MITSUBISHI ELECTRIC RESEARCH LABORATORIES
http://www.merl.com

Discontinuity Preserving Stereo with Small Baseline Multi-Flash Illumination

Rogerio Feris, Ramesh Raskar, Longbin Chen, Kar-Han Tan, Matthew Turk

TR2005-146 October 2005

Abstract

Currently, sharp discontinuities in depth and partial occlusions in multiview imaging systems pose serious challenges for many dense correspondence algorithms. However, it is important for 3D reconstruction methods to preserve depth edges as they correspond to important shampe features like silhouettes which are critical for understanding the structure of a scene. In this paper we show how active illumination algorithms can produce a rich set of feature maps that are useful in dense 3D reconstruction. We start by showing a method to compute a qualitative depth map from a single camera, which encodes object relative distances and can be used as a prior for stereo. In a multiview setup, we show that along with depth edges, binocular hal-occluded pixels can also be explicitly and reliably labeled. To demonstrate the usefulness of these feture maps, we show how they can be used in two different algorithms for dense stereo correspondence. Our experimental results how that our enhanced stereo algorithms are able to extract high quality, discontinuity preserving correspondence maps from scenes that are extremely challenging for conventional stereo methods.

IEEE International Conference on Computer Vision (ICCV)

Discontinuity Preserving Stereo with Small Baseline Multi-Flash Illumination

Rogerio Feris[1] Ramesh Raskar[2] Longbin Chen[1] Kar-Han Tan[3] Matthew Turk[1]

UC Santa Barbara[1]
{rferis,lbchen,mturk}@cs.ucsb.edu

Mitsubishi Electric (MERL)[2]
raskar@merl.com

Epson Palo Alto Lab[3]
tan@erd.epson.com

Abstract

Currently, sharp discontinuities in depth and partial occlusions in multiview imaging systems pose serious challenges for many dense correspondence algorithms. However, it is important for 3D reconstruction methods to preserve depth edges as they correspond to important shape features like silhouettes which are critical for understanding the structure of a scene. In this paper we show how active illumination algorithms can produce a rich set of feature maps that are useful in dense 3D reconstruction. We start by showing a method to compute a qualitative depth map from a single camera, which encodes object relative distances and can be used as a prior for stereo. In a multiview setup, we show that along with depth edges, binocular half-occluded pixels can also be explicitly and reliably labeled. To demonstrate the usefulness of these feature maps, we show how they can be used in two different algorithms for dense stereo correspondence. Our experimental results show that our enhanced stereo algorithms are able to extract high quality, discontinuity preserving correspondence maps from scenes that are extremely challenging for conventional stereo methods.

1. Introduction

The establishment of visual correspondence in stereo images is a fundamental operation that is the starting point of most geometric algorithms for 3D shape reconstruction. Intuitively, a complete solution to the correspondence problem would produce the following:

- A mapping between pixels in different images where there is a correspondence, and
- Labels for scene points that are not visible from all views – where there is *no correspondence*.

In the past two decades, intense interest in the correspondence problem has produced many excellent algorithms for solving the first half of the problem. With a few exceptions, most algorithms for dense correspondence do not address occlusions explicitly [15]. The occlusion problem is difficult partly because distinguishing image intensity variations caused by surface geometry from those caused by reflectance changes remains a fundamental unsolved vision problem [13].

A promising method for addressing the occlusion problem is to use active illumination [8, 14, 17]. In this paper we show how active lighting can be used to produce a rich set of feature maps that are useful in dense 3D reconstruction. Our method uses multi-flash imaging [14] in order to acquire important cues, including: (1) depth edges, (2) the sign of the depth edge (which tells the side of the foreground object), and (3) information about object relative distances.

Using these cues, we show how to produce rich feature maps for 3D reconstruction. We start by deriving a qualitative depth map from a single multi-flash camera. In a multiview setup, we show how binocular half-occluded pixels can be explicitly and reliably labeled, along with depth edges. We demonstrate how the feature maps can be used by incorporating them into two different dense stereo correspondence algorithms, the first based on local search and the second based on belief propagation.

1.1. Contributions

Our technical contributions include the following:

- A technique to compute a qualitative depth map which encodes information about the object relative distances (Section 2).
- A method for detection of binocular half-occlusions (Section 3).
- Algorithms for enhanced local and global depth edge preserving stereo (Sections 4 and 5).

1.2. Depth Edges with Multi-Flash

Before introducing our techniques, we briefly review the basic idea of detecting depth edges with multi-flash imaging [14].

