Efficient Disparity Estimation Using Region based Segmentation and Multistage Feedback

LESLIE JIA[1], MRINAL MANDAL[1], and THOMAS SIKORA[2]

[1]Dept. of Electrical and Computer Engineering, University of Alberta, Edmonton, CANADA
[2]Institut für Telekommunikations Systeme, Technische Universität Berlin, Berlin, GERMANY
ljia@ualberta.ca, mandal@ece.ualberta.ca, sikora@nue.tu-berlin.de

Abstract: - Stereoscopic analysis is widely used in machine vision applications. Local and global methods are two main branches of stereoscopic analysis. The global methods typically minimize a cost function over the entire scene. Although these methods provide high estimation accuracy, because of its high complexity, they are not suitable for real-time implementation. The local methods typically use window-correlation approaches, and the associated complexity is generally low. However, the estimation accuracy is sensitive to the selected window size. In this paper, we propose a multistage local method that operates on image segments instead of traditional rectangular windows. This new approach exploits the unique characteristics of image segments, and reduces occlusion through a feedback system. Experimental results show that it is very effective for natural images. In addition, it has a low computational complexity which may be suitable for real-time implementation.

Keywords: Stereo Correspondence, Disparity Estimation, Progressive Processing, Local Methods.

1. Introduction

Depth estimation is an important tool in several applications such as machine vision, robotics, and satellite terrain mapping. With recent advances in 3D consumer video communications technology [1], application of depth estimation is likely to grow significantly in near future. The depth estimation techniques are of two types. Depth estimation using Laser or infrared ranging techniques are precise and popular. However, their applications are limited for certain tasks. For example, it is not advisable to laser-scan a live human. Stereoscopic methods, on the other hand, are purely passive and uses a pair of cameras (left and right) to map a scene. The lateral shift between the images captured by the two cameras, known as disparity, is used to estimate the depth of different parts of a scene. The stereoscopic depth estimation (SDE) is very convenient in image processing applications such as range detection, image rendering and 3-D scene reconstruction.

The SDE methods can be broadly classified into two categories: local methods and global methods. The global approaches typically define an energy function [2, 3] that balances the uniqueness and smoothness constraints. Optimal results are obtained when the energy function is minimized. Computational techniques such as dynamic programming and simulated annealing are typically used in the optimization algorithm. Although the problem formulation of global approaches is simple and elegant, the computational complexity is very high. The local methods, on the other hand, generally use area based matching for disparity estimation. Here, a window of pixels in one image is matched with a slightly shifted window of pixels in the other image. The disparity is estimated to be the shift that results in the minimum matching cost. Typically, the local methods have much less complexity compared to global methods.

A major difficulty of the window-based local methods is the selection of a proper window size. If the window size is small, there may not be enough image features in a small window, and the estimated disparity may not reliable at low texture area. In the extreme case, a window of 1×1 pixels will likely have many potential aliases, and the uniqueness constraint may not be satisfied. On the other hand, if the window size is large, the smoothness constraint may not be satisfied as the window may contain two or more sub-regions of different disparities. It generally results in blurred disparity boundaries. Therefore, it is a challenging task to choose windows that are homogenous in disparities and yet contain sufficient features.

Several techniques have been proposed to reduce the limitation of fixed windows. Fusiello *et al.* proposed a multi-window approach [4]. Instead of analyzing only one window centered at the subject pixel, this method compares nine candidate windows of equal sizes. The candidate that offers the minimum matching cost is selected. Hirschmüller *et al.* [5] have proposed a multiple window approach to decrease estimation error at object borders. Veksler proposed an approach where the windows can take on different sizes [6]. For each pixel, many windows of varying sizes are examined. Similar to the multi-window approach, the window of the highest correlation is chosen. Sun proposed a multi-resolution approach [7]. Starting with a coarse version of the image, obtained by decimation, this method refines the disparity map step by step, as the image resolution improved. At each step, an intermediate disparity map is first obtained. It is then divided into horizontal stripes and adjacent stripes of similar disparity are merged into one. Thus, each sub-region reflects the size and shape of the object much better. Agrawal *et al.* [8] investigated the relationship between window sizes and disparity levels, and found that most windows have at most two disparities. Therefore, they introduced the concept of bi-labeled windows where each pixel is assigned two disparity candidates. A global optimization is then performed to estimate the disparity.

