

# Dense Stereo Matching Based on Propagation with a Voronoi Diagram

Li Tang

H.T. Tsui

C.K. Wu

Dept. of Electronic Engineering Dept. of Electronic Engineering Dept. of Electronic Engineering  
Chinese University of Hong Kong Chinese University of Hong Kong Xidian University of China  
ltang@ee.cuhk.edu.hk httsui@ee.cuhk.edu.hk chkwu@xidian.edu.cn

## Abstract

*A new dense matching algorithm is proposed in this paper. It is based on propagation from  $N$  seed points, which have been matched reliably by feature tracking. The whole image is first divided into  $N$  cells by the Voronoi diagram of the seed feature points. Then corresponding relations are propagated from the seed in each cell until all pixels within this cell are processed. Modified sum of squared differences (SSD) is employed as the cost function in the propagation of matching according to a statistic model of disparity distribution within the window. The size of the window is adaptive. It is made inversely proportional to the texture density inside the window to increase the reliability of matching. A significant merit of the algorithm is that it can be applied to a wide range of image pairs including those with large disparities, with or without rectification. The algorithm has been verified with real images and the results show that it is both robust and accurate.*

## 1. Introduction

Traditional dense matching falls into two categories of approaches: an approach based on a local method; another based on a global optimization. The former compares intensity similarity of pixels within a window between a pair of images to decide whether the centre points of the windows are a pair of corresponding points. In this approach, the selection of an appropriate window size is critical to achieve a smooth and detailed disparity map. The optimal choice of a window size depends on the local amount of variation in texture and disparity. As for the global optimization, sometimes better results can be achieved. Its objective is to find a solution that minimizes a disparity function, which includes a data term and a smoothness term. The data term measures how well the disparity function agrees with the input image pair while the smoothness term encodes the smoothness assumptions made by the algorithm. An important problem for global algorithms, however, is to find the right balance between data and smoothness terms [10].

Among the first category is the adaptive window method proposed by Kanade [3]. The window size and shape are iteratively changed according to the local textures and current depth estimates. Though it improved the matching result significantly, it is extremely computationally expensive. In the second category, attempts at using dynamic programming for solving stereo matching problem [7] using edges as the basic primitives have been reported. A generalization of the dynamic programming algorithm transforms the stereo correspondence problem into a maximum-flow problem [9]. Once solved, the minimum-cut associated to the maximum-flow yields a disparity surface for the whole image at once. This algorithm provides a more accurate and coherent disparity map than the traditional line-by-line stereo. By introducing global constraints and adopting a cooperative iterative algorithm, a fairly good dense disparity map is obtained in [15]. Its computational complexity is less than that of the adaptive window method. But both the amount of memory needed and the complexity of computation are still high. This is especially so when the disparities are large. Later, this algorithm is extended in choosing local support areas by enforcing the image segmentation information [14].

Dense matching algorithm using region growing [4] has shown good performances. However, the methods developed so far can only be applied to images rich in textures. Propagation is no good in regions that are too smooth. Recently a quasi-dense matching algorithm is proposed [5] which propagates the matches from the most textured pixels to less textured ones. In [11], the stereo matching problem is formulated as a Markov network consisting of three coupled Markov random fields. A new matching cost based on the reconstructed image signals is derived in [12]. Okutomi et al. [8] proposed to detect the region where the object boundary is likely to occur and adopts appropriate methods for these regions. Hierarchical stereo algorithm has also been researched [6]. The images are down-sampled an optimal number of times and the disparity map for a lower level is used as 'offset' to guide the computation of disparity map at a higher level. Its complexity is independent of the disparity range. A genetic algorithm [2] optimizes both the compatibility between corresponding points and the continuity of the

disparity map, which removes mismatches caused by false targets.

In this paper, a new dense matching algorithm is proposed. The Voronoi diagram of a number of feature points, which have been matched reliably, divides the whole image into many cells. Each cell contains a feature point, which is taken as the seed for propagation inside this region. Correspondences of the 8 neighbouring points of each seed are found using the disparity of this centred point under the continuity constraint. In this way, corresponding relations propagate from the seeds towards boundaries of the Voronoi diagram. A new cost function is employed in the measurement of intensity similarity with an adaptive window. The size of the window is changed dynamically according to the texture density within it. Full search is performed at possible regions of sharp depth discontinuities where continuity constraint might be violated. These techniques improve the performance of existing matching algorithms based on propagation. The details are described in Section 2. Some special issues in implementation are discussed in Section 3 and the algorithm is summarized in Section 4. Section 5 provides experimental results with real image pairs, which demonstrate the accuracy of the algorithm. Finally, some concluding comments are given in Section 6.