The main observation is that when a flash illuminates a scene during image capture, thin slivers of cast shadow are created at depth discontinuities. Thus, if we can shoot a

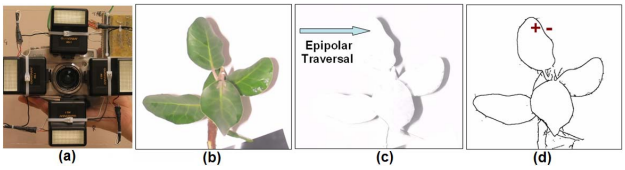


Figure 1: *(a) Multi-flash camera (b) Image taken with left flash. (c) Correspondent ratio image and traversal direction. (d) Computed depth edges. Note that we can obtain the sign of each depth edge pixel, indicating which side of the edge is the foreground.*

sequence of images in which different light sources illuminate the subject from various positions, we can use the shadows in each image to assemble a depth edge map using the shadow images.

Shadows are detected by first computing a *shadow-free image*, which can be approximated with the maximum composite image, created by choosing at each pixel the maximum intensity value among the image set. The shadow-free image is then compared with the individual shadowed images. In particular, for each shadowed image, a *ratio image* is computed by performing a pixel-wise division of the intensity of the shadowed image by the intensity of the maximum image.

The final step is to traverse each ratio image along its epipolar rays (as given by the respective light position) and mark negative transitions as depth edges. We use an implementation setup with four flashes at left, right, top and bottom positions, which makes the epipolar ray traversal aligned with horizontal and vertical scanlines. Figure 1 illustrates the main idea of the depth edge detection algorithm. Note that the sign of the edge is also obtained, indicating which part is the background and which part is the foreground in a local neighborhood.

2. Qualitative Depth Map

In this section, we use a single multi-flash camera to derive a qualitative depth map based on shadow width information. Our method is related to shape from shadow techniques [4], but differs significantly in methodology. At this point we are not interested in quantitative depth measurements. Rather, we want to segment the scene, while simultaneously establishing object depth-order relations and approximate relative distances. This turns out to be a valuable prior information for stereo.

2.1. Shadow Width Estimation

A natural way of extending our depth edge detection method to estimate shadow width is to measure the length of regions delimited by a negative transition (which corresponds to the depth edge) and a positive transition along the

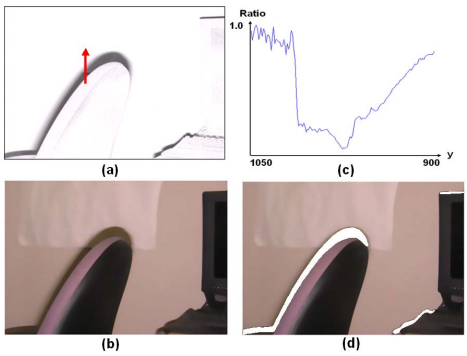


Figure 2: *(a) Ratio Image. (b) Original Image. (c) Intensity plot along the vertical scanline depicted in (a). Note that there is no sharp positive transition. (d) Meanshift segmentation to detect shadow, shown in white color.*

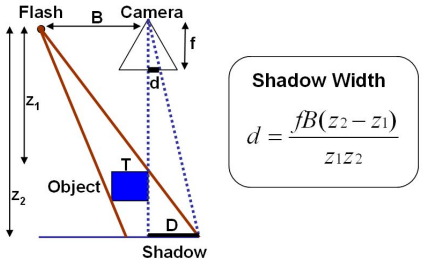


Figure 3: *Relationship of shadows and relative depth.*

epipolar ray in the ratio images. However, finding the positive transition is not an easy task, due to interreflections and the use of a non-point light source.

Figure 2a-c illustrates this problem: note that the intensity profile along the vertical scanline depicted in the ratio image has spurious transitions due to interreflections and a smooth transition near the end of the shadow. Estimation of the shadow width based on local-area-based edge filtering leads to unrealiable results. In contrast, we take advantage of the global shadow information. We apply the mean-shift segmentation algorithm [3] in the ratio image to segment the shadows, allowing accurate shadow width estimation (see Figure 2d).

