

# **A Volumetric Iterative Approach to Stereo Matching and Occlusion Detection**

C. Lawrence Zitnick, Takeo Kanade

CMU-RI-TR-98-30

The Robotics Institute  
Carnegie Mellon University  
Pittsburgh, Pennsylvania 15213

December 1998

© 1998 Carnegie Mellon University

This research has been supported in part by the Advanced Research Projects Agency under the Department of the Army, Army Research Office grant number DAAH04-94-G-0006. Views and conclusions contained in the document are those of the authors and should not be interpreted as necessarily representing official policies or endorsements, either expressed or implied, of the Department of the Army or the United States Government.

Keywords: Stereo vision, occlusion detection, volumetric, 3-D vision.

## **Abstract**

This paper presents a stereo algorithm for obtaining disparity maps with explicitly detected occlusion. To produce smooth and detailed disparity maps, two assumptions that were originally proposed by Marr and Poggio are adopted: uniqueness and continuity. That is, the disparity maps have unique values and are continuous almost everywhere. A volumetric approach is taken to utilize these assumptions. A 3D array of match likelihood values is constructed with each value corresponding to a pixel in an image and a disparity relative to another image. An iterative algorithm updates the match likelihood values by diffusing support among neighboring values and inhibiting others. After the values have converged, the region of occlusion is explicitly detected.

To demonstrate the effectiveness of the algorithm we present the processing results from synthetic and real image pairs, with comparison to results by other methods. The resulting disparity maps are smooth and detailed with occlusions detected. Disparity values in areas of repetitive texture are also found correctly.



## 1. Introduction

Stereo vision can produce a dense disparity map. The resultant disparity map should be smooth and detailed; continuous and even surfaces should produce a region of smooth disparity values with their boundary precisely delineated, while small surface elements should be detected as separately distinguishable regions. Though obviously desirable, it is not easy for a stereo algorithm to satisfy these two requirements at the same time. Algorithms that can produce a smooth disparity map tend to miss the details and those that can produce a detailed map tend to be noisy.

For area-based stereo methods [8], [12], [22], [5], [2], [7], which match neighboring pixel values within a window between images, the selection of an appropriate window size is critical to achieving a smooth and detailed disparity map. The optimal choice of window size depends on the local amount of variation in texture and disparity [14], [2], [4], [15], [7]. In general, a smaller window is desirable to avoid unwanted smoothing. In areas of low texture, however, a larger window is needed so that the window contains intensity variation enough to achieve reliable matching. On the other hand, when the disparity varies within the window (i.e., the corresponding surface is not fronto-parallel), intensity values within the window may not correspond due to projective distortion. In addition to unwanted smoothing in the resultant disparity map, this fact creates the phenomena of so-called fattening and shrinkage of a surface. That is, a surface with high intensity variation extends into neighboring surfaces with less variation across occluding boundaries.

Many attempts have been made to remedy this serious problem in window-based stereo methods. One earlier method is to warp the window according to the estimated orientation of the surface to reduce the effect of projective distortion [17]. A more recent and sophisticated method is an adaptive window method [7]. The window size and shape are iteratively changed based on the local variation of the intensity and current depth estimates. While these methods showed improved results, the first method does not deal with the difficulty at the occluding boundary, and the second method is extremely computationally expensive. A typical method to deal with occlusion is bi-directional matching. For example, in the paper by Fua [4] two disparity maps are created relative to each image: one for left to right and another for right to left. Matches which are consistent between the two disparity maps are kept. Inconsistent matches create holes, which are filled in using interpolation.

The fundamental problem of these stereo methods is that they make decisions very locally; they do not take into account the fact that a match at one point constrains others due to the stereo and scene geometry. One constraint commonly used by feature-based stereo methods is edge consistency [16], [9]. That is, all matches along a continuous edge must be consistent. While constraining matches using edge consistency improves upon local feature-based methods [1], [19], they produce only sparse depth maps. The work by Marr and Poggio [10] is one of the first to apply global constraints or assumptions while producing a dense depth map. Two assumptions about stereo were stated explicitly: uniqueness and continuity of a disparity map. That is, the disparity maps have unique values and are continuous almost everywhere. They devised a simple cooperative algorithm for diffusing support among disparity estimates to take advantage of the two assumptions. They demonstrated the algorithm on synthetic random-dot images. Due to memory and processing constraints, similar methods using real images have been left largely unexplored. Recently, Scharstein and Szelski [20] proposed a Bayesian model of stereo matching. In creating continuity within the disparity map, support among disparity estimates is non-linearly diffused. The derived method has results similar to that of adaptive window methods.

In this paper we present an volumetric iterative stereo algorithm using global constraints to find a dense depth map. The uniqueness and continuity assumptions by Marr and Poggio are adopted. A 3D set of match likelihood values is constructed with each value corresponding to a pixel in the reference image and a disparity relative to another image. An iterative update function generates continuous and unique values by diffusing support among neighboring match likelihood values and by inhibiting values along similar lines of sight. Initial match likelihood values, possibly obtained by pixel-wise correlation, are used to retain details during each iteration. After the match likelihood values have converged, occluded areas are explicitly detected.

To demonstrate the effectiveness of the algorithm we provide data from several synthetic and real scenes. The resulting disparity maps are smooth and detailed with occlusions detected. Also, disparity values in areas of repetitive texture are found correctly. A comparison with multi-baseline and multi-baseline with adaptive window methods is made.

## 2. A Volumetric Iterative Stereo Algorithm

Marr and Poggio [10] presented two basic assumptions for a stereo vision algorithm. The first assumption states that at most a single unique match exists for each pixel; that is, each pixel corresponds to a single surface point. When using intensity values for matching this uniqueness assumption may be violated if surfaces are not opaque. A classic example is a pixel receiving contribution from both a fish and a fish bowl. The second assumption states that disparity values are generally continuous, i.e. smooth within a local neighborhood. In most scenes surfaces are relatively smooth and discontinuities occur only at object boundaries.

We propose a volumetric [10], [3], [11], [20], [21] approach to utilize these two assumptions. Given a pair of stereo images, we construct a 3D voxel space with dimensions row, column and disparity. Assuming (without loss of generality) that the images have been rectified, each voxel  $(r, c, d)$  projects to the pixel  $(r, c)$  in the left image and to the pixel  $(r, c+d)$  in the right image. Within each voxel, the estimated likelihood value of a match between the pixels is held.

The uniqueness assumption implies there can only exist one match within a set of voxels which project to the same pixel in an image. As illustrated in figure 1 by dark voxels, let  $P(r, c, d)$  denote the set of voxels which overlap voxel  $(r, c, d)$  when projected to an image. That is, each voxel in  $P(r, c, d)$  projects to pixel  $(r, c)$  in the left image or to pixel  $(r, c+d)$  in the right image.

