J. Inst. Eng. India Ser. B (December 2012–February 2013) 93(4):223–229
DOI 10.1007/s40031-012-0030-z

ORIGINAL CONTRIBUTION

Depth Estimation from Stereo Images based on Adaptive Weight and Segmentation

R. Verma · H. S. Singh · A. K. Verma

Received: 12 March 2012 / Accepted: 19 October 2012 / Published online: 25 April 2013

Abstract Stereo vision is one of the most researched areas to develop human like vision capability into machines for the purpose of automatic navigation and reconstruction of the real world from images. Point correspondence matching for disparity map calculation is a vital research issue of stereo vision system. Window-based cost aggregation methods used for the correspondence problem have attracted researches as it can be implemented in real time. In this paper, a new hybrid cost aggregation strategy for similarity evaluation based on adaptive weight and color segmentation of stereo image is proposed. In this strategy, pixels which lie on same segment and are spatially closer are given higher weight. Experimental results show that the proposed method is effective in improving overall disparity map and at depth discontinuity.

Keywords Stereo vision · Correspondence · Disparity · Adaptive weight · Segmentation

Introduction

Stereo vision (SV) uses two images of a scene captured by a stereo camera consisting of two cameras placed side by side as shown in Fig. 1 for the purpose of determining depth of scene points from the cameras. The fundamental basis of stereo vision is that a visible scene point is projected at two different locations in the two images. The difference between the distances of these image points from the corresponding camera axis is referred as disparity. The disparity is inversely proportional to the depth of the scene point. As shown in Fig. 1 the left and right image planes are coplanar and their optical axes are parallel. A point P in the scene is projected by the two cameras at different positions P_l and P_r in the corresponding images. The depth of the scene point P from the camera origin, z, is inversely proportional to the difference between P_l and P_r as given by Eq. (1) [1]. The scene points depth are essential for various down line activities such as 3D reconstruction from scenes, automatic mobile robot navigation, terrain mapping and many other applications in engineering and medicine where human vision like capability is required.

$$z = \frac{bf}{(x_l - x_r)} \tag{1}$$

where b is the baseline distance between two cameras; f, the focal length of the cameras; z, the depth of the scene point from the camera origin; x_l, the distance of scene point projection from left camera axis in the left image; and x_r is the distance of scene point projection from right camera axis in the right image.

One of the vital research issues in SV is accurately and efficiently finding the projected locations of a scene point in both images. This is referred to as point correspondence matching. This seemingly simple issue is eluding the researchers for the last three decades due to noise and errors in the image and imaging devices. The problem is further compounded due to occlusion of some points. The occluded points are visible in only one image, hence no correspondence

R. Verma (✉)
Department of CSE, Faculty of Engineering, J.N.V. University, Jodhpur, Rajasthan, India
e-mail: rachna_mbm@yahoo.co.in

H. S. Singh
Department of Physics, Faculty of Science, J.N.V. University, Jodhpur, Rajasthan, India

A. K. Verma
Department of PI Engineering, Faculty of Engineering, J.N.V. University, Jodhpur, Rajasthan, India

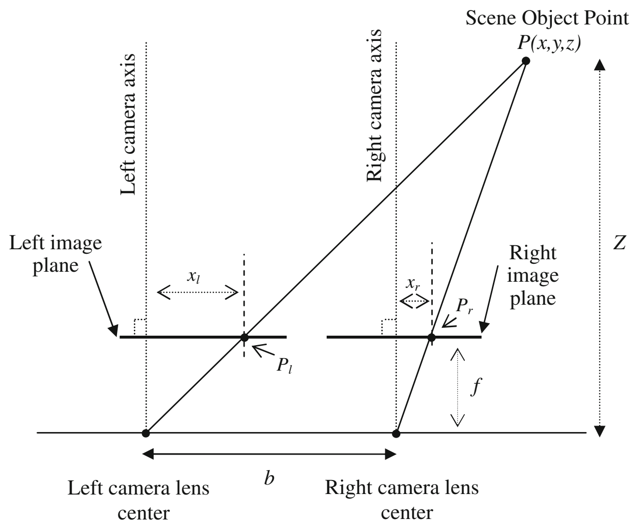


Fig. 1 Depth estimation from a stereo system

exist in other image. To reduce the complexity of the problem, the stereo images are rectified before finding the correspondence. In rectification images are translated and rotated 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 lie on the same horizontal scan-line in both images.