2.2. Shadows and Relative Depth

We now look at the imaging geometry of the shadows, depicted in Figure 3, assuming a pinhole model. The variables involved are f (camera focal length), B (camera-flash baseline), z_1, z_2 (depths to the shadowing and shadowed edges), D (shadow width) and d (the shadow width in the image plane). For now, assume that the background is flat and whose distance z_2 from the camera is known. We have that $\frac{d}{f} = \frac{D}{z_2}$ and $\frac{D}{z_2 - z_1} = \frac{B}{z_1}$. It follows that the shadow width in the image can be computed as:

$$d = \frac{fB(z_2 - z_1)}{z_1 z_2} \qquad (1)$$

Working on this equation, we have:

$$\begin{aligned} \frac{dz_2}{fB} &= \frac{(z_2 - z_1)}{z_1} \\ \frac{dz_2}{fB} &= \frac{z_2}{z_1} - 1 \\ log(\frac{dz_2}{fB} + 1) &= log(\frac{z_2}{z_1} - 1 + 1) \end{aligned}$$

$$log(\frac{dz_2}{fB} + 1) = log(z_2) - log(z_1) \quad (2)$$

Note that for each depth edge pixel, we can compute the left hand side of equation 2, which encodes the relative object distances (difference of log depth magnitudes). This allows us to create a gradient field that encodes sharp depth changes(with gradient zero everywhere except at depth discontinuities) and perform 2D integration of this gradient field to obtain a qualitative depth map of the scene. This idea is described with more details below.

2.3. Gradient Domain Solution

In order to construct a sharp depth gradient map, we need to know the direction of the gradient at each depth edge pixel. This information can be easily obtained through the sign of the depth edge pixel in each orientation, which tells us which part of the edge is the foreground and which part is the background.

Le E be the set of depth edge pixels and $G = (G_h, G_v)$ the sharp depth gradient map, where G_h and G_v correspond to its horizontal and vertical components, respectively, with:

$$\begin{aligned} G_h(x,y) &= 0 \text{ if } (x,y) \notin E \quad (3) \\ &= log(\frac{d_h(x,y)z_2}{fB} + 1)s_h(x,y) \text{ otherwise} \end{aligned}$$

where $s_h(x,y)$ is the sign $(-1,+1)$ of the depth edge pixel (x,y) and $d_h(x,y)$ is the shadow width along the horizontal direction. The component G_v is calculated in the same way as equation 3 for the vertical direction.

Our qualitative depth map can be obtained with the following steps:

- Compute the sharp depth gradient $G(x,y)$.
- Integrate G by determining M which minimizes $|\nabla M - G|$.
- Compute the qualitative depth map $Q = exp(M)$.

It is important to note that the gradient vector field G may not be integrable. In order to determine the image M, we use a similar approach as the work of Fattal et al. [6]. The observation is that the optimization problem to minimize $|\nabla M - G|^2$ is equivalent to solving the Poisson differential equation $\nabla^2 M = div\ G$, involving a Laplace and a divergence operator. We solve this partial differential equation using the standard full multi-grid method, which involves discretization and the solution of a linear system in different grid levels. For specifying boundary conditions, we pad the images to square images of size the nearest power of two, and then crop the result image back to the original size. The final qualitative depth map is obtained by exponentiating M, since M contains the logarithm of the real depth values.

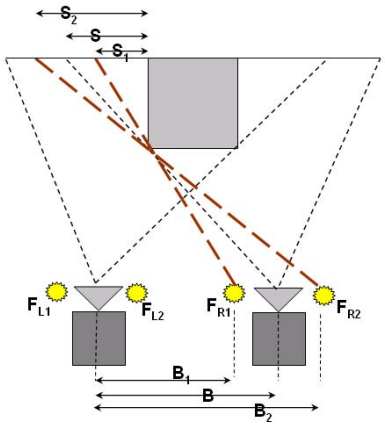


Figure 4: *The length of the half-occluded region is bounded by shadows created by flashes surrounding the other camera.*

For many applications, the background may be not flat and its distance to the camera unknown. In this case, we can set $\frac{z_2}{FB}$ to 1.0. In this case we cannot obtain the absolute distances from the background. Instead we get relative distances proportional to the shadow width and a qualitative depth map with segmented objects. We will show in Section 5 that this is a very useful prior for stereo matching.

3. Occlusion Detection

Binocular half-occlusion points are those that are visible in only one of the two views provided by a binocular imaging system [5]. They are a major source of error in stereo matching algorithms, due to the fact that half-occluded points have no correspondence in the other view, leading to false disparity estimation.

Current approaches to detect occlusion points are passive (see [5] for a comparison among five different techniques). They rely on the correspondence problem and thus are unable to produce accurate results for many real scenes. In general, these methods report a high rate of false positives and have problems to detect occlusions in areas of the scene dominated by low spatial frequency structure.