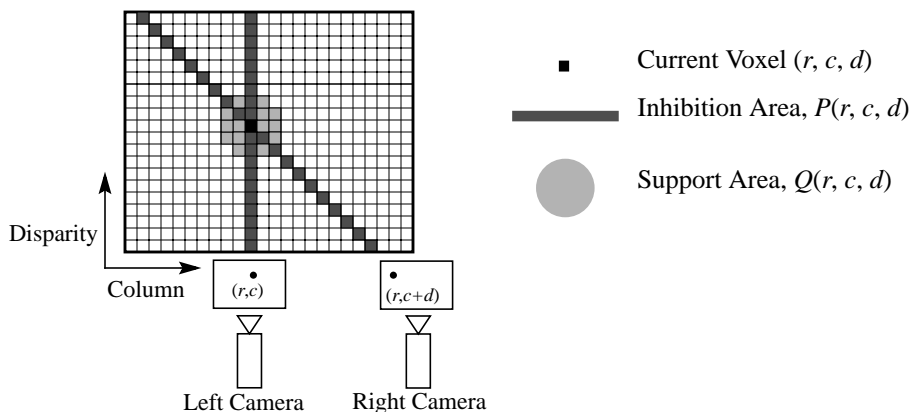


Figure 1; Illustration of the inhibitory and averaging relations between voxels for a 2D slice of the 3D voxel space with the row number held constant.

The continuity assumption implies neighboring voxels have consistent match likelihood values. Marr and Poggio proposed iteratively averaging their values to increase consistency. When averaging neighboring match likelihood values we need a concept of local support. The local support for a voxel determines which and to what extent neighboring voxels should contribute to averaging. Ideally, the local support should include all and only those neighboring voxels which correspond to a correct match if the current voxel corresponds to a correct match. Since the correct match is not known beforehand, some assumption is required on deciding the extent of the local support. Marr and Poggio, for example, used voxels having equal disparity values for averaging - that is, their local support area spans a 2D area ( $d=const.$ ) in the  $r$ - $c$ - $d$  space. This 2D local support corresponds to the parallel plane assumption. However, sloping and more general surfaces require using a 3D area for local support. Many 3D local support assumptions have been proposed [6], [18], [19], [7].

Let  $L_n(r, c, d)$  denote the match likelihood value assigned to voxel  $(r, c, d)$  at iteration  $n$ , and  $S_n(r, c, d)$  to be the sum of all match likelihood values within a 3D local support  $Q$ .

$$S_n(r, c, d) = \sum_{(r', c', d') \in Q(r, c, d)} L_n(r + r', c + c', d + d')$$

To obtain a smooth and detailed disparity map an iterative function is used to refine the match likelihood values. The initial values  $L_0(r, c, d)$  can be computed from images for example, by using intensity correlation. The update function should enforce the uniqueness assumption within the inhibition set  $P(r, c, d)$  and increase continuity using  $S_n(r, c, d)$ . We propose that the update function be proportional to the following:

$$L_{n+1}(r, c, d) \approx \left( \frac{S_n(r, c, d)}{\sqrt{\sum_{(r', c', d') \in P(r, c, d)} S_n(r', c', d')^2}} \right)^\alpha \quad (2)$$

Note the following properties of this function: Neighboring match likelihood values are averaged to create consistency. Inhibition, resulting from normalizing the sums over the inhibition area  $P(r, c, d)$ , creates a unique high valued voxel within the set. Normalization also restricts the range of  $L_n(r, c, d)$  to lie between 0 and 1. The rate of convergence is controlled by  $\alpha$  ( $\alpha > 1$ ).

However, averaging the match likelihood values in each iteration may result in over-smoothing and thus a loss of details. To maintain details we propose restricting the match likelihood values by their initial values; in this way, since the initial match likelihood values are based on intensity correlation, only the voxels which project to pixels with similar intensities can have high values. Therefore, we may multiply (2) by the initial values  $L_0(r, c, d)$ .

$$L_{n+1}(r, c, d) = L_0(r, c, d) \left( \frac{S_n(r, c, d)}{\sqrt{\sum_{(r', c', d') \in P(r, c, d)} S_n(r', c', d')^2}} \right)^\alpha \quad (3)$$

Using (3) we can achieve our goal of producing a smooth and detailed depth map.

While our method is similar to Marr and Poggio's, it does differ for two important points.

The first is we use a 3D local support when averaging voxels. Real images have sloping disparities so the local support must cover a 3D area. Secondly, Marr and Poggio used discrete values for their match likelihood values. Our method uses continuous values, allowing for interpolation between disparity estimates.

We will now discuss two additional properties that equation (3) provides us: one is explicit identification of the areas of occlusion within the images, and the other is elimination of mismatches due to repetitive textures.

### 3. Explicit Detection of Occluded Areas

Occlusion is a critical and difficult phenomena for stereo algorithms [13]. With any reasonably complex scene there exists occluded pixels with no correct match. Unfortunately most stereo algorithms do not consider this important case explicitly, and therefore they produce gross errors in areas of occlusion or find disparity values similar to the foreground or background. Our method, in contrast, can explicitly detect occlusion by examining the magnitude of the converged match likelihood values.

Since no correct match exists in areas of occlusion, all match likelihood values corresponding to occluded pixels should be small. Consider a pixel  $p$  in the left image, whose correct corresponding point is not visible in the right image. For a voxel  $v$  along the line of sight of  $p$ , there are two cases that occur for its projection  $q$  on the right image. The first case, depicted in figure 2(a), is when  $q$ 's correct corresponding point is visible in the left image. Then there exists a voxel  $v'$  which corresponds to the correct match between a pixel  $p'$  in the left image and  $q$ . Since voxels  $v$  and  $v'$  both project to pixel  $q$ , their match likelihood values will inhibit each other due to the uniqueness assumption. Generally, the correct voxel  $v'$  will have a higher match likelihood value, causing the value for voxel  $v$  to decrease. The second case, depicted in figure 2(b), is a more difficult case. This occurs when  $q$ 's true corresponding voxel is occluded in the left image. Since neither  $p$  nor  $q$  has a correct match, the likelihood value of a match between  $p$  and  $q$  will receive no inhibition from voxels corresponding to correct matches. Thus false matches could have high likelihood values if occluded areas between images happen to have similar intensities.