The point correspondence issue is an active research problem and many algorithms have been developed during last three decades. These algorithms can be broadly classified into local and global methods [2]. In local methods pixels in a small window, called support window, surrounding pixel of interest, called centre pixel, are used whereas in global methods complete scan-lines or the entire image is used for the purpose of correspondence matching. Accuracy of disparity maps generated by the global methods is superior as compared with the local methods. However, the global methods are computationally very expensive and do not lend themselves for real-time implementation whereas local method can be implemented in real-time [2]. Hence, current focus of researchers is on quality improvement of disparity map generated by local methods.

In local methods, a support window is selected in the reference image (normally left image) and a number of similar windows are selected in target image (normally right image) on the corresponding scanline. Based on some similarity criterion the best match of the reference window in the target windows is searched. The corresponding element is given by the window that maximizes the similarity criterion within the search region. Local methods generate poor quality disparity maps specifically in textureless areas and at object boundaries [2].

This paper aims to improve the disparity map particularly at object boundaries by utilizing segmentation information in similarity evaluation criterion. The proposed method is an enhancement to the method proposed by Tombari et al. [3] which used segmented image for disparity computation. The basic idea of [3] was that any pixel that is not in the segment of center pixel is considered as an outlier and is given a very small weight in similarity evaluation. In the proposed method idea of segmentation support [3] and adaptive weight based on spatial closeness and color similarity as proposed by Yoon and Kweon [4] are combined. The term adaptive weights indicate that the support weights of the pixels in a given window are adjusted according to their photometric and geometric relationship with the centre pixel [4]. Thus pixels which lie in the same segment and are spatially closer to the centre pixel are considered as inliers and have higher weights in similarity evaluation. On the other hand, the pixels which fall outside the segment and are spatially apart from the center pixel have lower weights in similarity evaluation. The proposed method is effective in improving the disparity map on all the test bed stereo images [5].

Related Works

The local stereo matching algorithms were criticized for inaccurate disparity map estimation than global stereo matching algorithms. This gap has been reduced by recent local stereo algorithms [6]. Yoon and Kweon proposed the adaptive weight aggregation method based on color similarity and spatial proximity which outperformed earlier local stereo algorithms. Gong et al. [7] and Richardt et al. [8] modified the original adaptive weight algorithms to make it suitable for real time applications. It has been observed that depth discontinuity at object boundaries is also accompanied by color discontinuity. Based on this observation many methods employed segmentation in their approaches [3, 9]. Gerrits and Bekaert [9] proposed a support aggregation method based on segmentation of the reference image only and assigned a very small weight to a pixel that lies outside the image segment of the center pixel. Effect of various cost aggregation approaches on quality of disparity has been reported by Tombari et al. [10]. Scharstein and Szeliski [5], Brown et al. [2] and De-Maeztu et al. [6] has presented extensive reviews of stereo vision research carried out during the last two decades.

Yoon and Kweon [4] reported a method that can generate quite accurate disparity map at depth discontinuity as well as in homogeneous textureless regions. They proposed to adjust the support weight of the pixel in a given support window based on color similarity and spatial proximity to the pixel of interest. The color similarity is based on the Euclidian distance between two colors and the spatial proximity is geometric distance in the image domain.

According to gestalt principle of proximity, the support weight of a pixel decreases as the spatial distance to the reference pixel increases [4]. The adaptive support weight of a pixel is given by Eq. (2)

$$wt_a(p,q) = \exp\left(-\left(\frac{\Delta c_{pq}}{\gamma_c} + \frac{\Delta g_{pq}}{\gamma_P}\right)\right) \tag{2}$$

where p is any pixel of the support window and q is the centre of the window. Δc_{pq} represents strength of grouping by similarity and is equal to the Euclidian distance between p and q in CIELab color space. Δg_{pq} represents grouping by proximity and is equal to the Euclidian distance between p and q in the image space. γ_c and γ_p are scaling constants. For detailed explanation of these parameters kindly refer Yoon and Kweon [4].