Although, the above methods improve upon the basic fixed windows, these approaches do not address discontinuity situations well enough. Recently, region based correlation matching techniques [9] have been proposed to improve the estimation performance. In these techniques, an image is first segmented and each segment is then matched to obtain the disparities. However, the segmentation approach also has some limitations. For example, it is difficult to decide correct size of the segments. In addition, it is difficult to accurately determine the border of a segment. Another strategy for improving the performance of stereo matching is to use progressive techniques [10]. In this method, the disparity map is first calculated for pixels with highly reliable matches. In next iterations more pixels are added progressively.

In this paper, we propose a multistage segmentation based local method for disparity estimation. In the initial stages, larger segments are used to calculate the disparity values. The disparity is refined iteratively with smaller segments. A feedback loop is employed to enforce directional consistency in each stage, and improve the robustness.

The organization of the paper is as follows. A brief review of background work is presented in section 2. The proposed technique is presented in section 3. The performance of the proposed technique is presented in Section 4, which is followed by the conclusions.

2. Review of Background Work
In this section, we present a brief review of works related to SDE with an emphasis on the local methods.

2.1 Geometry
In a stereoscopic system, the physical distance (or depth) L between an object and the camera follows the relationship $L = f * D / d$, where f is the focal length, D is the distance between the two lenses and d is the spatial displacement (known as *disparity*) between the object's projected images in the two views. As f and D are constants, $L \propto 1/d$. Thus, the estimation of depth L essentially becomes the estimation of disparity d. Therefore, the terms "depth estimation" and "disparity estimation" are used interchangeably in stereoscopic literature. This simple model also works well for more complex systems, such as human eyes, as long as $L \gg d$.

2.2 Local Methods
The local methods generally use area/region based matching for disparity estimation. Here, one assumes that a pixel is surrounded by a window of pixels with equal disparity. To estimate the disparity for pixel m in the left image, a window of pixels around m is selected, and let it be denoted by $\Delta(m)$. This window is matched with a corresponding window $\Delta'(m)$ in the right image. The center of $\Delta'(m)$ is shifted horizontally by d to get a better match with $\Delta(m)$. Let the shifted window be denoted by $\Delta'(m+d)$. The matching cost between these two windows can be expressed as

$$\partial\left(L_{\Delta(m)}, R_{\Delta'(m+d)}\right)$$

where $\partial(.)$ is a distance function, $L_{\Delta(m)}$ are the pixel gray values in the left image, and $R_{\Delta'(m+d)}$ are the pixel gray values in the right image.

The shift distance d for which the cost is minimum is generally considered as the correct disparity. In other words, the disparity corresponding to the pixel m is given by

$$d(m) = \underset{d_{\min} \le d \le d_{\max}}{\arg\min} \ \partial\left(L_{\Delta(m)}, R_{\Delta'(m+d)}\right) \qquad (1)$$

where $d_{\min}, d_{\max}$ are the minimum and maximum allowable disparities.

For simplicity, many SDE techniques use rectangular window. The distance functions are typically Euclidian or Manhattan distance. The disparity estimation using Eq. (1) requires much less computational complexity compared to the global methods. However, a major limitation of this approach is that if the window size is small there may not be enough image features in the window, and the estimated disparity may not reliable at low texture area. On the other hand, if the window size is large, the pixels in the window may have different disparities, resulting in blurred disparity boundaries.

3. Proposed Technique

In this section, we present the details of the proposed algorithm.

3.1 Proposed Multistage Approach

The overall architecture of the proposed multistage feedback algorithm is shown in Fig. 1. Here, several stages of processing are executed hierarchically. Each stage extracts a certain subset of the original test image and processes that subset with regard to its characteristics. The general rule is that large, uniform segments are processed first, which is somewhat similar to the processing of low-frequency signals first in multistage image processing.
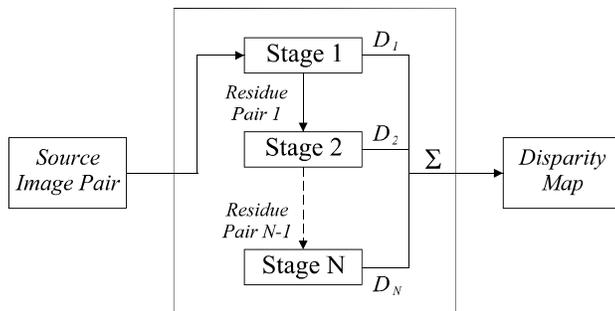

Fig. 1. Schematic of the proposed multistage disparity estimation algorithm.