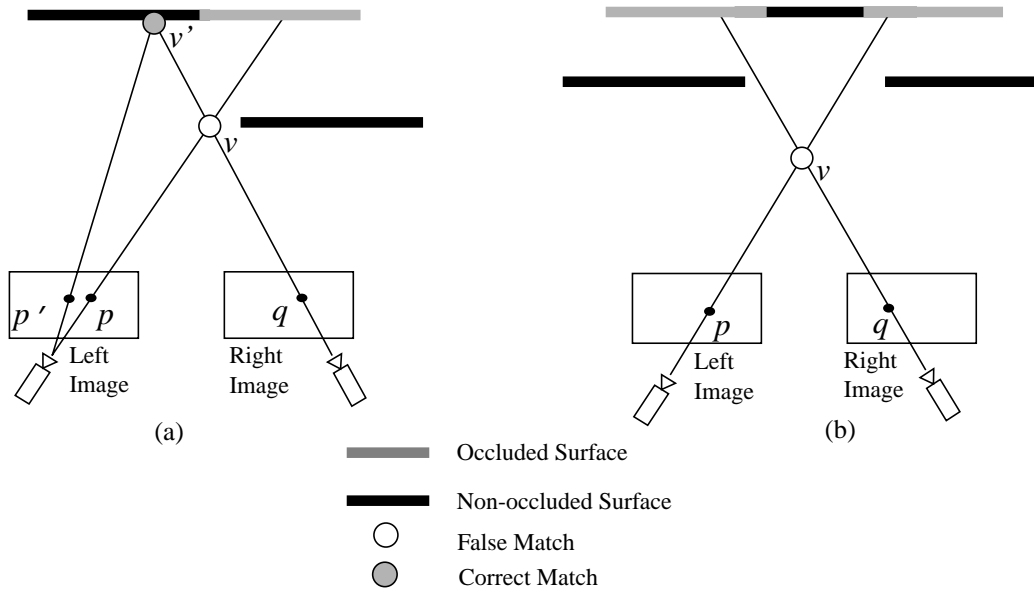


Figure 2;(a) If  $q$  is not occluded there is a correct match with  $p'$  which will inhibit the false match with  $p$ ; (b) If  $q$  is occluded it is possible for a false match to occur with  $p$ .

Provided occluded areas between images do not have similar intensities, all match likelihood values corresponding to occluded pixels will be small. Therefore, after the match likelihood values have converged, we can determine if a pixel is an occluded pixel by summing the values of its voxels along its line of sight. If the sum of the match likelihood values is below a threshold the pixel is labeled as occluded.

To illustrate this method of detecting occlusion we have created a pair of random dot stereo images shown in figure 3(a). The correct disparity map is shown in figure 3(b) with occluded pixels to the right of the upper block and to the left of the lower block. At iteration  $n=0$ , the occluded areas have smaller likelihood values since no correct matches exist. Figure 3(c) shows as the iterations proceed the match likelihood values of the occluded areas decrease due to inhibition from the correct matches. After four iterations the match likelihood values have converged and the occluded areas are detected as shown in figure 3(d).

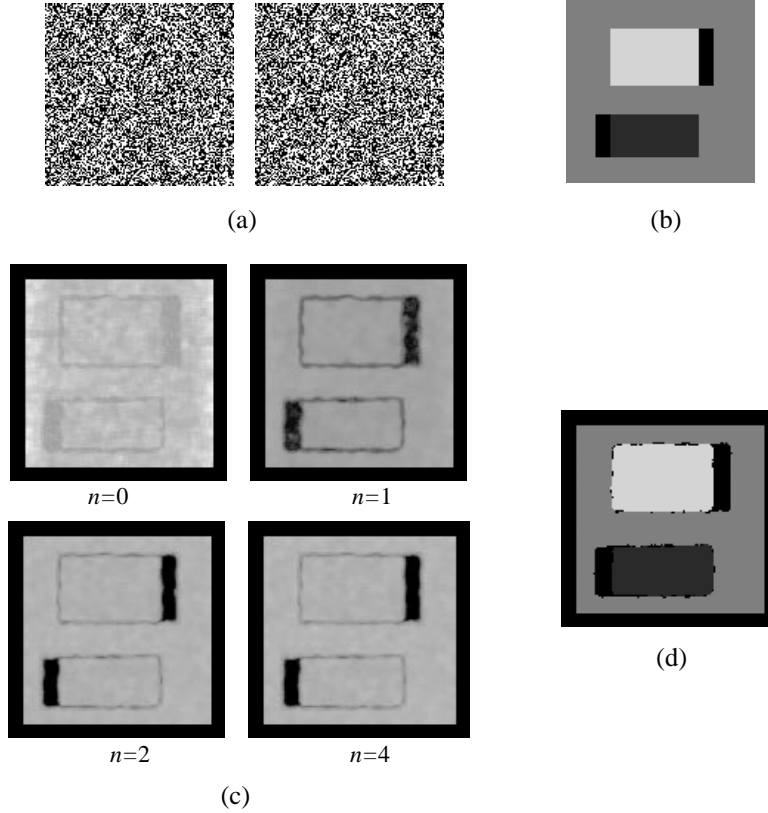


Figure 3;(a) Left and right random dot stereo images, 50% density; (b) Correct disparity map, where black regions are occluded; (c) Sum of match likelihood values for each pixel at  $n$  iteration; (d) Computed disparity map; black regions are identified as occluded.

#### 4. Summary of Algorithm

The summary of the volumetric iterative algorithm is as follows:

1. Set initial match likelihood values  $L_0$  using intensity correlation, appendix A.
2. Update match likelihood values  $L_n$  using (3).
3. Repeat step 2 until the match likelihood values converge.
4. For each pixel, find the voxel with the maximum match likelihood value.
5. Threshold the sum of the match likelihood values to detect occlusion within the image.

The running time for steps 1 through 5 is on the order of  $N^2 * D * I * C$ , where  $N^2$  is the size of the image,  $D$  is the number of disparity voxels for each row and column,  $I$  is the number of iterations and  $C$  is the number of cameras. The amount of memory needed is on the order of  $N^2 * D$ .

## 5. Finding Correct Matches with Repetitive Textures

Repetitive textures lead to a unique difficulty in stereo matching. They may result in multiple voxels with high match likelihood values for a pixel in a repetitive texture. It is difficult, if not impossible, to locally determine which voxel corresponds to the correct match when using a small number of images.

Consider the projection of a four post picket fence shown in figure 4(a). If each post of the picket fence is identical to the others, there will be four voxels with high match likelihood values for each post. We can group the neighboring high valued voxels into potential surfaces. Each potential surface has a consistent interpretation of the scene, i.e. each post in the left image is mapped to a unique post in the right image or interpreted as occluded.

What happens to the match likelihood values as the number of iterations increase? During each iteration the match likelihood value  $L_n(r, c, d)$  varies based on its  $S_n(r, c, d)$  value and the number of high valued voxels inhibiting it relative to others in  $P(r, c, d)$ . The match likelihood values of voxels 1 and 16 will decrease because they have fewer neighboring voxels with high values, and thus lower  $S_n$  values. As the values of voxels 1 and 16 are inhibited, the match likelihood values on the edge of the other potential surfaces will increase or decrease based on the size of their potential surfaces. Voxel 7 within the larger potential surface has only four inhibiting voxels, 2, 4, 11 and 14, as compared to the voxels, such as 11, within the smaller potential surfaces having five. Then during an iteration the match likelihood values of the less inhibited voxel 7 will increase while the values of voxels 2, 4, 11 and 14 decrease. Similarly, as iterations proceed, the remaining high valued voxels within the smaller potential surfaces will decrease due to inhibition from the larger potential surface. After enough iterations the smaller potential surfaces will be removed.