Using this concept, overall weight for similarity evaluation of reference window and target window is calculated by Eq. (3). Let subscript r refers to reference window and subscript t refers to target window. The matching cost between reference and target centre pixels as shown in Fig. 2 is computed by Eq. (3).

$$ct(q_r,q_t) = \frac{\sum\limits_{p_r \in W_r, p_t \in W_t} wt_r(p_r,q_r)wt_t(p_t,q_t)\mathrm{TAD}(p_r,p_t)}{\sum\limits_{p_r \in W_r, p_t \in W_t} wt_r(p_r,q_r)wt_t(p_t,q_t)} \tag{3}$$

where, TAD is Truncated Absolute Difference of color intensities of reference and target pixels in RGB color space with reference to a predefined threshold t as given in Eq. (4). The TAD operator increases the robustness of matching cost calculation against the outliers [3, 4].

$$\mathrm{TAD}(p_r,p_t) = \mathrm{Min}\left(\sum_{c \in \{r,g,b\}} |I_c(p_r) - I_c(p_t)|, t\right) \tag{4}$$

Tombari et al. [3] applied segmentation information within the weight cost function. The key idea of [3] was that each pixel lying on the same segment of the center pixel of the window must have a similar disparity value, and its weight has to be equal to the maximum value of the range. They applied segmentation on both reference and target images and proposed a modified weight function as given in Eq. (5).

$$wt(p,q) = \begin{cases} 1.0\ldots p_q \\ \exp\left(-\frac{\Delta c_{pq}}{\gamma_c}\right)\ldots p \notin S_q \end{cases} \tag{5}$$

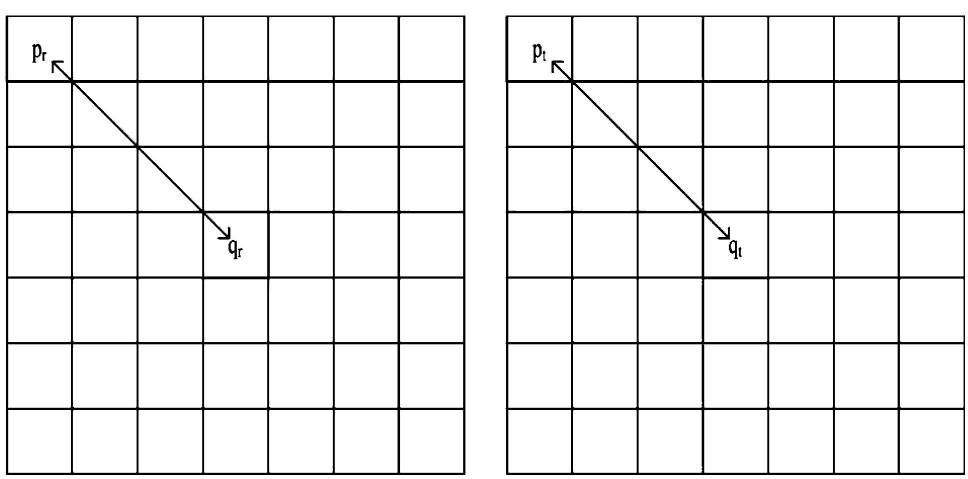
where S_q is the segment in which the centre pixel q lies. The proximity term was eliminated from the overall weight computation. Eq. (3) was used to calculate the overall weight of the window. Tombari et al. [3] method outperformed Yoon and Kweon [4] method.

Proposed Method

This paper brings together the strengths of the segmentation based [3] and the adaptive weight based [4] disparity map calculation algorithms. Tombari et al. [3] proposed that each pixel lying on the same segment of the central pixel of the correlation window have disparity value close to that of the central pixel. This is not true for segments corresponding to slanted objects where there can be large variation in disparities among segment pixels. However, change in disparity values in adjacent pixels of a segment will be gradual. Based on this observation, following two changes are proposed and incorporated into the method presented in Tombari et al. [3].

(i) The adaptive weight is combined with the segmentation weight. Thus all the pixels of a segment will not have same weight. Thus pixels of a segment which are

Fig. 2 Support weight windows



(a) Reference window(W_r) **(b) Target window(W_t)**

nearer to the center pixel will have more weight as compared to the pixels away from the center pixel.

(ii) The pixels of a correlation window that lie outside the segment of the central pixel will be assigned weight as given in Yoon and Kweon [4].