Each stage, by itself, is essentially an implementation of a local stereo method. However, instead of a "winner-take-all" decision, the proposed algorithm does not force to make a disparity selection. A "no-winner" situation is allowed, when no good match is found. The disparity of a local region is asserted only if it is considered to be of high confidence. The "no-winner" regions are passed onto the next stage, where a different algorithm has a chance to perform better. The output of each stage would be a partially finished disparity map plus a residue image pair, whose disparity is yet to be decided. The partial

disparity maps are eventually aggregated to form a complete disparity map.

All different stages in the proposed multistage concept share the same basic layout. The common structure is illustrated in Fig. 2. We use segment-correlation algorithms as the basic building blocks in our system. The disparity computation and aggregation is virtually identical for each stage. Their main difference is in the segmentation processes.
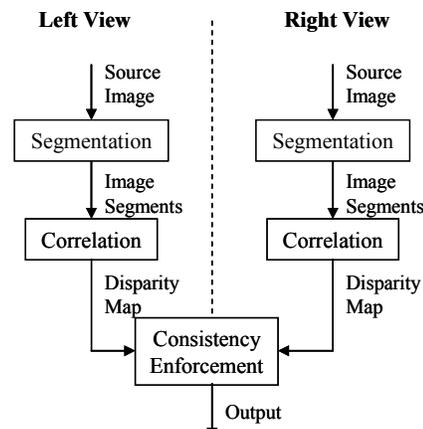

Fig. 2. Basic layout of a stage.

3.2 Image Segmentation

In this paper, we use segmented image regions instead of rectangular window for feature matching. We assume that segments of uniform intensities have uniform disparities also. In other words, a segment contour is a disparity boundary as well. The size and position adjustments of a segment come naturally with a properly selected segmentation algorithm. Therefore, the smoothness constraint is nicely enforced.

There is another advantage of using segments. A traditional rectangular window needs to contain a substantial amount of feature to function properly. However, a window may contain little or no feature, and therefore may not be able to provide good performance. The segmented regions do not have this limitation as the segmented boundaries may be considered as features. This is illustrated in Fig. 3 where there are two featureless objects. As the objects do not have features (or intensity variation) inside them, it is difficult to estimate disparities with small windows such as C (in Fig. 3(b)). However, the feature contours (in Fig. 3(c)) A′ and B′ can be used to match the objects. In other words, the use of segmentation can effectively handle the presence of featureless regions. The large featureless objects may fall into the "unique match" category or, if

occlusion and boundary situation are dominating, the "no match" category.

In this paper, we have used standard image segmentation algorithms such as region growing, split and merge, and vector quantization [11] to segment the images. Although, we have not found great advantage of one technique over others, the segmentation parameters, however, need to be selected carefully to obtain good results.

3.3 Consistency Enforcement

This section elaborates the concept of disparity consistency enforcement previously introduced in the stage layout diagram. Note that window-correlation method is a directional process, and matching is not a symmetric process [4]. When we match from both directions (from left to right, and from right to left, in most cases), we expect to obtain different sets of conjugate pairs. These inconsistencies may occur frequently, due to signal noise and different segment/window definition, and are observed even when $1 \rightarrow 1$ correspondence between the left and right images can be established.

Original Image   Rectangular Window   Segment Contour
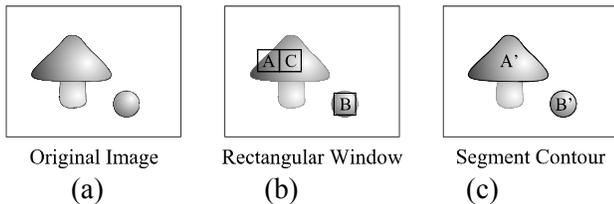(a)   (b)   (c)

Fig. 3. Comparison of rectangular region and image segments.