Of course, an interesting question is whether the larger potential surface is correct. As illustrated in figure 4(a), if the object with the repetitive texture is not occluded, the largest potential surface is correct. Intuitively this makes sense with the four post picket fence. However, the largest surface is not always the correct surface. As shown in figure 4(b) it is possible for a false surface to be the largest surface when it is occluded. As discussed in [23], this case is rare and its occurrence can be further minimized if smaller baselines are used.

Figure 5 demonstrates how well the algorithm can deal with the locally ambiguous stereo matching problem with repetitive texture. The synthetic example scene is a slanted roof, figure 5(a), with a repetitive texture with random noise. Figure 5(b) shows the left and right images. Figure 5(c) shows the results obtained by using simple window correlation with a fixed small (3x3) and large (15x15) window. The results for both have many mismatches. Figure 5(d) shows the progression of our volumetric algorithm. At iteration 0, we observe that the disparity possesses many mismatches. As the iterations proceed the false potential surface is inhibited and the mismatches are removed. The final disparity map, shown in figure 5(e), is smooth and possesses only a few mismatches.

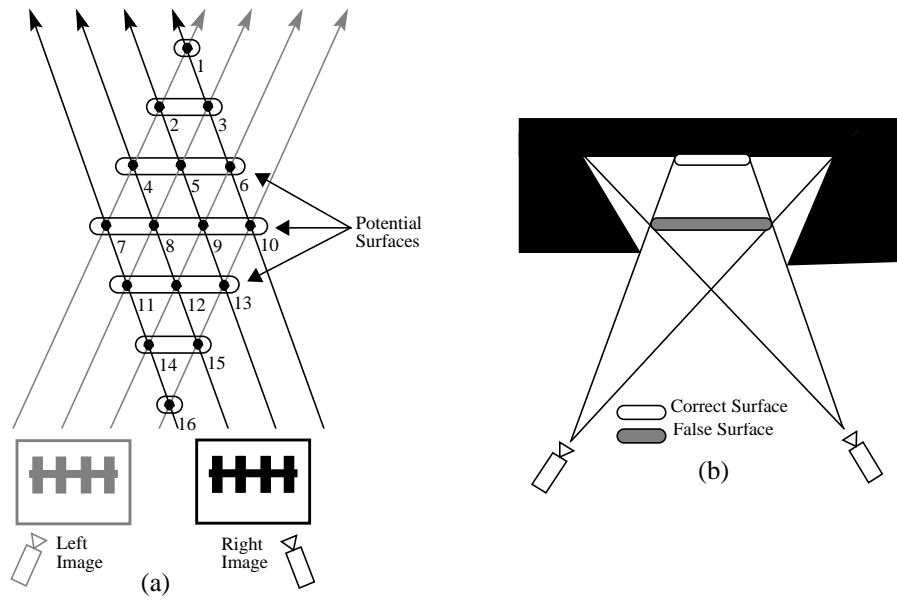


Figure 4; (a) When the surface with a repetitive texture is not occluded the largest potential surface is the correct surface; (b) When the surface with the repetitive texture is occluded the largest potential surface may not be the correct surface.

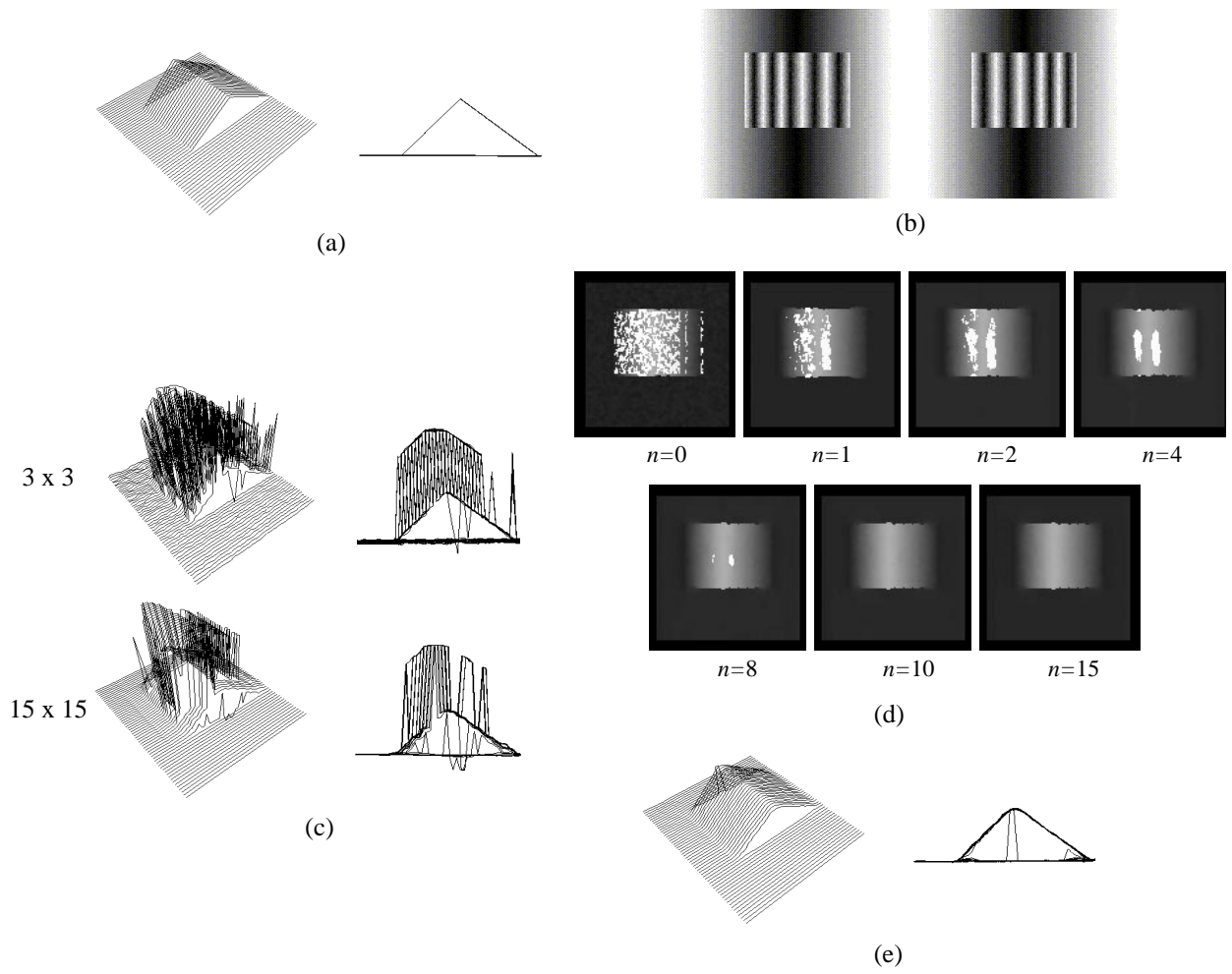


Figure 3: Results using synthetic data; (a) Oblique and front view of true disparity map; (b) Synthetic left and right images; (c) Oblique and front view of disparity map using small window(3x3) and large window(15x15); (d) Disparity maps for volumetric method after iteration  $n$ ; (e) Oblique and front view of disparity map using volumetric method.