The proposed method has combined the strength of methods developed by Tombari et al. [3] and Yoon and Kweon [4] i.e. segmentation information and adaptive weight respectively. This helps in handling of the object boundaries and depth discontinuities in better way and experimentally it is found that proposed integration has outperformed both the method as given in Tombari et al. [3] and Yoon and Kweon [4] and produced better results at depth discontinuities as well. The method is composed of following three steps:

A. Support weight aggregation.
B. Dissimilarity computation.
C. Disparity selection.

Support Weight Aggregation

The pixels in the correlation windows are assigned adaptive weight as well as segmentation weight. The adaptive weight of a pixel is given by the Eq. (2). In contrast to [3] in which pixels lying outside the segment were given weight based on only color similarity whereas it is proposed to include color similarity as well as spatial proximity for the same. Experimentally, it is found to outperform [3] and [4] on object boundaries and depth discontinuities and is consistent with the gestalt principle of grouping. The segmentation weight is now given by modifying Eq. (5) to include spatial weight and is given by Eq. (6)

$$wt_s(p,q) = \begin{cases} 1.0\ldots p \in S_q \\ \exp\left(-\frac{\Delta c_{pq}}{\gamma_c} + \frac{\Delta g_{pq}}{\gamma_p}\right)\ldots p \notin S_q \end{cases} \tag{6}$$

The symbols used in Eq. (6) carry identical meaning as defined in Eqs. (2) and (5). The overall support weight of a pixel in a support window is sum of values calculated by Eqs. (2) and (6). The weight is calculated for reference and target windows separately. For image segmentation, the software EDISION is used, which is available at http://coewww.rutgers.edu/riul/research/code/EDISON/. This uses mean shift segmentation [11] to get homogeneous regions of reference and target image. The segmentation parameters such as spatial radius (h_s), range radius (h_r) and minimum region size (M) control image segmentation as well as number of segments generated.

Dissimilarity Computation

The dissimilarity between pixels is computed by aggregating the costs in the support windows. The support weights based on color similarity, spatial proximity and segmentation information of both the reference and target window are considered during computation. The dissimilarity $ct(q_r,q_t)$ between reference pixel q_r and target pixel q_t is given by Eq. (7) where subscript r is used for the reference window and subscript t is used for the target window. Equation (3) is modified to include the segmentation weight as well as the adaptive weight for matching cost calculation as given in Eq. (7). The segmentation weight and the adaptive weight are added together pixel-wise and aggregated over support window for their combined effect on the matching cost.

$$ct(q_r,q_t) = \frac{\sum\limits_{p_r \in W_r,\, p_t \in W_t} \{wt_{ar}(p_r,q_r) + wt_{sr}(p_r,q_r)\}\{wt_{at}(p_t,q_t) + wt_{st}(p_t,q_t)\}\,\mathrm{TAD}(p_r,\ p_t)}{\sum\limits_{p_r \in W_r,\, p_t \in W_t} \{wt_{ar}(p_r,q_r) + wt_{sr}(p_r,q_r)\}\{wt_{at}(p_t,q_t) + wt_{st}(p_t,q_t)\}} \tag{7}$$

where wt_{ar} and wt_{at} are the adaptive weights of reference window and target window respectively and are computed using Eq. (2). The wt_{sr} and wt_{st} are the segmentation weights for the reference window and target window respectively and are computed using Eq. (6). TAD $(p_r,\ p_t)$ is computed using Eq. (4).

Disparity Selection

Finally, the disparity is obtained for each pixel p of the reference image using Eq. (8).

$$D_p = \operatorname*{Min\ Index}_{d \in N} ct_d(q_r,q_t) \tag{8}$$

where, N represents the set of all possible disparities which varies from $d_{\min}$ to $d_{\max}$. The values of $d_{\min}$ and $d_{\max}$ are predefined to reduce the search space. $ct_d(q_r,\ q_t)$ is the matching cost of the reference window and the target window corresponding to the disparity d. The Min Index function returns the disparity value corresponding to the minimum matching cost in the disparity range. The disparity obtained for all points of the left image can be displayed as an image and is known as disparity map.