3.1. Occlusions Bounded by Shadows

Rather than relying on the hard correspondence problem, we exploit active lighting to detect binocular half-occlusions. Assume we have a stereo pair of cameras with horizontal parallax and light sources arranged as in Figure 4. By placing the light sources close to the center of projection of each camera, we can use the length of the shadows created by the lights surrounding the other camera to bound the half-occluded regions.

This idea is illustrated in Figure 4. Note that the half-occluded region S is bounded by the width of the shadows S_1 and S_2. Observing the figure, let I_{L_1}, I_{R_1} and I_{R_2} be the images taken by the left camera with light sources F_{L_1}, F_{R_1} and F_{R_2}, respectively. The width of S_1 and S_2 can be determined by applying the meanshift segmentation algorithm in the ratio images $\frac{I_{R_1}}{I_{L_1}}$ and $\frac{I_{R_2}}{I_{L_1}}$ (as described in Section 2.1). We then determine the half-occluded region by averaging the shadowed regions: $S = \frac{B}{B_1+B_2}(S_1 + S_2)$, where B, B_1, and B_2 are the baselines of the camera and each light source, as shown in the figure.

The occluded region is determined with precision for planar shadowed region and with close approximation for non-planar shadowed region. In the non-planar case, the linear relationship between baseline and shadow width does not hold, but the length of the occluded region is guaranteed to be bounded by the shadows.

We could also use Helmholtz stereopsis [17] by exchanging the position of a multi-flash camera with a light source. The shadowed region caused by the light source in this configuration would denote exactly the half-occluded region. However, the device swapping needs precise calibration and would be difficult to implement as a self-contained device.

4. Enhanced Local Stereo

In this section, we enhance window-based stereo matching using automatically detected depth edges and occlusions. Our method requires very few computations and shows great improvement over traditional correlation-based methods.

A major challenge in local stereo is to produce accurate results near depth discontinuities. In such regions, the main assumption of local methods is violated: the same window (aggregation support) contains pixels that significantly differ in disparity, often causing serious errors in the matching process, due to perspective distortions. In addition, windows that include half-occluded points near depth discontinuities are another source of error, since they do not have correspondence in the other view.

The central problem of local methods is to determine the optimal size, shape, and weight distribution of the aggregation support for each pixel. There is a trade-off in choosing the window size: if the window is too small, a wrong match might be found due to ambiguities and noise. If the window is too large, problems due to foreshortening and depth discontinuities occur, with the result of lost detail and blurring of object boundaries. Previous solutions to this problem include the use of adaptive windows [10] and shiftable windows [11], but producing clean results around depth discontinuities still remains a challenge.

4.1. Varying Window Size and Shape

We adopt a sliding window which varies in shape and size, according to depth edges and occlusion, to perform local correlation. Given the quality of the detection of depth edges and half-occluded points, results are significantly improved.

In order to determine the size and shape of the window for each pixel, we determine the set of pixels that has aproximatelly the same disparity as the center pixel of the window. This is achieved by a region growing algorithm (starting at the center pixel) which uses depth edges and half-occluded points as boundaries.

Only this set of pixels is then used for matching in the other view. The other pixels in the window are disconsidered, since they correspond to a different disparity.

5. Enhanced Global Stereo

The best results achieved in stereo matching thus far are given by global stereo methods, particularly those based on belief propagation and graph cuts [12, 16]. These methods formulate the stereo matching problem as a maximum a posteriori Markov Random Field (MRF) problem. In this section, we will describe our enhanced global stereo method, which uses belief propagation for inference in the Markov network.

Some current approaches explicitly model occlusions and discontinuities in the disparity computation [1, 9], but they rely on intensity edges and junctions as cues for depth discontinuities. This poses a problem in low-contrast scenes and in images where object boundaries appear blurred. However, we want to suppress smoothness constraints only at occluding edges, not at texture or illumination edges. Our method makes use of the prior information to circumvent these problems, including the qualitative depth map and the automatically detected binocular half-occlusions described earlier.

5.1. Inference by Belief Propagation

The stereo matching problem can be formulated as a MRF with hidden variables $\{x_s\}$, corresponding to the disparity of each pixel, and observed variables $\{y_s\}$, corresponding to the matching cost (often based on intensity differences) at specific disparities. By denoting $X = \{x_s\}$ and $Y = \{y_s\}$,