## 6. Experimental Results with Real Images

To demonstrate the effectiveness of our algorithm we have applied it to several real images. The input images are rectified and normalized when necessary. The initial match likelihood values are computed using a sigmoid function of sums of absolute differences, appendix A. For all images, unless stated otherwise, the number of iterations is 10 with  $\alpha$  set to 2. We use a 3D local support with a fixed width, height and depth (7x7x3) in (3). Linear interpolation between disparity estimates is also done to achieve higher accuracy. The threshold for detecting occlusion is set constant for all image sets at 0.1. The running time of the algorithm on an Indigo 2ex for 256x256 images and 80 disparity estimates is about 20 seconds per iteration.

Figure 6 presents the stereo image pair of the “Coal Mine” scene and the processing results. The multi-baseline [15] method and the adaptive window [7] method are applied to the stereo for comparison. The multi-baseline result (Figure 6 (f)) that uses three input images is clearly the noisiest of the three. The result of the adaptive window approach (Figure 6 (g)) is smooth in general, but a few errors remain, especially around the small building next to the tower in the center of the image and the slanted roof in the upper corner of the scene is overly smoothed. The result of the volumetric approach (Figure 6 (e)) is smooth while recovering several details at the same time. The slanted roof of the lower building and the water tower on the roof top are clearly visible. Depth discontinuities around the small building next to the tower are preserved.

The second scene is a “model town,” figure 7. The results using multi-baseline and multi-baseline with adaptive window are also given. The same number of images were used for each approach as in the coal scene. The multi-baseline approach has several spikes and is generally noisy. The adaptive window method produced smooth results, especially on the faces of buildings, but details on the lower center building are lost. We observe, however, it behaves poorly in areas of changing disparity, such as the ground in the lower portion of the image. There is also a erroneous spike of disparity in the middle of the image. The volumetric approach is precise with no gross errors. Details are found on the face of the lower center building and the ground is correctly captured. The volumetric method also produces no errors in the presence of the repetitive background texture in the upper part of the image. Due to the repetitive texture 15 iterations were needed instead of 10. If only two images were used with a non-iterative area-based method the background would produce many mismatches [15].

Figure 8 shows results from a pair of aerial images of the Pentagon. Both of the two bridges to the right of the image are found. Details on the Pentagon’s roof are visible. Also noticeable are correctly detected occlusions within the Pentagon’s roof and to its right. Occlusions are also detected to the lower right of the Pentagon in areas which vehicles moved between the two images.

Figure 9 is a SRI lab scene. The thin white tape running up the center of the image is largely, though not all, captured. The plants are also found, but the occluded wall between the plant’s leaves is incorrectly matched. A specular reflection on the lower table causes disparity values to be warped.

The last scene is a SRI sequence of a group of trees shown in figure 10. This sequence presents difficulty in finding correct disparity estimates due to the high disparity variation within the scene. Both the results using a small and large baseline stereo pair are detailed with few mismatches. The tree and tree stump are clearly visible. In the small baseline example a small group of leaves in the top right of the image is correctly found. When a large baseline is used, the area to the right of the tree is correctly found to be occluded.

## 8. Conclusion

The volumetric iterative algorithm for stereo matching is capable of producing smooth and detailed disparity maps. First proposed by Marr and Poggio [10], two assumptions are made about the disparity maps: the disparity maps have unique values and are continuous almost everywhere, which implies objects within the scene are opaque and smooth. A 3D set of match likelihood values is constructed with each value corresponding to a pixel location in the reference image and a disparity relative to another image. The iterative algorithm updates the match likelihood values are updated taking advantage of the two assumptions. To find a unique match, the values of voxels on the same line of sight inhibit each other. To find a smooth match, the values among the neighboring support region are averaged. Details are maintained during each iteration by restricting the growth of match likelihood values based on their initial values. Finally, after the values have converged, occlusions are explicitly detected.

As demonstrated by using several synthetic and real image examples, the resulting disparity map is smooth and detailed. Disparity values in areas of repetitive textures were also found correctly.