Disparity Refinement

To improve the accuracy of the proposed method, disparity maps for both the reference and the target images are calculated separately. Left–right disparity consistency check is used to segregate consistent and inconsistent disparities [2]. This is done by retaining the disparity values that are same in both the disparity maps and mismatched disparities are discarded and marked with a negative value. Figure 3a shows the images after applying left–right consistency for testbed images.

The inconsistent disparities are handled with following three steps.

1. The first step uses neighborhood disparities to estimate disparity for an inconsistent disparity. Calculate the disparity density of an initial 3×3 window centered at pixel of interest whose disparity is to be determined. The disparity density is defined as percentage of consistent disparity pixels with respect to total number of pixels of the center pixel segment inside the window. If disparity density is greater than 50 % then the mode of the consistent disparities is assigned to the center pixel. If the disparity density is less than 50 % then the window size is increased by one pixel on each side i.e. now the window size is 5×5 is taken and disparity density is calculated again. If disparity density is greater than 50 % then the mode of the consistent disparities is assigned to the center pixel. Otherwise the above process is repeated again by increasing the size of the window until the disparity density of the segment inside the window is greater than 50 % or selected window touches segment boundary in any direction.
2. The unassigned disparities left after the first step is filled by linear interpolation. Effort is made to locate pixels with consistent disparities in left and right of the inconsistent disparity pixels of a scanline of a segment. The left and right disparities are used to linearly fill the inconsistent disparities.
3. The unassigned disparities left after step 2 are assigned minimum of the left and the right disparity maps. Figure 3b shows the disparity map after disparity refinement.

Implementation and Results

The proposed algorithm is implemented in Matlab. The algorithm is tested for its performance on benchmark rectified stereo pairs available in the website [12]. The parameters are kept constant for all four image pairs. Segmentation parameters (h_s, h_r, M), γ_c and window size (w) are taken from Tombari et al. [3]. γ_p is determined empirically. The result presented in the Table 1 are for the window size $w = 51 \times 51$, $t = 35$, $h_s = 3$, $h_r = 3$, $M = 35$, the constants $\gamma_c = 22.0$ and $\gamma_p = 25.0$. Table 1 compares the disparity map of [3, 4] and the proposed method. Figure 4 shows that the proposed approach generates improved disparity map in general and particularly at