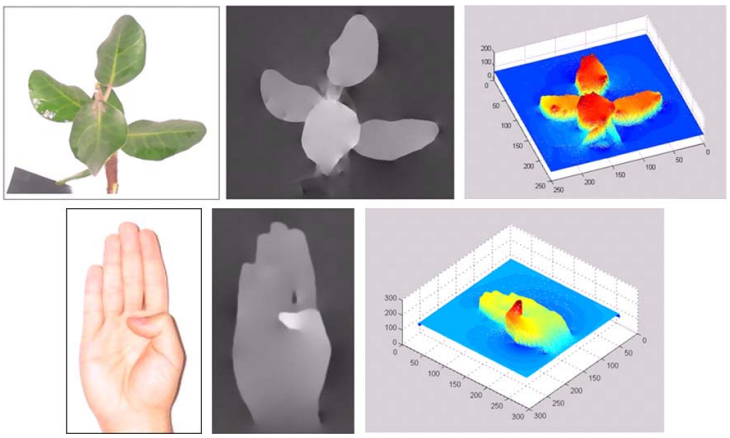


Figure 5: *From left to right: original image, qualitative depth map and the corresponding 3D plot. Note that our method captures small changes in depth and is robust in the presence of low intensity variations across depth contours.*

the posterior $P(X|Y)$ can be factorized as:

$$P(X|Y) \propto \prod_s \psi_s(x_s, y_s) \prod_s \prod_{t \in N(s)} \psi_{st}(x_s, x_t) \qquad (4)$$

where $N(s)$ represents a neighborhood of s, ψ_{st} is called the compatibility matrix between nodes x_s and x_t (smoothness term), and $\psi_s(x_s, y_s)$ is called the local evidence for node x_s, which is the observation probability $p(y_s|x_s)$ (data term). The belief propagation algorithm gives an efficient approximate solution in this Markov network [16].

5.2. Qualitative Depth as Evidence

We can potentially use our computed depth edges to suppress smoothness constraints during optimization. However, the depth contours may have gaps. Fortunately, our qualitative depth image shows a desirable slope in intensity when gaps occur (as we will show in our experiments), and hence it is a good choice to set the compatibility matrix ψ_{st}. In addition, the qualitative depth map encodes the object relative distances via the shadow width information, and we use the map to encourage discontinuities at a certain disparity difference.

Let P be the qualitative depth scaled to match the set of possible disparities d_i, $i = 1..L$. We define $\psi_{st}(x_s, x_t) = C^{st}_{LxL}$, where C^{st}_{ij} is defined as:

$$C^{st}_{ij} = exp(-\frac{|d_i - d_j - \Delta P_{st}|}{F}) \qquad (5)$$

where ΔP_{st} is the intensity difference between pixels s and t in the qualitative map (which was scaled to match possible disparities) and F is a constant scaling factor. Intuitively, if $\Delta P_{st} = 0$, there is no sharp discontinuity for neighboring pixels s and t and the compatibility matrix will have larger values along its diagonal, encouraging neighboring pixels to have the same disparity. In contrast, if $\Delta P_{st} \neq 0$, the larger values will be shifted to the disparity encoded by ΔP_{st}. The direction of this shift depends on the sign of ΔP_{st}, which is the sign of the correspondent depth edge.

We have also included the half-occlusion information in our method. Nodes correspondent to pixels that have no match in the other view are eliminated, while a penalty is given for matching a given pixel with an occluded point in the other view.

5.3. Signed Edge Matching

We also consider depth edges as part of the matching cost computation. This is very useful in low-contrast scenes, where occluding boundaries may not correspond to intensity edges. Using signed depth edges to improve matching is significantly more reliable than using intensity edges. Our approach could be also used in techniques based on dynamic programming, where the matched edges would correspond to a priori ground control points.

6. Experiments

In this section we describe our experiments, showing results for the computation of a qualitative depth map, detection of binocular half-occlusions, and enhanced local and global stereo algorithms.

Qualitative Depth Map. Figure 5 illustrates results obtained for the qualitative depth map computation from a single camera. We assume we do not know the camera-background distance, since our interest is to use this map as prior for stereo. As we can see, our method effectively segments the scene, encoding object relative distances through the shadow width information. Note that the images have low intensity variation and small depth changes, a challenging scenario for most 3D reconstruction methods.