The left-right inconsistency is very important for handling occluded regions. Though this property has been used for windows [4], the significance of this property is more evident for image segments. Fig 4 illustrates a stereo pair: the left image represents the two views. Segments A and B are corresponding conjugates. As shown, we can either shift segment A rightward to match B, or shift segment B leftward to match A.

The domain of this new disparity map is not identical to the left disparity map. Yet, they overlap for the most part. Ideally, where they do overlap, the disparity values should agree exactly. If not, then a voting mechanism must be used to decide which one is better. Generally speaking, this step belongs to the aggregation process. When occlusion is present, the cost curve is likely to be the "no match" type as shown in the centre image in Fig 4. In this case, the aggregated result is likely to be wrong.

The combination of symmetric processing and segmentation can solve some simple cases of occlusion.
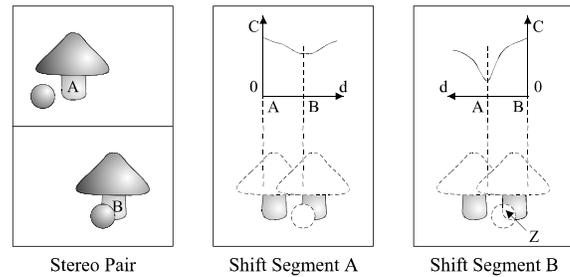

Stereo Pair   Shift Segment A   Shift Segment B

Fig. 4. Asymmetric cost curves.

3.4 Feedback System

The consistency enforcement is difficult to achieve in one pass. Hence, in this paper, we use a feedback system to resolve the inconsistencies. Let us designate $D_L(x, y)$ and $D_R(x, y)$ as the disparity maps for the left image and the right image, respectively. Then, the right-view equivalent of the left disparity map can be defined as $D_{R/L}(x - D_L(x, y), y) = D_L(x, y)$. Similarly, the left-view equivalent of the right disparity map can be defined as $D_{L/R}(x + D_R(x, y), y) = D_R(x, y)$. As shown in Fig. 5, left and right disparity maps are first transformed into the other view. Under each view, the two disparity maps are merged into one. When inconsistency occurs, the disparity with the smaller matching cost is selected.
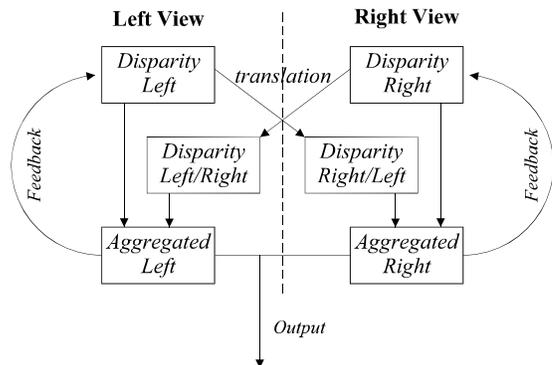

Fig. 5. The feedback system.

Note that during implementation, the view translation is performed from the minimum disparity value, in ascending order, to the maximum disparity value. A disparity value changed in one pass may themselves induce another change in the next pass. On the other hand, convergence is guaranteed for two reasons. First, matching costs are positive, by definition. Second, each disparity change will result in either a no-change or a

decrease of the total matching cost, where the decrement is always a multiple of a pre-defined constant.

Experimental results show that 30-70% of all the pixels are affected by the consistency enforcement during the first pass. The amount of affected pixels decays by about a factor of 100 for each subsequent pass. Since the test pictures we use typically have 100,000 to 200,000 pixels, at most three passes are needed before the output is stabilized. Among the affected pixels, most keep their disparity value and only have their matching costs reduced. A few of them have both their disparity and matching cost values updated.

The consistency enforcement is typically used for segments adjacent to occluded areas. For the occluded areas themselves, we use the smoothness constraint to calculate disparity. For tiny and uncluttered segments, the estimation is generally unreliable. To reduce the problem, we use 1-D scan line segmentation instead of 2-D segments and process horizontal and vertical lines separately. Once a scan line is broken into several segments, the prevailing disparity value in each segment is assigned to all the pixels in that segment.