## 2. Dense matching based on propagation with Voronoi Diagram

There are mainly three stages in our algorithm. At the start, a number of seed feature points are extracted and matched by feature tracking. Then the Voronoi diagram of these seeds is computed to divide the image into many cells so that each seed is contained in a cell. Finally, matching relations are propagated from the correspondences of the seed feature points towards boundaries of these cells until all of the matched regions are merged together.

Propagation makes use of the continuity constraint. Surfaces of objects are assumed to be smooth under this constraint, which means its disparity varies continuously. Propagation is a simple and effective way of using this constraint to solve the problem of dense matching. Its basic idea is as follows. If a point A in the left image is corresponding to a point A' in the right image, the matching point B' of a point B in the neighbourhood of point A must be in the neighbourhood of the point A'. Thus the search area of B' is restricted in the neighbourhood of A'. Once a pair of matching points has been found with accuracy, the search areas of 8 points surrounding it are reduced to small regions if there are no large depth discontinuities. If the disparity level of the new correspondence is within a reasonable range, they are added to the set of the seeds to produce more matches.

The basic propagation strategy from a seed to its two neighbouring points is shown in Figure 1. In this figure, s and S is a pair of seed points while r and b are the two neighbouring points of point s respectively in the left image. The expected positions of their correspondences R and B in the right image are localized according to their relative positions with point s as is shown in Figure 1(b), in which the search windows for point R and B are shown by two rectangles with bold lines. Here the size of the search window is 4×4 pixels. It can also be seen from Figure 1 that the order constraint is incorporated implicitly in the process of propagation.

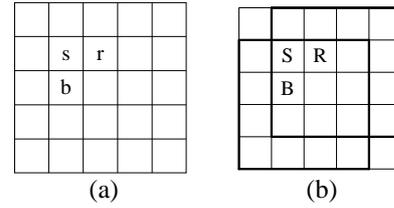


Figure 1: Propagation strategy from a seed point to two points in the neighbourhood of it. (a) Left image. (b) Right image.

There are two popular measurements of the intensity similarity of two locations: correlation coefficient and SSD. Both matching costs are defined over a certain area of support. The standard SSD in intensity between  $N_1(i_0, j_0)$  and  $N_2(i_0, j_0)$  is:

$$E(i_0, j_0) = \sum_{(i,j) \in N(i_0, j_0)} [I_1(i, j) - I_2(i, j)]^2 \quad (1)$$

where  $N_k(i_0, j_0)$ ,  $k=1,2$ , is a neighbourhood of the pixel  $(i_0, j_0)$  for image  $I_k$ ,  $k=1,2$ .  $I_1(i, j)$  and  $I_2(i, j)$  represent intensity values of the image pair at point  $(i, j)$  respectively. Computation of correspondences at each pixel is to find the position associated with the minimum SSD in intensity within the area of a search window:

$$[\hat{i}, \hat{j}]^T = \arg \min_{(i_0, j_0) \in \Phi} E(i_0, j_0) \quad (2)$$

where  $\Phi$  represents the area of the search window. In fact, the method of measuring intensity similarity of corresponding pixels in two windows to find matching points is valid only when all the pixels in the window have the same depth and foreshortening. The foreshortening problem can be tackled by rectification. If their depths are different, the disparity in the centre of the window must be different from other points. Using these points to support the matching of the centre point will cause some errors, which will blur edges in the resulting depth maps. However, disparities of the image pair can be assumed to change approximately continuously under the continuity constraint. A statistical model of disparity distribution within the window as described by Kanade et al. [3] can be adopted to decrease this systematic error. A typical model is as follows:

$$d(r, s) - d(0, 0) \sim N(0, \alpha \sqrt{r^2 + s^2}) \quad (3)$$

where  $\alpha$  is a constant that represents the amount of fluctuation of the disparity:

$$\alpha = \frac{1}{N} \sum_{(i, j) \in W(i_0, j_0)} \frac{(d_0(r_i, s_j) - d_0(0, 0))^2}{\sqrt{r_i^2 + s_j^2}} \quad (4)$$