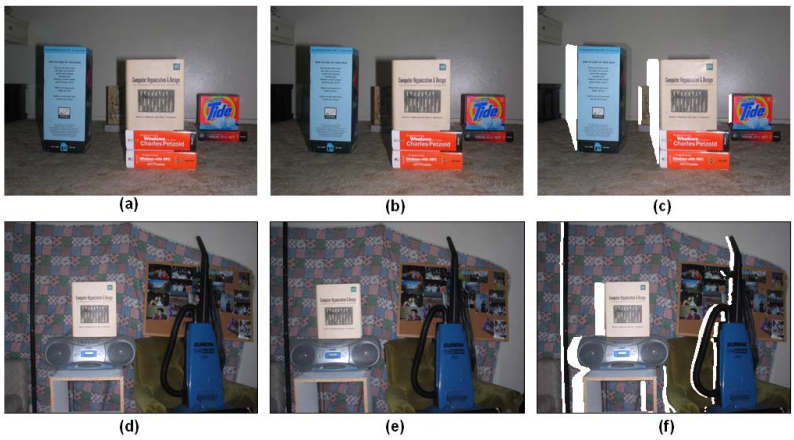


Figure 6: *Detection of binocular half-occlusions in both textured and textureless regions. (a)-(b) Images taken with light sources surrounding the other camera. (c) Our occlusion detection result marked as white pixels.* 0.65% *of false positives and* 0.12% *of false negatives were reported. (d) Left view. (e) Right view. (f) Occlusion detection (white pixels).*

Our qualitative depth map also offers the advantage of creating a slope in intensity when there are gaps in the depth contours. Note in the hand image the smooth transition between the thumb finger and the palm of the hand. This is a useful property for setting smoothness constraints in stereo matching.

Clearly, our method is not able to handle slanted surfaces or rounded objects, since the depth variation is smooth without a sharp discontinuity. This is not a problem if we use it as a prior for stereo reconstruction.

Occlusion Detection. We used two Canon G3 cameras with light sources arranged as Figure 4 to test our half-occlusion detection algorithm. Figure 6 demonstrates the reliable performance of our method. The images contain occlusion points in both textured and textureless regions, which is a challenging problem for passive algorithms that rely on pixel correspondence. For quantitative evaluation, we selected a piecewise planar scene (Figure 6a-c), since it is easier to obtain the occlusion ground truth (computed from the known disparity map). For this scene, our method reports 0.65% of false positives and 0.12% of false negatives. For very large depth differences our method may not give a precise estimation (for non-planar shadowed regions, due to larger bounded regions) and it might fail due to detached shadows with thin objects.

Stereo Matching. We used a horizontal slide bar for acquiring stereo images with a multi-flash camera. Occlusions were estimated by moving the flashes properly to the shooting camera positions.

Figure 7a shows one of the views of a difficult scene we used as input. The image contains textureless regions, ambiguous patterns (e.g., the background close to the book), a geometrically complex object and thin structures. The resolution of the images is 640x480. We rectified them so that epipolar lines are aligned with horizontal scanlines. We adopted a small baseline between the cameras (maximum disparity equals 10), so that we can obtain a hand-labeled disparity ground truth (Figure 7b).

Figure 7c shows our computed depth edges and half-occluded points. Note that some edges do not appear in the ground truth (due to range resolution) and we also have some gaps in the edges due to noise. This data was considered to test our algorithms under noisy conditions.

Traditional local-correlation approaches perform very poorly in this scene, as we show in Figures 7d and 7e, using windows of size 9x9 and 31x31. In addition to noise, there are major problems at depth discontinuities - corners tend to become rounded and thin structures often disappear or expand. In contrast, our method preserve discontinuities with large windows (Figure 7f). We show a quantitative analysis of the two methods with respect to the window size in Figure 7g. The axis of the graph correspond to the root-mean-squared error (RMS) and the window size in pixels. The error decreases significantly as the window grows for our method (solid line). At some point, it will start growing again with larger windows due to gaps in the depth edges. We could use our qualitative depth map here, but this would add an undesirable computational load, since local-based approaches are attractive because of their efficiency.

Global Stereo Matching. We use the qualitative depth map as prior for belief propagation stereo matching. The computed map is shown in Figure 8a. The results for the standard belief propagation algorithm and our enhanced method are shown in Figures 8b and 8c, respectively. The passive method fails to preserve discontinuities due to matching ambiguities (we used the implementation available at http://cat.middlebury.edu/stereo/ with different weight and penalty parameters). Our results clearly show significant improvements with a RMS of 0.4590 compared to 0.9589 for this input. It is important to note that (although