3.5 The Last Stage

Previous sections proposed a multi-stage concept. There are not a fixed number of stages quoted. Actually, the implementation can be flexible to suit the unique need of each application. There can be as many stages as needed, as long as the performance is still improving. However, the most performance improvements are accomplished in the first several stages. After that, the improvement saturates very fast, and more stages are no longer justified.

The last stage is quite different from all other stages. Since all residue pixels must be processed in the last stage, image segmentation is no longer effective. Thus a traditional rectangular window technique is employed. To improve its performance with minimum additional complexity, a fixed-sized and position-varying technique as in [4] is used. For each residual pixel, an appropriate window within a pre-defined distance is selected to achieve the smallest matching cost.

4. Experimental Setup

In this section, we present the details of algorithm implementation and experimental set-up.

4.1 Pre-processing

Before the matching process actually starts, a few preparation steps need to be done. First, the sources images are low-pass filtered to be smoother. This would make more accurate segmentations. The filter is a typical 3x3 Gaussian matrix. To preserve contrast at edges, all the smoothing are performed at non-boundary locations. For simplicity, we choose to use the algorithm provided in [3] to define image "boundaries".

For any matching algorithm implementation, a range of possible disparity values are assumed. Unfortunately, most literatures in this area do not explicitly state the detailed algorithms involved. In our previous work, we have set the range to be the known range of the ground truth map. In other words, we have avoided the problem, in order to focus on the multistage algorithms.

We have improved our algorithm by adding a very simple yet robust way to estimate the range. The source image pair first undergoes a typical window matching—the algorithm is actually shared with the last stage of matching. All non-featureless pixels with better than average matching cost are considered to be of high confidence. The disparity values of these pixels are built into a histogram with Gaussian distribution. When the insignificant tails are truncated, the remaining histogram makes good disparity range estimation.

4.2 Segmentation

The implementation of a good segmentation algorithm is crucial to obtain a good depth estimation performance. Since the proposed depth estimation algorithm is intended to be fully automatic, the algorithm must operate with no human input. In this paper, we have implemented two image segmentation algorithms: region growing (first stage), and vector quantization (second stage). The implementation details are as follows.

Region Growing

Seed selection: The image is scanned from left to right and from top to bottom in a zigzag path. The first pixel encountered that is not part of an existing region is regarded as a seed.

Growing: The seed itself can be viewed as a region with only one pixel. For any region, all of its neighbouring pixels not yet grouped are investigated. Ones that satisfy the pre-defined threshold are included to the region and labelled as grouped.

Threshold: A pixel can be added, or grown into, an existing adjacent segment if intensity criteria are met. Let T be a predefined scalar threshold value. Let the average intensity of a segment be (r', g', b'). An adjacent color pixel with components (r, g, b) is included into the segment if the following two conditions are satisfied.

i) $\max\left(|r - r'|, |g - g'|, |b - b'|\right) < T$ and

ii) $Y = 0.299r + 0.587g + 0.114b < 0.8T$

In our software implementation, T is set to be 12.

Vector Quantization
In vector quantization (VQ), the first step is to generate the codebook, which is a suitable set of N color vectors. In our implementation, we have used $N = 9$. Initially, the color vectors in the codebook are set as $\{\phi_i, \phi_i, \phi_i\}$, $0 \leq i \leq 8$ where $\phi_r = r \cdot \Delta + \Delta/2, \Delta = 256/(N-1)$. Each pixel (r, g, b) in the image is assigned the color vector with minimum Euclidian distance $\left(\sqrt{(r-r')^2, (g-g')^2, (b-b')^2} \right)$ where (r', g', b') is a color in the codebook. The vector itself is then updated to be the centroid of these assigned pixels. This process is executed iteratively until there is no more change in vector assignment or when a preset number of updates are reached.

After the codebook is calculated, the entire image is divided into multiple segments. Each segment contains pixels corresponding to only one vector. Any two adjacent segments are to be merged, if both of them correspond to the same color vector.

4.3 Post Processing
After all the stages of matching are finished, the disparity map still needs more refinement. We apply Gaussian smoothing to both horizontal and vertical scan lines of the disparity map. Secondly, to remove artefacts, we segment the obtained disparity map. Each segment contains only pixels of the same disparity and no adjacent segments shall have the same disparity. The disparity segments with less than 60 pixels are then considered unnaturally small and thus an artefact. The disparity values for these pixels are then discarded and the void is filled by low-pass filtering.