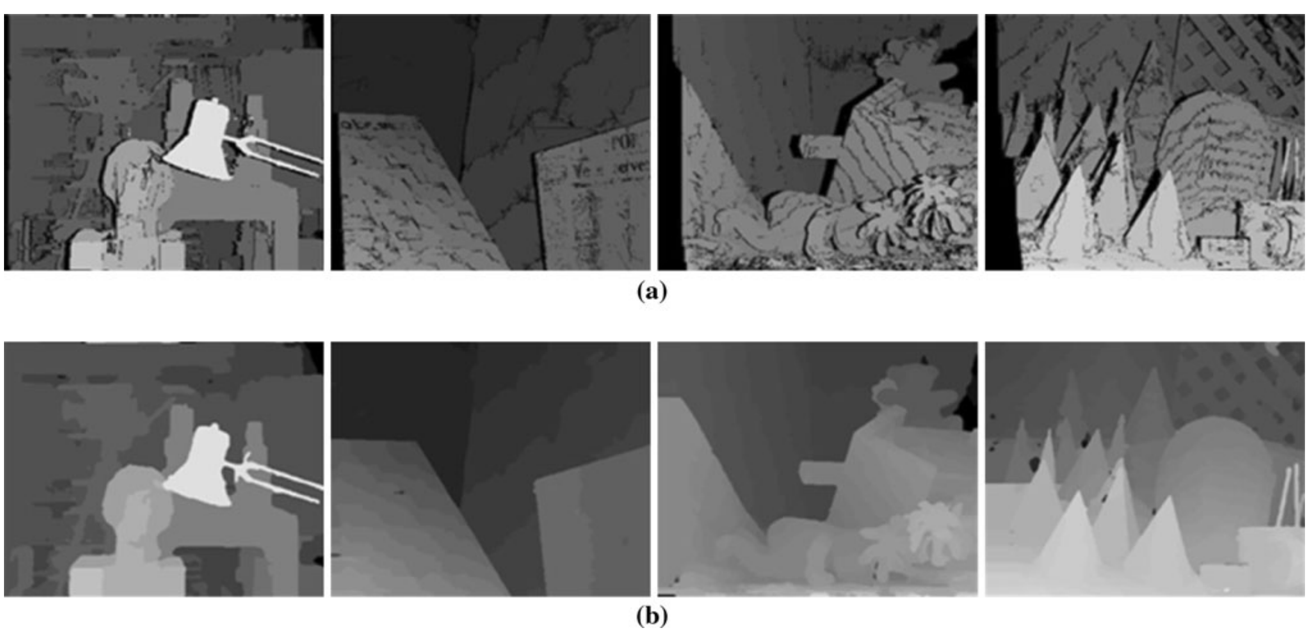
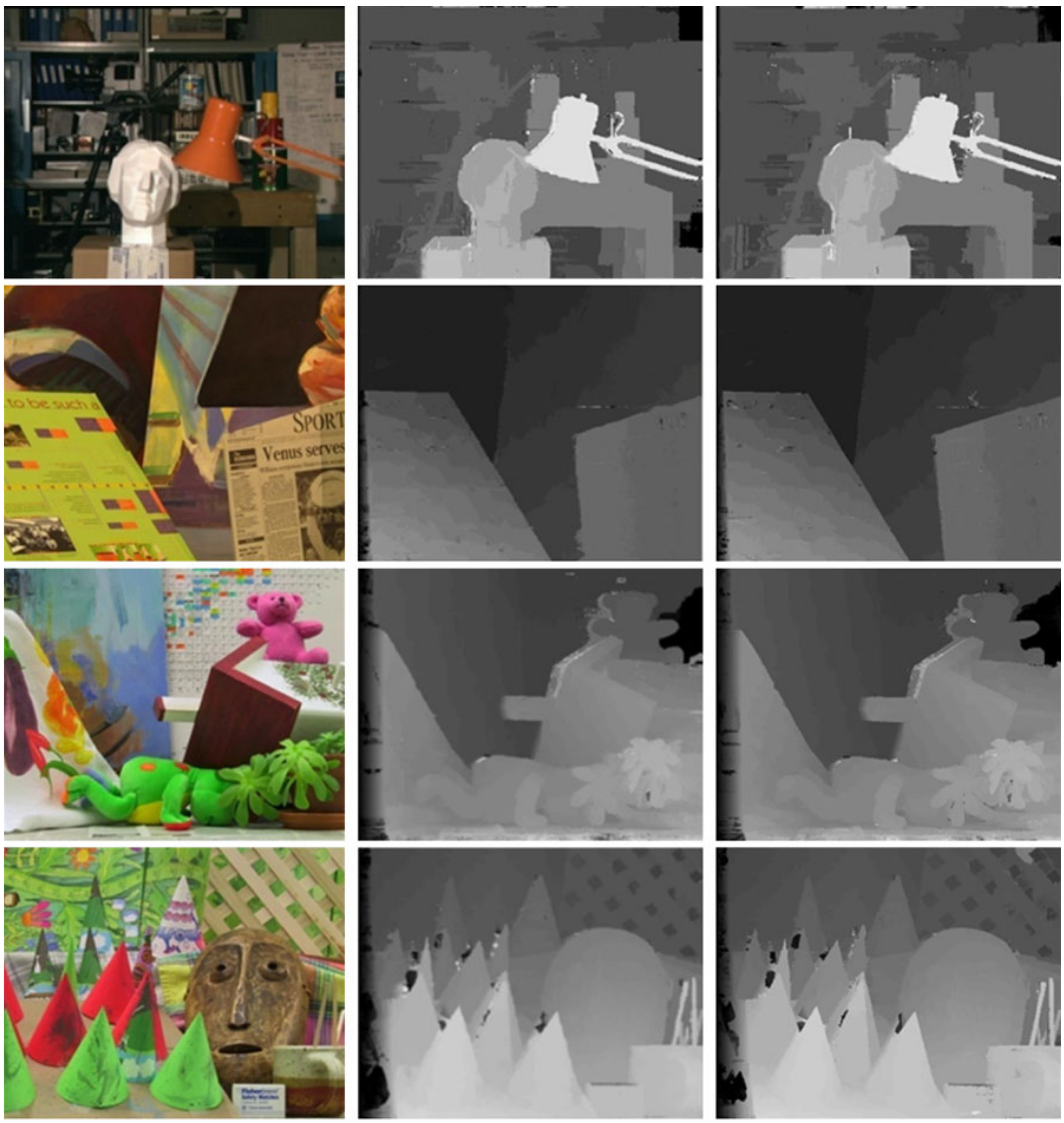