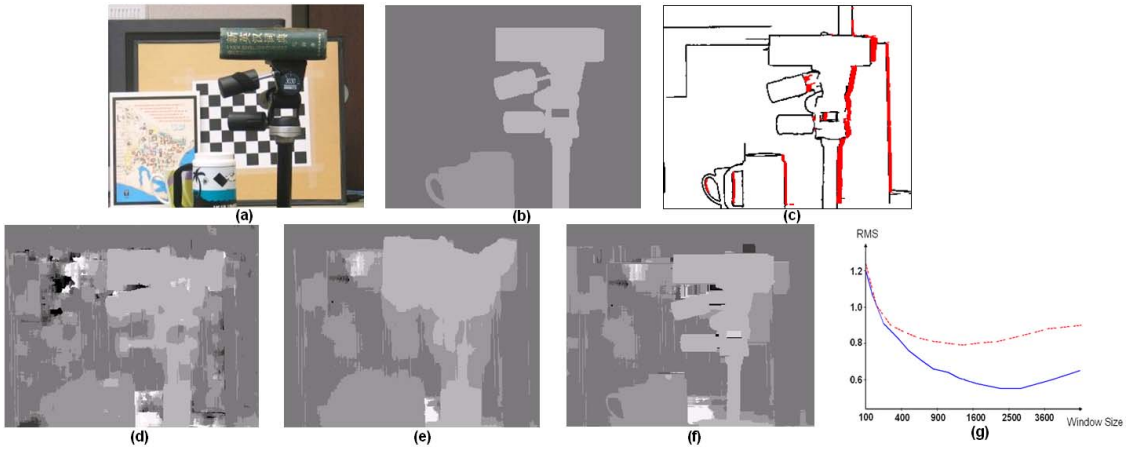


Figure 7: Enhanced Local Stereo *(a) Original image. (b) Hand-labeled ground truth. (c) Detection of depth edges and binocular half-occlusions. (d) Local correlation result with a 9x9 window. (e) Local correlation result with a 31x31 window. (f) Our multi-flash local stereo result with a 31x31 window. (g) Analysis of the root-mean-squared error with respect to window wize. The dashed line corresponds to traditional local correlation, while the solid line corresponds to our approach.*

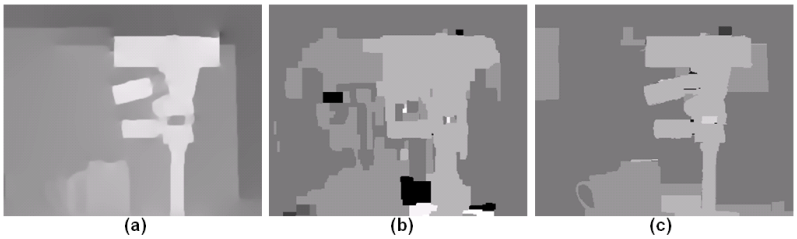


Figure 8: Enhanced Global Stereo *(a) Qualitative depth map. (b) Standard passive belief propagation result (RMS: 0.9589). (c) Our enhanced global stereo method (RMS: 0.4590).*

we do not show in this scene) our method handles slanted surfaces exact in the same way as standard global methods. In other words, we do not sacrifice slanted surfaces to preserve discontinuities as opposed to [2].

Figure 10 illustrates a simple example to show the importance of signed edge matching in disparity computation. The scene is challenging because the objects have the same color and occlude each other. Thus, the assumption that depth discontinuities are associated with intensity edges is not valid. In this case the most prominent features to use are the detected depth edges. We match signed edges in the two views and use belief propagation to propagate information according to our qualitative depth map, leading to the result shown in Figure 10b. For larger baseline scenarios, problems may occur with view-dependent edges (which are depth discontinuities from one view but normal discontinuities from the other).

Efficiency. Our qualitative depth map takes about two seconds to compute on a Pentium IV 1.8 GHz for 640x480 resolution images. Our enhanced local-based stereo algorithm requires very few computations since depth edges can be computed extremely fast [14]. Our enhanced global method computation time is the sum of the time for the qualitative depth map computation plus the time for belief propagation procedure. We refer to [7] for an efficient implementation of the belief propagation algorithm.

Comparison with other techniques. Figure 9 shows a comparison of our multi-flash stereopsis approach with other stereo methods. Note that small baseline flash setup means we do not need a laboratory setup as in photometric stereo and the cost and complexity of a flash attachment is very low. In addition, for non-intrusive applications, we can use readily available infra-red flash lighting but projecting high frequency structured patterns requires an infra-red projector.

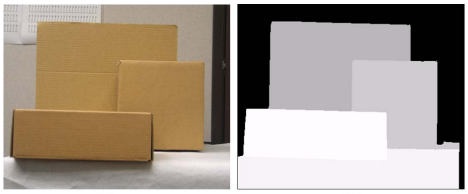
Remarks. All the experiments reported above were carried out on indoor, static scenes. A method for detecting depth edges in dynamic scenes was demonstrated in [14]. This requires high frame rates, but we are currently working on using light sources with different wavelength (triggered all in the same time) to tackle this problem.