4.4 Performance Evaluation Criteria
In this paper, we use the *percentage of wrongly matching pixels* (PWMP) as the performance criterion. For a stereo image pair, we estimate the disparity for each pixel in units of pixels. Let the estimated disparity, and the true disparity of pixel (m, n) be denoted by $d_e(m, n)$ and $d_t(m, n)$, respectively. For an image with size $M \times N$, the PWMP is defined as:

$$\eta = \frac{100}{M \cdot N} \sum_{(m,n)} \left(\left| d_e(m, n) - d_t(m, n) \right| > \delta_d \right) \qquad (2)$$

where $1 \leq m \leq M$, $1 \leq n \leq N$, and δ_d is a threshold for error tolerance. In this paper, we set δ_d as 1.

Note that the performance of an algorithm is sensitive to many image parameters such as texture, and edges.

Therefore, for a given stereo image pair, we calculate η for three types of regions.

i) Entire Image. This provides the overall performance. This is referred to as "ALL" in Tables 1, 2, and 3.

ii) Textureless regions: Here, the horizontal gradient is small. This is referred to as "TL" in Tables 1, 2, and 3.

iii) Depth discontinuity regions: Here, the neighboring disparities differ substantially. This is referred to as "DISC" in Tables 1, 2, and 3.

5. Performance Evaluation
In this section, we present the performance of the proposed technique. Simulations have been carried out using four *Middlebury* stereoscopic image pairs [3]. The images represent different content types. Highly accurate (ground truth) disparity maps for these images are known a priori. One of these image pairs (the Tsukuba) is shown in Fig. 6 along with its ground truth disparity map. We estimate the disparity using the proposed technique and calculate η using Eq. (2).
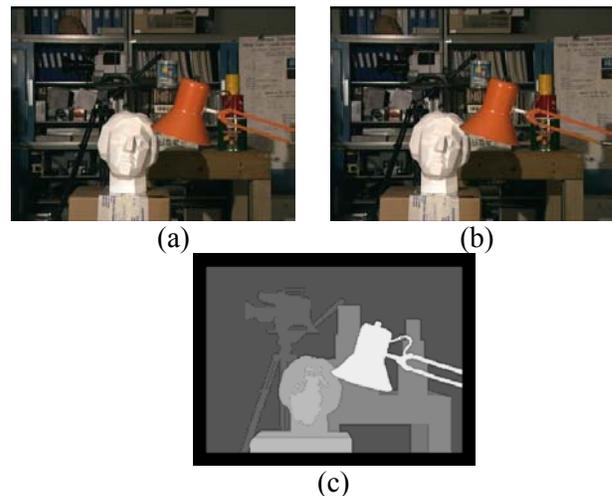

Fig. 6. Tsukuba test image. a) Left view, b) Right view, and c) ground truth disparity map. The ground truth disparity map corresponds to the left view image.

The performance of the proposed technique is expected to improve as the number of stages increase. However, in our experiments, it has been found that performance does not improve much beyond three stages. Therefore, three stages are used in our implementation. A region-growing segmentation technique [11] is used in the first stage and a vector-quantization segmentation technique [12] is used in the second stage. Fig. 7 shows the segmented Tsukuba

images in segments 1 and 2. In these two stages, only large uniform segments are used. The minimum size of the segments depends on the stages as well as the left and right images. In our implementation, we have set the minimum size to about 75 pixels.

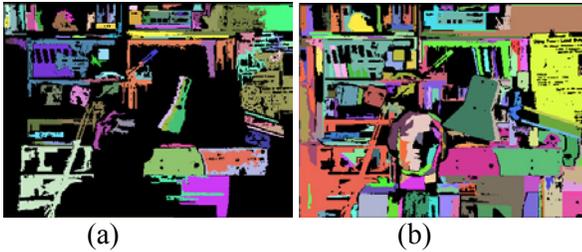<div align="center">(a)   (b)</div>

Fig. 7. Segmented Tsukuba images. a) Stage-1, b) Stage-2.