where  $N$  is the number of the samples within the window.  $d(r, s)$  is the disparity which represents the matches as offsets to the points in the left image. This model assumes that the difference in disparity at a point  $(r, s)$  in the window from that of the centre point  $(0, 0)$  has a 2D zero-mean Gaussian distribution with variance proportional to the distance between these points. In other words, the expected value of the disparity at  $(r, s)$  is the same as the centre point, but it is expected to fluctuate more as the point is farther from the centre. Alternatively, the small surface corresponding to the window in the image is expected to be locally flat and parallel to the baseline statistically. However, the expectation becomes less certain as the window becomes larger. Thus we modified the above SSD according to this model:

$$E(i_0, j_0) = \sum_{(i, j) \in N(i_0, j_0)} w(i, j) [I_1(i, j) - I_2(i, j)]^2 \quad (5)$$

where  $w(i, j)$  is the added weight which has a 2D zero-mean Gaussian distribution:

$$w(i, j) = \frac{w_g(i, j)}{\sum_{(i, j) \in N(i_0, j_0)} w_g(i, j)} \quad (6)$$

$$\text{where } w_g(i, j) = e^{-((i^2 + j^2) / (2\sigma^2))} \quad (7)$$

The accuracy of the matching is improved significantly by adopting the above cost function since it fits the reality better. At the same time, the computational complexity introduced by the added weight is not so significant.

The local support area for an element determines which and to what extent neighbouring elements should contribute to averaging. Ideally, the local support area should include only those neighbouring elements that correspond to a correct match if the current element corresponds to a correct match. Since the correct match is not known beforehand, some assumption is required on deciding the extension of the local support. After edge extraction, we can assume that points in a window from which no edges can be extracted are on the same plane. As there are little textures within the window, a small window size may cause mismatch. So a larger one is adopted to improve the accuracy of the matching. In our algorithm, the size of the window is increased until the textures within it are above a fixed threshold or it has reached a maximum value. This will not lose many details and will not blur object boundaries in the resulting disparity map. At the same time, errors introduced by the large window will also decrease by adopting the modified

SSD as the cost function. It turns out to be able to deal with propagation in less textured areas in experiments.

### 3. Several Issues in Implementation

The Voronoi diagram of a collection of seed feature points is a partition of an image space into cells, each of which consists of those image points which are closer to one particular feature point than to any others. The boundaries of these diagrams are the so-called medial axes and their duals are Delaunay triangulation. The Voronoi diagrams are involved in situations where a space should be partitioned into "spheres of influence". So it is a good choice in our propagation algorithm. An example of the Voronoi diagram of a real image is shown in Figure 2.



Figure 2: Example of the Voronoi diagram of an image

In our algorithm, corresponding relations grow from each seed until the boundaries of the diagram are reached. Since in the process of 3D reconstruction, feature matching is the first step for estimation of epipolar geometry, a number of (100 – 150) matched feature points are available at this stage. So this requirement is reasonable and will not bring additional computation. It is obvious that it is better if we can have more corresponding feature points. After the first stage of feature tracking, more corresponding points can be added manually to make all of them distribute evenly in the image plane so that they cover the entire depth range in the appropriate parts of the image. It is true that our division of the image does not coincide with the edges of different depth. This can only be achieved by depth segmentation, as is mentioned in [14]. Perfect results can be obtained by propagation within the area of the same depth. But this method depends heavily on a good segmentation algorithm, which is now still being researched. So in this sense, our algorithm is more feasible than [14].

The key point of propagation is the assumption of depth continuities. The surface should be smooth since pixels are assigned disparities close to the disparities of already

matched pixels. However, continuity constraint is violated at sharp depth discontinuities. Search windows with just  $4 \times 4$  or  $5 \times 5$  pixels may not be enough to achieve correct matches in these areas. So we must detect these areas and enlarge the search ranges in these regions. One possible way of finding areas of depth discontinuities is as follows. Firstly, we apply operators of edge detection to find the edges in the images, which are then dilated by a  $3 \times 3$  square mask. That is, if any of the 9 data elements centred on the element of interest are unequal to zero, then a 1 is returned for the element of interest, and a 0 otherwise. In this way, the possible areas of depth discontinuities are detected to show when a full search should be performed. This will avoid the risk of bad propagation. Here a full search just refers to the maximum disparity range. This process is demonstrated in Figure 3. Figure 3(a) is the result of edge extraction and Figure 3(b) dilation. Although we cannot distinguish the true depth discontinuities from textures or shades, it is an effective and simple method to find possible areas of depth discontinuities.