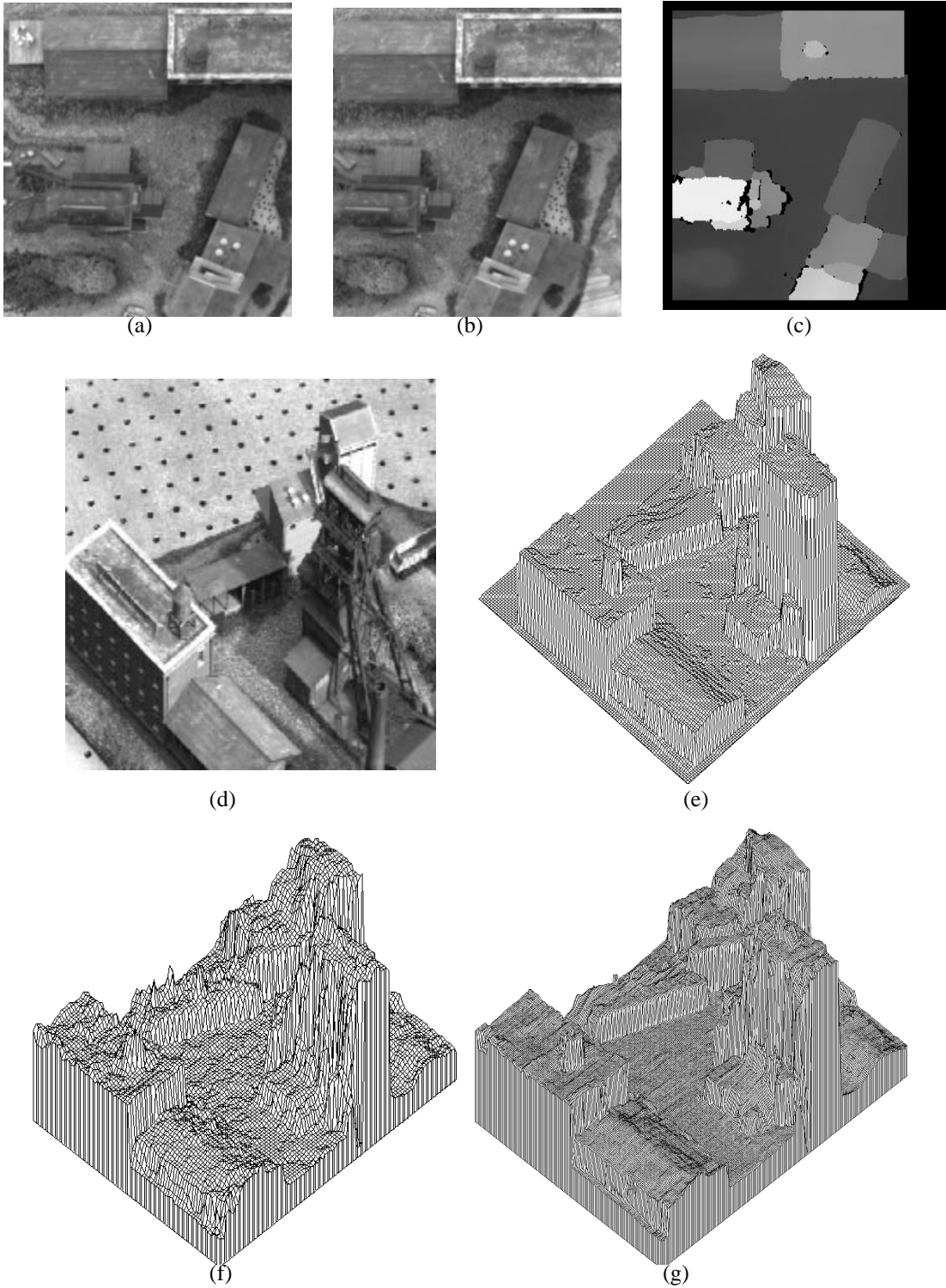


Figure 6: Coal Mine Scene; (a) Reference (left) image; (b) Right image; (c) Disparity Map, black areas are detected occlusions; (d) Oblique view of the coal mine model; (e) Isometric plot of the disparity map using volumetric iterative method with two images; (f) Isometric plot of the disparity map using multi-baseline stereo with three images as presented in [15]; (g) Isometric plot of disparity map using multi-baseline stereo with adaptive window with three images as presented in [7].



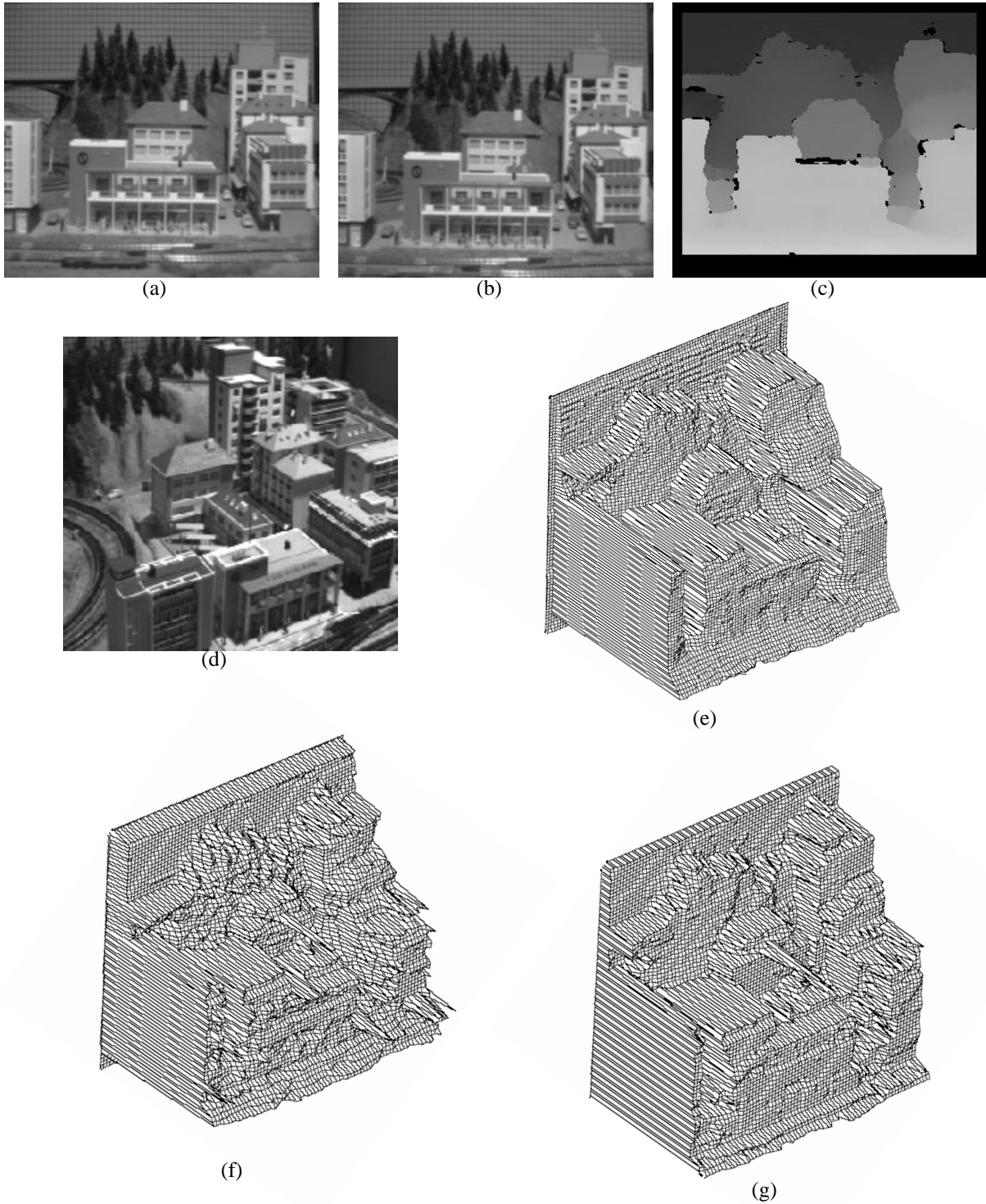


Figure 7: Town Scene; (a) Reference (top) image; (b) Bottom image; (c) Disparity Map, black areas are detected occlusions; (d) Oblique view of the town model; (e) Isometric plot of the disparity map using volumetric iterative method with two images; (f) Isometric plot of the disparity map using multi-baseline stereo with three images as presented in [15]; (g) Isometric plot of disparity map using multi-baseline stereo with adaptive window with three images as presented in [7].