The mean square difference function is selected as the cost function. The disparity corresponding to the minimum cost is obtained for each segment. All residual pixels after the first two stages are processed in the third stage, where a typical position-varying window algorithm is used. The aggregated disparity map is processed by a smoothing algorithm for a last refinement.

Quantitative Comparison

Since we allow a partially finished disparity map from each stage (except the last one), it is important to know exactly what percentage of the test images are determined. Table 1 shows the percentage of the completion for each stage and its corresponding error rate. The estimation efficiency η is calculated before applying smoothing. It is obvious that early segmentation stages are very well suited to un-textured regions with a very low error rate.

Table 1: Percentage of finished disparity map at different stages for Tsukuba image. In entry $\alpha(\eta)$, α represents the percentage of finished disparity map, and η represents the PWMP. "TL" refers to the Texture-less regions, and "DISC" refers to the Discontinuity regions.

	Stage1	Stage2	Stage3
ALL	34.02(0.038)	38.96(0.033)	100(4.88)
TL	51.35(0.016)	57.41(0.014)	100(7.23)
DISC	15.52(0.487)	23.60(0.320)	100(11.73)

Table 2 reveals the performance advantage of the multistage method. Since stage 3 is the last stage in our implementation, it is intended to process all residue information. Therefore, stage 3 is always used. We can

see a clear pattern: with the help of segmentation stages, the percentage errors decrease substantially.

Table 3 compares the performance of the proposed algorithm with three other window-based methods: Real-Time Correlation (RTC) based method [5], fast variable window (FVW) method [6], and Windows-based discontinuity preserving (WDP) method [8]. The performance of the RTC and FVW methods was obtained from [6] and WDP performance was obtained from [8].

Table 2: Improvement in performance (η) at different stages.

Stage 1	Stage 2	Stage 3	ALL	TL	DISC
×	×	√	1.484%	0.542%	7.63%
√	×	√	1.144%	0.302%	6.29%
√	√	√	1.136%	0.302%	6.25%

It is observed in Table 3 that the proposed algorithm provides a superior performance for the Tsukuba image compared to other algorithms. The performance for the Map image is also very good (error is less than 1%), but not better than the other methods. This is primarily because the map image is not a natural image, and the image segmentation does not work well, resulting in performance degradation. However, as most real images are likely to be natural (like the Tsukuba image), this limitation should not pose much problem.

Table 3: Performance (η) comparison of the proposed method with other methods.

Image	Pixels	FVW [6]	WDP [8]	RTC [8,6]	Proposed Method
Tsukuba	ALL	2.35	1.78	4.25	1.14
	TL	1.65	1.22	4.47	0.30
	DISC	12.17	9.71	15	6.25
Map	ALL	0.24	0.32	0.81	0.70
	TL	NA	NA	NA	NA
	DISC	2.9	3.33	11.4	8.39

Qualitative Analysis

Fig. 8 shows disparity maps resulting from the "Tsukuba" test images. It is easy to see that the segment correlation approach does generate large segments of high confidence. Furthermore, the performance is especially good at sharp disparity discontinuities. The visual quality of the lamp's supporting rods is among the best in all results submitted to the Middlebury test bed website [13].

The results for "Map" is shown in Fig. 9. Note that the "Map" test image pair is black and white, with significant amount of white noise. The patterns are quite abstract and therefore the segmentation stages hardly yield any useful data. Most of the processing work is done by the circular window matching stage. The errors occur mostly at the left side of the map, due to occlusion. The smoothing procedures failed to propagate correct disparity values to the occluded areas mostly because of the complicated nature of the image.
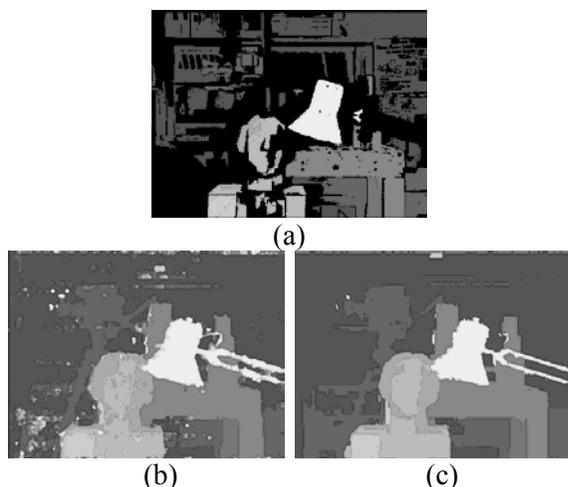