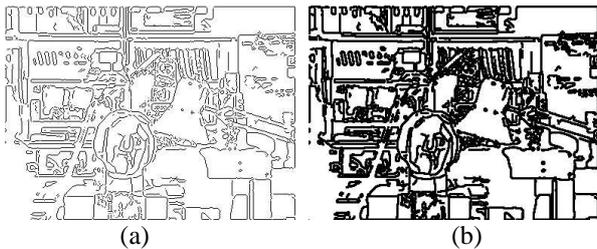


Figure 3: (a) Edge extraction. (b) Dilation of edges marks possible areas of depth discontinuities.

Since windows create problems at depth discontinuities, a simplified version of adaptive windows similar to that of Bobick et al. [1] can be adopted. At every pixel location different windows are used to perform the matching. Windows are located so that at an occlusion boundary, some of these windows will match across the boundary and some will not. At each pixel, only the best matching result will be stored. The effect of trying all shifted windows around a pixel is the same as taking the minimum matching score across all centred windows in the same neighbourhood. In this way, bad matches resulting from occlusion tend to be discarded.

When our algorithm is tested on image pairs without rectification, depth maps cannot be used to display the results. So a global way to visualize dense matches for arbitrary images is used [4]. Pixels of the first image are colored with a black-white checkerboard. For each matched pixel of this image, the corresponding pixels of the second image are set with the same color. This makes it easy to visualize the match of each square and its distortion. An even better way for displays is to blend a checkerboard with the original images.

When we deal with images without rectification, we only need to add the epipolar constraint to restrict the search area further or take it as a term in the cost function to compute the match values. After feature tracking, the fundamental matrix is available. So it is very flexible to do rectification or not when our algorithm is applied.

## 4. Summary of the algorithm

The algorithm is summarized as follows.

1. Feature tracking. Obtain reliable correspondences of  $N$  seed feature points.
2. Estimate the epipolar geometry if the image pair is not rectified.
3. Find the Voronoi diagram of these seed feature points. There is one Voronoi cell for each seed.
4. Take out a seed to generate the correspondences of its 8 neighbours. For each neighbour, the matching point is searched at the neighbourhood of the corresponding point of the seed. Modified SSD is used in the adaptive search window with epipolar constraint.
5. Correspondences generated from the already matched points are used to produce more matching points.
6. Correspondences propagate from the seed feature point in the middle of each cell until the boundaries of the Voronoi diagram are reached.

## 5. Experimental Results

Experiments are performed to evaluate the new dense matching algorithm. First, it has been applied to real images. The initial image pair of a church is shown in Figure 4. This is a difficult example due to the existence of repetitive patterns and areas with little texture. In spite of these, most of the points have been matched correctly. This is due to the new algorithm we developed. It is the propagation that avoids errors between repetitive patterns with a large variation in position. In order to illustrate the performance of the algorithm, no further processing for eliminating outliers has been carried out. In fact, a large part of the outliers could be eliminated by any commonly used techniques, such as RANSAC. The result of the visual matching checkerboard blend with the original image is shown in Figure 5. We can see that most of the errors occur in the middle part of the image where there is a glass window. This is because the basic assumption of Lambertian surfaces is violated in these regions. That means the appearance or intensity of the same point varies with viewpoint, as can be seen clearly from Figure 4. This accounts for the main errors in Figure 5. Another error in the right-bottom corner is caused by a lamp. As stated above, order constraint is enforced implicitly in the process of propagation. This constraint requires that the relative ordering of pixels remain the same between the

two views, which may not be the case in scenes containing narrow foreground objects such as this lamp.

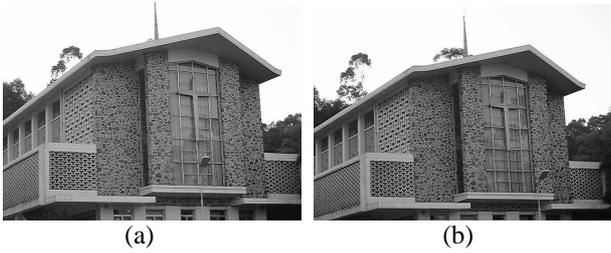


Figure 4: Initial image pair of a church. (a) Left image. (b) Right image.

