SR) under major research project scheme.

8. REFERENCES

[1] Scharstein D. and Szeliski R., "A Taxonomy and Evaluation of Dense two Frame Stereo Correspondence Algorithms", Int. Journal Computer vision, vol.47,pp.7-42, 2002.

[2] Brown M. Z., Burschka D. and Gregory d. Hager, "Advances in Computational Stereo" ,IEEE Trans. Pattern Analysis and Machine Intelligence ,vol. 25(8), pp 993,1008, 2003.

[3] Tombari, F., Mattoccia, S., Stefano L D. and Addimanda E.:, 'Classification and Evaluation of Cost Aggregation Methods for Stereo Correspondence', Proc. Computer Vision and Pattern Recognition, pp. 1-8,2008.

[4] Gerrits M.and Bekaert P., "Local Stereo Matching withSegmentation-based Outlier Rejection", Proc. Conf. Computer and Robot Vision,pp 66-73,2006.

[5] Yoon K. and Kewon S. , "Adaptive Support-Weight Approach for Correspondence Search", IEEE Trans. Pattern Analysis and Machine Intelligence, vol.28 no.4 pp. 650-656, 2006.

[6] Tombari F., Mattoccia S. and Di Stefano L., "Segmentation-Based Adaptive Support for Accurate Stereo Correspondence ",IEEE Pacific-Rim symposium on Image and Video technology(PSIVT),LNCS , vol4872,pp-427-438 Springer, Heidelberg (2007).

[7] Gong, M., Yang, R. Wang, L. and Gong, M., 'A Performance Study on Different Cost Aggregation Approaches used in Real-Time Stereo Matching', Int. Journal Computer vision, 2007, 75,(2),pp. 283-296.

[8] Comaniciu D. and Meer P.,"Mean Shift: A Robust Approach Toward Feature Space Analysis", IEEE Trans. Pattern Analysis and Machine Intelligence,vol.24(5),pp603-619,2002.

Algorithm	Avg.	Tsukuba ground truth			Venus ground truth			Teddy ground truth			Cones ground truth			Average Percent Bad Pixels
	Rank	nonocc	all	disc	nonocc	all	disc	nonocc	all	disc	nonocc	all	disc	
FLTG-DDE [90]	84.0	3.03 71	5.28 80	15.0 81	3.39 91	5.02 92	25.0 96	11.0 83	19.5 88	26.3 91	5.78 74	16.0 87	14.2 74	12.5
DP [1b]	88.3	4.12 85	5.04 78	12.0 67	10.1 103	11.0 103	21.0 92	14.0 90	21.6 90	20.6 70	10.5 95	19.1 95	21.1 92	14.2
DPVI [67]	88.5	4.76 89	5.83 86	16.6 87	4.89 95	5.66 95	22.9 94	11.0 84	16.2 76	23.4 81	9.64 90	15.6 86	23.5 99	13.3
Bipartite [78]	89.9	2.54 61	4.41 71	13.6 73	6.62 96	7.46 96	18.6 88	16.9 97	24.1 97	30.2 96	15.1 104	21.8 102	23.0 98	15.4
YOUR METHOD	92.4	5.36 98	6.03 88	16.5 86	3.31 89	3.74 87	17.3 85	17.4 99	23.8 94	31.7 97	10.6 96	17.2 90	24.8 100	14.8
PhaseBased [31]	93.1	4.26 87	6.53 94	15.4 83	6.71 97	8.16 97	26.4 99	14.5 91	23.1 91	25.5 87	10.8 98	20.5 100	21.2 93	15.3
RegionalSup [38]	93.9	3.99 83	6.05 89	14.2 75	8.14 99	9.68 100	36.8 104	18.3 102	26.7 102	32.1 98	9.16 89	19.3 96	19.9 90	17.0
IMCT [62]	94.4	4.54 88	5.90 87	19.8 97	3.16 88	3.83 88	23.2 95	18.0 101	23.1 92	35.3 101	12.7 99	18.5 94	27.9 103	16.3
SSD+MF [1a]	94.9	5.23 97	7.07 95	24.1 103	3.74 92	5.16 93	11.9 70	16.5 96	24.8 98	32.9 100	10.6 97	19.8 97	26.3 101	15.7

TABLE 1: The results of the proposed method as reported by Middlebury Stereo Evaluation - Version 2. Values in the table represent error percentages measured in different image region(Partial table)

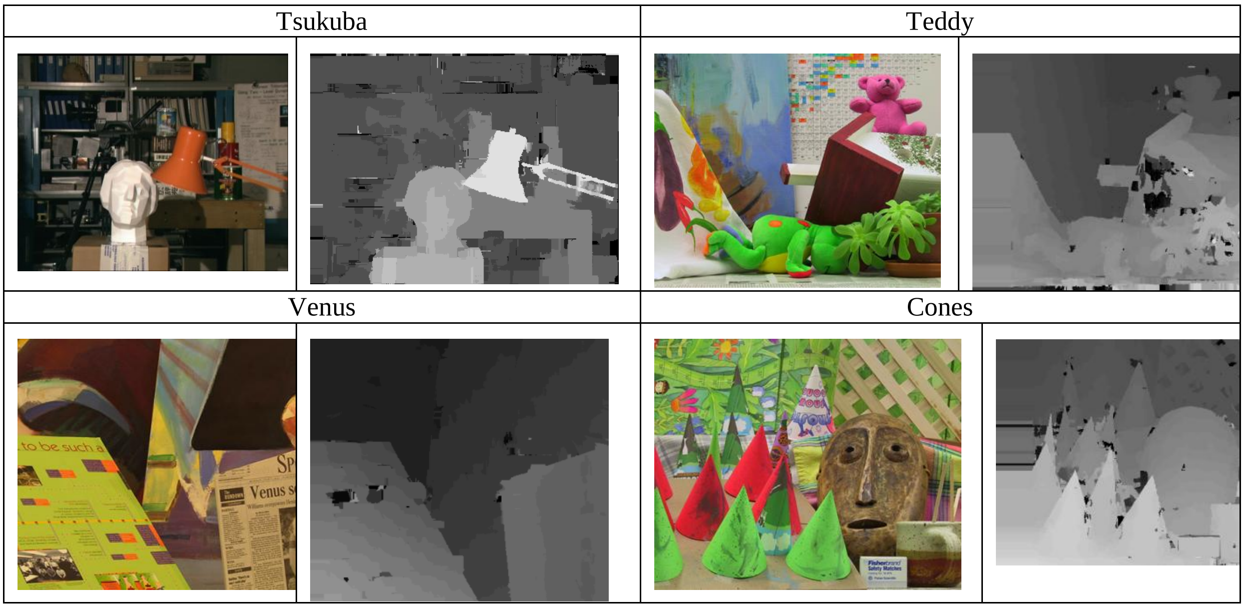

Fig. 1: Final disparity results computed by proposed method for testbed imagaes Tsukuba,Venus, Teddy and Cones.