Fig. 8. Estimated disparity maps of the "Tsukuba" image. Disparity maps a) after stage 2, b) after stage 3, and c) final.
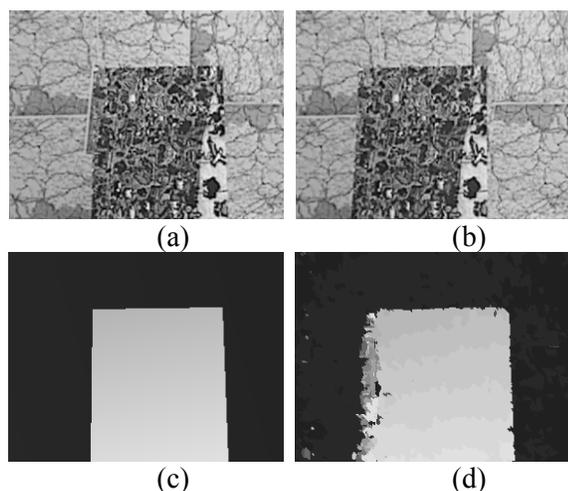

Fig. 9. "Map" image. a) Left image, b) right image, c) Ground truth disparity, and d) Estimated disparity map.

6. Conclusions

In this paper, we propose a multistage disparity estimation method. The proposed method segments a stereo image pair, and estimates the disparity for each segment. It also employs a multistage algorithm where consistency check of the estimated disparity is carried out at each stage. Experimental results show a good estimation performance with more than 98.5% accuracy for the test images considered. The method has low complexity and may be suitable for real-time analysis.

References
[1]   O. Schreer, P. Kauff, Peter, T. Sikora, *3D Videocommunication*, John Wiley & Sons, 2005.
[2]   D. Marr and T. Poggio, A theory of human stereo vision, *Proc. Royal Society of London*, Vol. B204, 1979, pp. 301-328.
[3]   D. Scharstein and R. Szeliski, A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms, *Intl. Journal of Computer Vision*, Vol. 47, No. 1, 2002, pp. 7-42.
[4]   A. Fusiello, V. Roberto, and E. Trucco, Symmetric stereo with multiple windowing, *International Journal of Pattern Recognition and Artificial Intelligence*, 14(8), Dec. 2000, pp. 1053-1066.
[5]   H. Hirschmuller, P. R. Innocent, and J. Garibaldi, Real-time correlation-based stereo vision with reduced border errors, *Intl. Journal of Computer Vision*, Vol. 47, No. 1/2/3, 2002, pp. 229-246.
[6]   O. Veksler, Fast variable window for stereo correspondence using integral images, *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition* (CVPR), Vol. 1, 2003, pp. 556-561.
[7]   C. Sun, "Fast Stereo Matching Using Rectangular Subregioning and 3D Maximum-Surface Techniques," *International Journal of Computer Vision*, vol. 47, No.1/2/3, pp.99-117, May 2002.
[8]   M. Agrawal and L. Davis, Window-based discontinuity preserving stereo, *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition* (CVPR), Vol. 1, 2004, pp. 66-73.
[9]   Y. Zhang and C. Kambhamettu, Stereo matching with segmentation-based cooperation, *Proc. of the 7th European Conf. on Computer Vision*, Copenhagen, Denmark, 2002.
[10] Z. Zhang and Y. Shan, "A progressive scheme for stereo matching," *Lecture Notes in Computer Science*, Vol. 2018, pages 68-85, Springer Verlag, March 2001.
[11] Gonzales and Woods, *Digital Image Processing*, Prentice Hall, 2002.
[12] P. C. Cosman, K. L. Oehler, E. A. Riskin and R. M. Gray, Using Vector quantization for Image Processing, *Proc. of the IEEE*, Vol. 81, Sept 1993, pp. 1326-1341.
[13] www.middlebury.edu/stereo Middlebury Testbed Website.