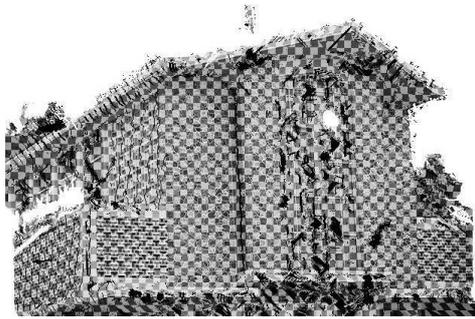


Figure 5: Result of the visual matching checkerboard blend with the original image

Our dense matching results are good enough to allow for a robust 3D reconstruction. RANSAC is applied to eliminate the outliers in dense matching. More than 70% of the correspondences are kept and they are enough to offer an acceptable reconstruction. The obtained 3D structure of the church is shown in Figure 6. The structure of the building is well recovered and the wall is smooth. This demonstrated the good performance of our dense matching algorithm. The top view is slightly distorted as we have no top view images.

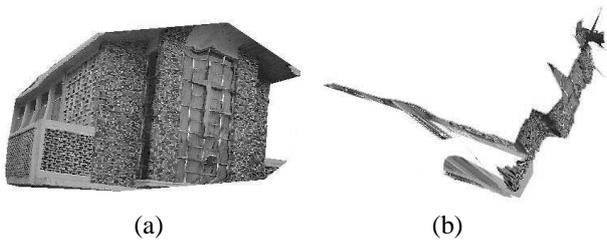


Figure 6: Reconstructed 3D structure of the church. (a) Front view. (b) Top view.

The new algorithm has also been applied to rectified images. The image pair of the church after rectification is shown in Figure 7. The size of the image is  $836 \times 490$  pixels and the range of the disparity is as large as 40 pixels. The resulting depth map is shown in Figure 8(a).

Similar to Figure 5, errors mainly occur on the glass windows. The algorithm performs well in other parts of the images in spite of the large disparity range. Many other algorithms cannot cope with this due to the large resources consumed.

To demonstrate the usefulness of the adaptive window and the modified SSD, we also computed the depth map with a constant window and the traditional SSD for comparison, which is shown in Figure 8(b). We can see that the performance is obviously worse than our algorithm, especially in areas with less texture and sharp depth discontinuities such as near the roof of the building.

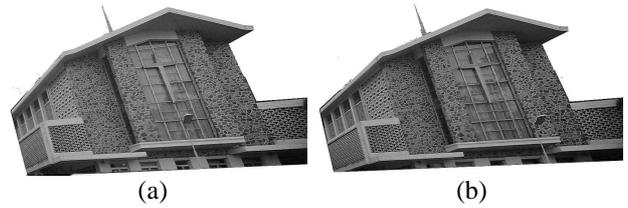


Figure 7: Rectified image pair of the church. (a) Left image. (b) Right image.

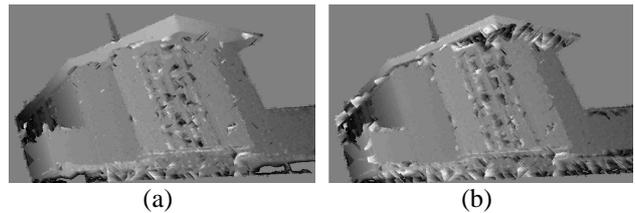


Figure 8: Depth map of the church obtained with (a) the new algorithm. (b) constant window and traditional SSD.

3D reconstruction has also been performed using dense matching results obtained from rectified images. The performance is approximately the same as images without rectification shown in Figure 6. In both cases, the percentage of the number of good points as a function of re-projection errors in pixels are shown in Figure 9 for comparison. It can be seen clearly that reconstruction with rectification has better performance.

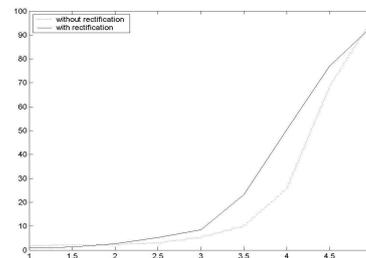


Figure 9: Percentage of the number of good points as a function of re-projection errors in pixels

To compare our algorithm with others, we also use the test data recommended in [10]. The University of Tsukuba “head and lamp” data set is selected. Figure 10 shows the reference image and the ground truth disparities respectively. The depth map obtained by our algorithm is shown in Figure 11. We can see that the depth map is smooth in most areas while the depth discontinuities around different objects are preserved at the same time. Depth maps obtained by other algorithms can be found in [10]. It should be noted that our dense matching algorithm is developed for 3D reconstruction from an image sequence captured by a hand-held camera. So its main advantage is that it handles images without rectification and makes good use of the correspondences of the feature points, which are available at the stage of dense matching. And it does well in our project of 3D reconstruction though it does not show better performance than some of the algorithms mentioned in [10] when dealing with stereo images selected from the testing data set.