Fig. 3 Disparity maps **a** after left–right consistency check **b** filled inconsistent disparities

Table 1 Comparison between proposed approach and method [3, 4] on the Middlebury dataset

	Tsukuba		Venus		Teddy		Cones	
	Nonocc	Disc	Nonocc	Disc	Nonocc	Disc	Nonocc	Disc
Proposed method	1.76	6.50	0.99	4.46	10.0	19.4	5.04	10.7
Segment support [3]	2.05	7.14	1.47	10.5	10.8	21.7	5.08	12.5
Adaptive weight [4]	4.66	8.25	4.61	13.3	12.7	22.4	5.50	11.9



(a) Reference Images (b) Disparity maps using [3] (c) Disparity maps by the proposed method

Fig. 4 Comparision of disparity maps generated by the proposed method and Tombari et al. [3]

depth discontinuity. The disparity maps computed by the proposed method are evaluated using the testbed online error checking system [12]. The website measures the percentage of bad matching pixels with absolute disparity error greater than one pixel. The *nonocc*—represents the disparity error computed only in non-occluded regions and *disc*—represents the error is evaluated near depth discontinuity.

Conclusions

In this paper an improved support aggregation strategy based on a hybrid system of adaptive weight and segmentation for disparity map computation is presented. The proposed method utilizes the color similarity, spatial proximity as well as segmentation information into an integrated framework. The system generates improved disparity in general and handled disparity at discontinuity boundaries in better manner as compared to other methods in its category. The future research direction will be to implement the algorithm for real-time implementation.

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

References

1. R. Jain, R. Kasturi, B.G. Schunck, *Machine Vision* (McGraw Hill International Edition, New York, 1995)
2. M.Z. Brown, D. Burschka, G.D. Hager, Advances in computational stereo. IEEE Trans. Pattern Anal. Mach. Intell. **25**(8), 993–1008 (2003)
3. F. Tombari, S. Mattoccia, L. Di Stefano, Segmentation-Based Adaptive Support for Accurate Stereo Correspondence, IEEE pacific-rim symposium on image and video technology (PSIVT), LNCS, vol. 4872, (Springer, Heidelberg, 2007), pp. 427–438
4. K. Yoon, S. Kweon, Adaptive support-weight approach for correspondence search. IEEE Trans. Pattern. Anal. Mach. Intell. **28**(4), 650–656 (2006)
5. D. Scharstein, R. Szeliski, A taxonomy and evaluation of dense two frame stereo correspondence algorithms. Int. J. Comput. Vis. **47**, 7–42 (2002)
6. L. De-Maeztu, S. Mattoccia, A. Villanueva, R. Cabeza, Linear stereo matching, *13th International Conference on computer vision*, Barcelona, Spain, 2011
7. M. Gong, R. Yang, L. Wang, M. Gong, A performance study on different cost aggregation approaches used in real-time stereo matching. Int. J. Comput. Vis. **75**(2), 283–296 (2007)
8. C. Richardt, D. Orr, I. Davies, A. Criminisi, N. Dodgson, Real-time spatiotemporal stereo matching using the dual-cross-bilateral grid, European conference computer vision, Springer Berlin, 2010, pp. 510–523
9. M. Gerrits, P. Bekaert, Local stereo matching with segmentation-based outlier rejection, in *Proceedings of Conference on Computer and Robot Vision*, 2006, pp. 66–73
10. F. Tombari, S. Mattoccia, L.D. Stefano, E. Addimanda, Classification and evaluation of cost aggregation methods for stereo correspondence, in *Proceedings of computer vision and pattern recognition*, 2008, pp. 1–8
11. D. Comaniciu, P. Meer, Mean shift: a robust approach toward feature space analysis. IEEE Trans. Pattern. Anal. Mach. Intell. **24**(5), 603–619 (2002)
12. http://www.middlebury.edu/stereo. Accessed 1 March 2012