7. Conclusions

We have presented a set of techniques based on active lighting for reliable, discontinuity preserving stereo matching.

	Recovered Information	Active / Passive	Handles Constant Albedo	Handles Depth Discontinuities	Compact, self-contained	Hardware Complexity
Structured Light	Depth	Active	Yes	Yes	No	High
Photometric Stereopsis	Normals	Active	Yes	Limited	No	Low
Multi-Flash Stereopsis	Depth	Active	Yes	Yes	Yes	Low
Passive Stereo	Depth	Passive	Limited	Limited	Yes	Low

Figure 9: *Comparison of our technique with other 3D reconstruction approaches.*



(a) (b)

Figure 10: *Usefulness of signed edge matching in low contrast scenes. (a) Left view. (b) Disparity map obtained by using belief propagation with matching costs including signed edge matching. This allows us to handle low-contrast scenes, where depth discontinuities may not correspond to intensity edges .*

Our methods include the derivation of a qualitative depth map from one single camera, detection of binocular half-occlusions, and enhanced local and global stereo algorithms based on these features.

Our techniques are reliable, simple, and inexpensive - the overall setup can be built into a self-contained device, no larger than existing 3D cameras. In the future, we plan to address the problem of specularities in stereo using the same framework and handle dynamic scenes.

References

[1] M. Agrawal and L. Davis. Window-based, discontinuity preserving stereo. In *Conference on Computer Vision and Pattern Recognition*, Washington, DC, 2004.

[2] S. Birchfield and C. Tomasi. Depth discontinuities by pixel-to-pixel stereo. *International Journal of Computer Vision*, 35(3):269–293, 1999.

[3] C. Christoudias, B. Georgescu, and P. Meer. Synergism in low level vision. In *International Conference on Pattern Recognition*, Quebec City, Canada, 2002.

[4] M. Daum and G. Dudek. On 3-D Surface Reconstruction using Shape from Shadows. In *International Conference on Computer Vision and Pattern Recognition (CVPR'98)*, pages 461–468, June 1998.

[5] G. Egnal and R. Wildes. Detecting binocular half-occlusions: Empirical comparisons of five approaches. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(8):1127–1133, 2002.

[6] R. Fattal, D. Lischinski, and M. Werman. Gradient Domain High Dynamic Range Compression. In *Proceedings of SIGGRAPH 2002*, pages 249–256. ACM SIGGRAPH, 2002.

[7] P. Felzenszwalb and D. Huttenlocher. Efficient Belief Propagation for Early Vision. In *International Conference on Computer Vision and Pattern Recognition (CVPR'04)*, 2004.

[8] P. Huggins, H. Chen, P. Belhumeur, and S. Zucker. Finding Folds: On the Appearance and Identification of Occlusion . In *Conference on Computer Vision and Pattern Recognition*, volume 2, pages 718–725, December 2001.

[9] H. Ishikawa and D. Geiger. Occlusions, Discontinuities, and Epipolar Lines in Stereo. In *European Conference on Computer Vision (ECCV'98)*, 1998.

[10] T. Kanade and M. Okutomi. A stereo matching algorithm with an adaptive window: Theory and experiment. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(9):920–932, 1994.

[11] S. Kang, R. Szeliski, and J. Chai. Handling occlusions in dense multi-view stereo. In *International Conference on Computer Vision and Pattern Recognition (CVPR'01)*, volume 1, pages 102–110, 2001.

[12] V. Kolmogorov and R. Zabih. Computing visual correspondence with occlusions using graph cuts. In *International Conference on Computer Vision*, Vancouver, Canada, 2001.

[13] M.Bell and W. Freeman. Learning Local Evidence for Shading and Reflectance. In *International Conference on Computer Vision (ICCV'01)*, volume 1, pages 670–677, 2001.

[14] R. Raskar, K. Tan, R. Feris, J. Yu, and M. Turk. A non-photorealistic camera: depth edge detection and stylized rendering using multi-flash imaging. *SIGGRAPH'04 / ACM Transactions on Graphics*, 2004.

[15] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. In *International Journal of Computer Vision*, volume 47(1), pages 7–42, 2002.

[16] J. Sun, N. Zheng, and H. Shum. Stereo matching using belief propagation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(07):787–800, 2003.

[17] T. Zickler, P. N. Belhumeur, and Kriegman D. J. Helmholtz Stereopsis: Exploiting Reciprocity for Surface Reconstruction . In *European Conference on Computer Vision (ECCV'02)*, 2002.