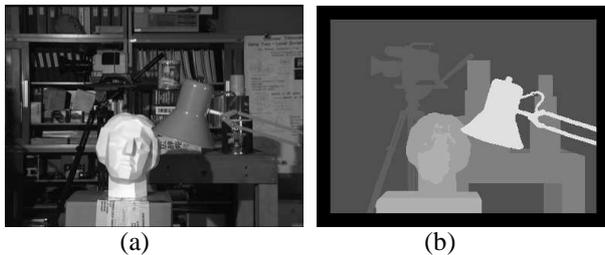


Figure 10: Testing image of the University of Tsukuba “head and lamp” data set. (a) Reference image. (b) Ground truth disparities.

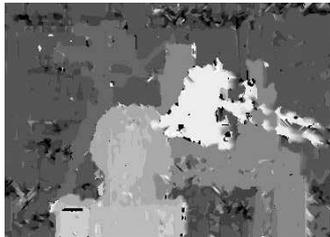


Figure 11: Depth map obtained by our algorithm

## 6. Conclusions

A novel dense matching algorithm based on propagation with Voronoi diagram is proposed in this paper. It is effective in eliminating large errors between repetitive patterns and in improving the efficiency. It works for a pair of images with or without rectification. In our experiments with real data, an overwhelming majority of the dense correspondences are correct. The results are good enough to allow for a robust 3D reconstruction. The algorithm has two steps. The first step uses Voronoi

diagram of the feature points to divide the image into a number of cells. The second step consists of dense matching by propagation from the matched seed feature point of each cell. A SSD in a window weighted by Gaussian coefficients is used as the matching score. Thus the weight in the centre is larger than those towards the edges of the window. All of these contribute to the improved performance of the proposed dense matching algorithm relative to most classical approaches.

## References

- [1] Bobick and S. Intille. Large Occlusion Stereo. *IJCV*, 33(3): 181-200, 1999.
- [2] M. Gong and Y. H. Yang. Multi-resolution Stereo Matching Using Genetic Algorithm. *IEEE CVPR. Workshop on Stereo and Multi-Baseline Vision*, 2001.
- [3] Kanade T. and M. Okutomi. A Stereo Matching Algorithm with an Adaptive Window: Theory and Experiment. *IEEE TPAMI*, Vol. 16, NO.9, 920-932, 1994.
- [4] M. Lhuillier. Efficient Dense Matching for Textured Scenes Using Region Growing. *INRIA Research Report*, RR n03382, 1998.
- [5] M. Lhuillier and L. Quan. Quasi-Dense Reconstruction from Image Sequence. *ECCV*. Vol. 2, 125-139, 2002.
- [6] G. Meerbergen et al. A hierarchical stereo algorithm using dynamic programming. *IEEE CVPR. Workshop on Stereo and Multi-Baseline Vision*, 2001.
- [7] Y. Ohta and Kanade T. Stereo by Intra- and Inter-Scanline Search Using Dynamic Programming. *IEEE TPAMI*, Vol.7, NO.2, 139-154, 1985.
- [8] M. Okutomi and Y. Katayama. A Simple Stereo Algorithm to Recover Precise Object Boundaries and Smooth Surfaces. *Proc. of IEEE CVPR. II-138*, 2001.
- [9] S. Roy and I. J. Cox. A Maximum-Flow Formulation of the N-camera Stereo Correspondence Problem. *Proc. of ICCV*, Bombay, 1998.
- [10] D. Scharstein and R. Szeliski (2002). A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms. *IJCV*, 47(1): 7-42; 2002.
- [11] J. Sun and H. Y. Shum and N. N. Zheng. Stereo Matching Using Belief Propagation. *ECCV*. Vol. 2, 510-524, 2002.
- [12] R. Szeliski and D. Scharstein. Symmetric Sub-pixel Stereo Matching. *ECCV*. Vol. 2, 525-540, 2002.
- [13] H. Tao, H. Sawhney, and R. Kumar. A global matching framework for stereo computation. *Proc. of ICCV*. Vol. I, 532-539, 2001.
- [14] Y. Zhang and C. Kambhamettu. Stereo Matching with Segmentation-Based Cooperation. *ECCV*. Vol. 2, 556-571, 2002.
- [15] L. Zitnick, and Kanade T. A Cooperative Algorithm for Stereo Matching and Occlusion Detection, *IEEE TPAMI*, Vol. 22, No. 7, 675 – 684, 2000.