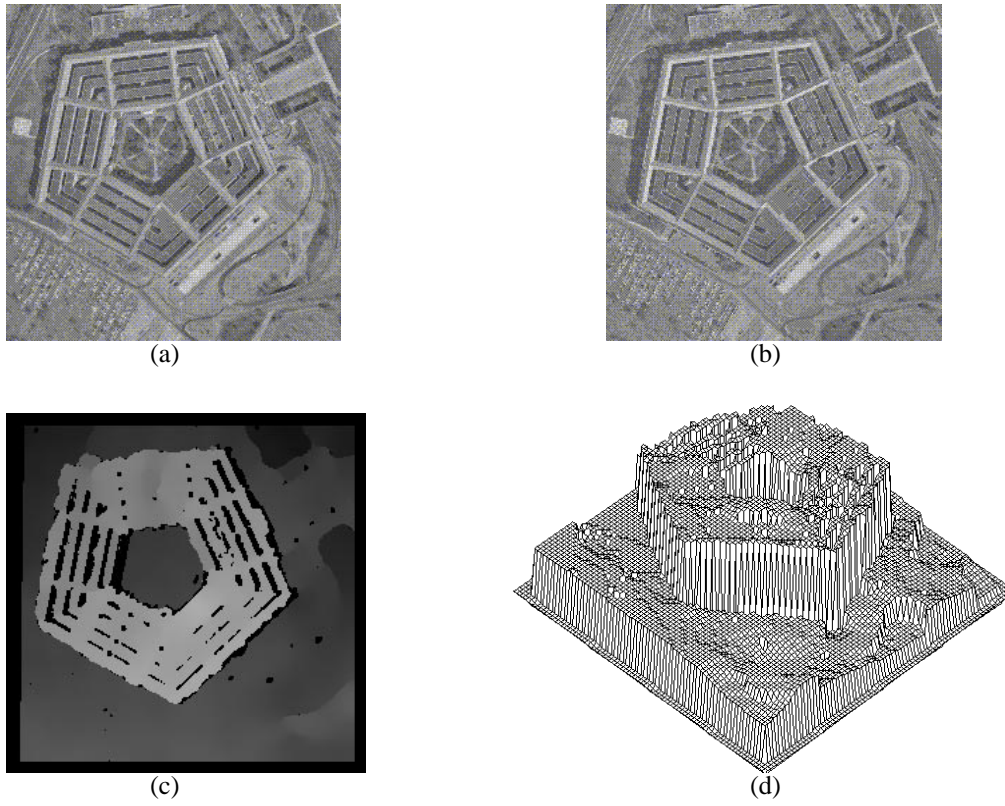


Figure 8: Pentagon scene; (a) Reference (left) image; (b) Right image; (c) Disparity map, black areas are detected occlusions; (d) Isometric plot of disparity values.

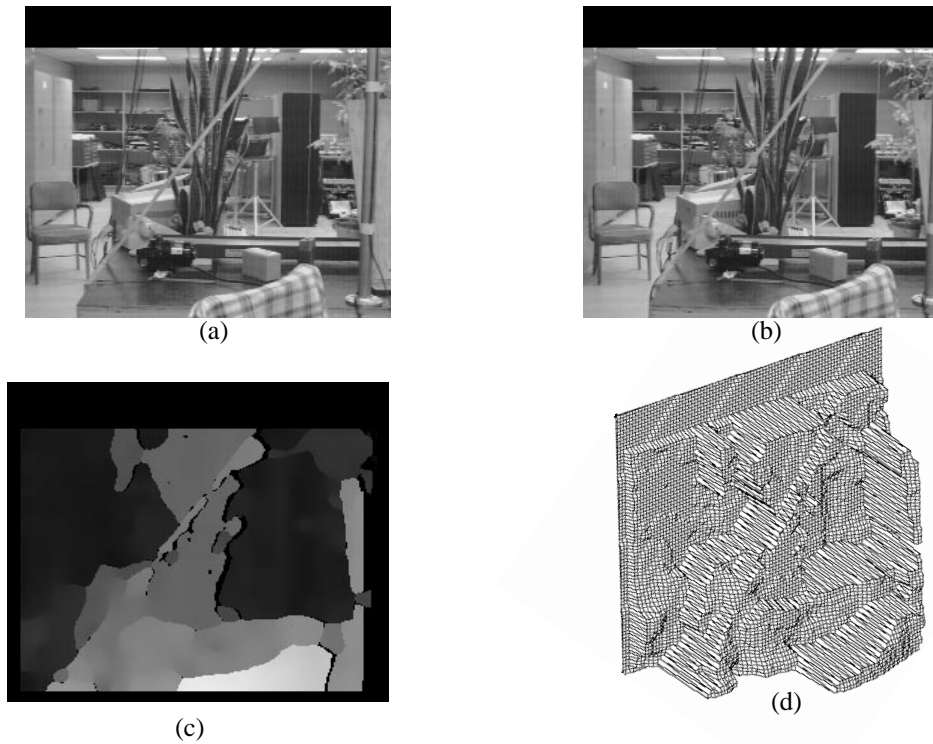


Figure 9: SRI lab scene; (a) Reference (left) image; (b) Right image; (c) Disparity map, black areas are detected occlusions; (d) Isometric plot of disparity values.

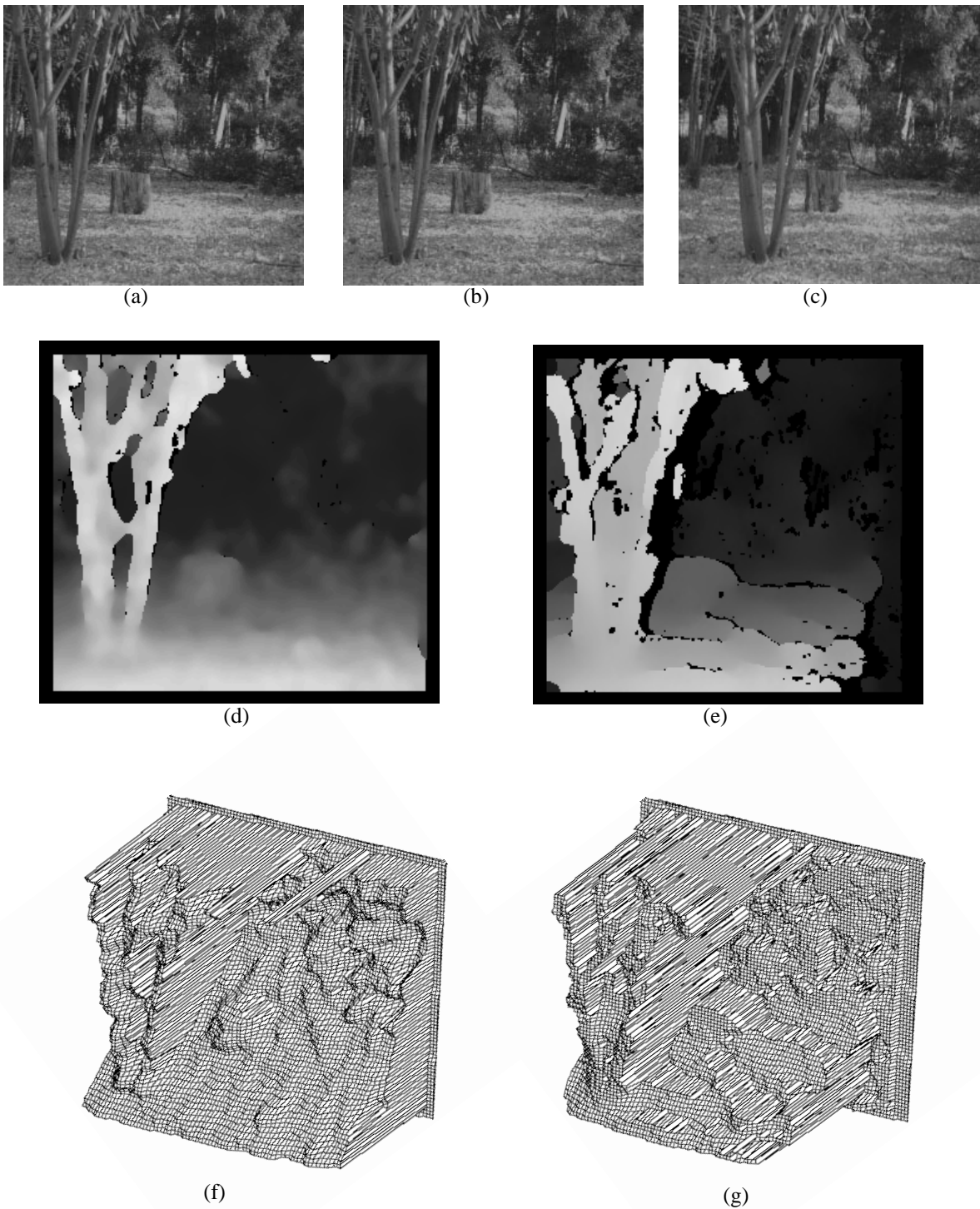


Figure 10: SRI tree scene; (a) Reference (left) image; (b) Right image for short baseline; (c) Right image for large baseline; (d) Disparity Map for small baseline; (e) Disparity map for large baseline; (f) Isometric plot of disparity using small baseline; (g) Isometric plot of disparity using large baseline.

## References

- [1] H. H. Baker and T. O. Binford, 1981, "Depth from edge and intensity based stereo," In *Proc. of the 7th International Joint Conference on Artificial Intelligence*, Vancouver, 1981, pp. 631-636.
- [2] S. T. Barnard and M. A. Fischler, "Stereo Vision," in *Encyclopedia of Artificial Intelligence*. New York: John Wiley, 1987, pp. 1083-1090.
- [3] R. Collins, "A space-sweep approach to true multi-image matching," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Francisco, CA, 1996, pp. 358-363.
- [4] P. Fua, "A parallel stereo algorithm that produces dense depth maps and preserves image features," *Machine Vision and Applications*, vol. 6, pp. 35-49, 1993.
- [5] W. Forstner and A. Perti, *Photogrammetric Standard Methods and Digital Image Matching Techniques for High Precision Surface Measurements*. New York: Elsevier Science Publishers B. V., 1986, pp. 57-72.
- [6] W. E. L. Grimson, "Computational experiments with a feature based stereo algorithm," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol 7, no. 1, pp. 17-34, 1985.
- [7] T. Kanade and M. Okutomi, "A stereo matching algorithm with an adaptive window: Theory and experiment," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, no. 9, pp. 920-932, 1994.
- [8] M. Levine, D. O'Handley and G. Yagi, "Computer determination of depth maps," *Computer Graphics and Image Processing*, vol. 2, no. 4, pp. 131-150, 1973.
- [9] S. A. Lloyd, E. R. Haddow and J. F. Boyce, "A parallel binocular stereo algorithm utilizing dynamic programming and relaxation labelling," *Computer Vision, Graphics and Image Processing*, vol. 39, pp. 202-225, 1987.
- [10] D. Marr and T. Poggio, "Cooperative computation of stereo disparity," *Science*, vol. 194, pp. 209-236, 1976.
- [11] H. Moravec, "Robot spatial perception by stereoscopic vision and 3D evidence grids," CMU Robotics Institute Technical Report CMU-RI-TR-96-34, 1996.
- [12] K. Mori, M. Kidode and H. Asada, "An iterative prediction and correction method for automatic stereo comparison," *Computer Graphics and Image Processing*, vol. 2, pp. 393-401, 1973.
- [13] Y. Nakamura, T. Matsuura, K. Satoh and Y. Ohta, "Occlusion detectable stereo --- Occlu-

sion patterns in camera matrix,” *IEEE Conf. on Computer Vision and Pattern Recognition*, 1996.

[14] H. K. Nishihara, and T. Poggio, “Stereo vision for robotics,” *ISRR Conference*, Bretton Woods, NH, 1983.

[15] M. Okutomi and T. Kanade, “A multiple-baseline stereo,” in *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 15, no. 4, pp. 353-363, 1993.

[16] Y. Ohta and T. Kanade, “Stereo by intra- and inter-scanline search using dynamic programming,” in *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 7, pp. 139-154, 1985.

[17] D. J. Panton, “A flexible approach to digital stereo mapping,” *Photogram. Eng. Remote Sensing*, vol. 44, no. 12, pp. 1499-1512, 1978.

[18] S. B. Pollard, J. E. W. Mayhew, and J. P. Frisby, “Pmf: A stereo correspondence algorithm using a disparity gradient limit,” *Perception*, vol. 14, pp. 449-470, 1985.

[19] K. Prazdny, “Detection of binocular disparities,” *Biological Cybernetics.*, vol. 52, no. 2, pp. 93-99, 1985.

[20] D. Scharstein, and R. Szeliski, “Stereo matching with nonlinear diffusion,” *International Journal of Computer Vision*, vol. 28, no. 2, pp. 155-174, 1998.

[21] R. Szeliski and P. Golland, “Stereo matching with transparency and matting,” in *Sixth International Conference on Computer Vision.*, Bombay, India, pp. 517-524, 1998.

[22] G. Wood, “Realities of automatic correlation problem,” *Photogrammetric Engineering and Remote Sensing*, vol. 49, pp. 537-538, 1983.

[23] C. L. Zitnick and J. A. Webb, “Multi-baseline stereo using surface extraction,” CMU Technical Report CMU-CS-96-196, 1996.

## Appendix A: Initial Match Likelihood Values

The initial match likelihood values are set to range from 0 to 1. A function which maps intensity correlation to match likelihood values must possess a couple properties. Areas of low intensity correlation should be close to 0. Areas with high correlation should gradually approach 1 to allow some smoothing of the data. Linear, sigmoid, exponential and locally adjusting functions were all tested. Of these the sigmoid function produced the best data.

Let  $\epsilon_{SAD}(r, c, d)$  be the sums of absolute differences (SAD) within a  $3 \times 3$  window  $W$  for pixel  $(r, c)$  at disparity  $d$ . If  $I_L(r, c)$  and  $I_R(r, c)$  are the intensity values for the left and right images respectively at pixel  $(r, c)$  then:

$$\epsilon_{SAD}(r, c, d) = \sum_{(r', c') \in W} |I_L(r + r', c + c') - I_R(r + r', c + c' + d)|$$

We set the initial likelihood value  $L_0(r, c, d)$  by:

$$L_0(r, c, d) = \frac{1}{1 + e^{\frac{-(\epsilon_{SAD}(r, c, d) - \lambda)}{\mu}}}$$

For the examples within this paper, the values  $\mu$  and  $\lambda$  are set equal to the standard variation of all SAD values.