Occlusion Detectable Stereo
– Occlusion Patterns in Camera Matrix –

Yuichi NAKAMURA, Tomohiko MATSUURA, Kiyohide SATOH, Yuichi OHTA

Institute of Information Sciences and Electronics, University of Tsukuba
Tsukuba, Ibaraki, 305, JAPAN (yuichi@is.tsukuba.ac.jp)

Abstract

In stereo algorithms with more than two cameras, the improvement of accuracy is often reported since they are robust against noise. However, another important aspect of the polynocular stereo, that is the ability of occlusion detection, has been paid less attention. We intensively analyzed the occlusion in the camera matrix stereo (SEA) and developed a simple but effective method to detect the presence of occlusion and to eliminate its effect in the correspondence search. By considering several statistics on the occlusion and the accuracy in the SEA, we derived a few base masks which represent occlusion patterns and are effective for the detection of occlusion. Several experiments using typical indoor scenes showed quite good performance to obtain dense and accurate depth maps even at the occluding boundaries of objects.

1 Introduction

Occlusion is one of the most difficult problems in stereo vision [3, 9, 13]. Especially in binocular case, when a target point can not be seen at one of the cameras, the triangulation is impossible. On the other hand, if more than two cameras are used, the target points that are visible from only one camera decrease as the number of cameras increases. Also the false targets can be reduced by using redundant information obtained by the third or successive cameras. The improvement of accuracy by using trinocular stereo has been reported in several papers [1, 8, 10, 14, 16].

The accuracy around occluding boundaries, however, is not simply improved, since the target points invisible from at least one camera may increase than the binocular case. The reason for this is that most algorithms previously proposed are not able to discriminate occlusion from noise. To cope with this problem, we intensively analyzed the occlusion in the camera matrix stereo SEA (Stereo by Eye Array) proposed by ourselves, and developed a method to discriminate the presence of occlusion from the presence of noise. By considering several useful statistics, we derived a few masks to represent occlusion patterns which are effective for the detection of occlusion. The experiments of SEA with 3×3 camera matrix and 5×5 camera matrix showed quite good performance to obtain a dense and accurate depth map even around occluding boundaries.

2 Detection of Occlusion

2.1 Occlusion in Stereo

The false matches in stereopsis are mainly caused by the following two reasons.

(a) The evaluation of the correct disparity becomes worse because of the presence of noise.

(b) The evaluation of the correct disparity becomes worse because the correct corresponding points are invisible at one or more cameras.

In binocular stereo, the discrimination between the above two cases is basically impossible without global information. If more than two cameras are used in stereopsis, the cases of (a) and (b) may be separable by using local spatial relation of the evaluation values. This is mainly because the occurrence of (b) is heavily dependent on the location of cameras, but the occurrence of (a) is almost independent. If the presence of occlusion is detected, the correct triangulation can be made by simply omitting the cameras to which the target point is invisible.

Then the problem is how to represent the spatial dependence of occlusion and how to utilize it in the correspondence search. Before describing this, we show the framework of SEA, the camera matrix stereo[11, 12].

2.2 SEA: Stereo by Eye Array

Coordinate System

Figure 1 illustrates the coordinate system of the camera matrix. The origin of the scene coordinates is located at

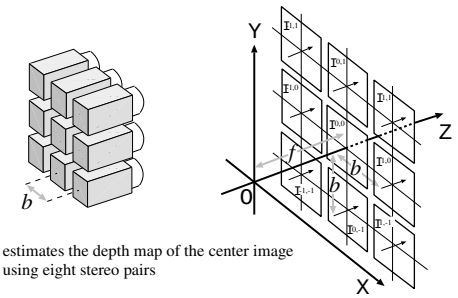


Figure 1: Geometry of SEA

the lens center of the center camera. Other cameras are located at a grid on the X-Y plane at an equal interval of b. The image captured by each camera in $M \times N$ matrix is labeled as $I^{m,n}$ $(m = -\frac{M-1}{2}, \cdots, 0, \cdots, \frac{M-1}{2};\ n = -\frac{N-1}{2}, \cdots, 0, \cdots, \frac{N-1}{2})$.

The optical axes of all cameras are set to be parallel with each other. The image coordinates x-y of each camera are set to be parallel to the X-Y axes of the scene coordinates. All cameras have a same focal length f.

For simplification of description, we assume below that the size of the camera matrix is 3×3, *i.e.* nine cameras are used. In this configuration, SEA uses eight stereo pairs constructed between the center image and each of the eight peripheral images. Target point $P(x, y, z)$ is observed on the center image $I^{0,0}$ at image point

$$I^{0,0}(f \cdot X/Z,\ \ f \cdot Y/Z). \tag{1}$$

When there is no occlusion, P is also observed on each peripheral image $I^{k,l}$ at image point

$$I^{k,l}(f \cdot (X - kb)/Z,\ f \cdot (Y - lb)/Z). \tag{2}$$

Or we can denote them as $I^{0,0}(x, y)$, and $I^{k,l}(x - kd,\ y - ld)$, where x, y, and d are defined as $x = f \cdot X/Z$, $y = f \cdot Y/Z$, $d = f \cdot b/Z$, respectively. d is called disparity.

Basic Algorithm

At first, the dissimilarity values $e^{k,l}(x, y, d)$ between $I^{0,0}(x, y)$ and $I^{k,l}(x - kd, y - ld)$ are computed for each stereo pair, assuming that the disparity is d. The dissimilarity between two points is evaluated as the summation of RGB distances within a small window whose center is located at each of the two points. For each d, $e^{k,l}(x, y, d)$ are summed up to make $e(x, y, d)$ as the penalty of the disparity d.

$$e(x, y, d) = \sum_{k,l} e^{k,l}(x, y, d) \tag{3}$$

The correct disparity $\hat{d}$ at $I^{0,0}(x, y)$ is estimated by choosing a value d which satisfies the following equation:

$$\hat{d} = \underset{d}{\operatorname{argmin}}\ e(x, y, d) \tag{4}$$

By using a number of image pairs similarly to the multiple-baseline stereo [10, 17, 18], we can greatly reduce the false targets caused by false corresponding points which unexpectedly indicate good similarity. This realizes a dense disparity map with high spatial resolution.

Occlusion Detectable Algorithm

In SEA, simple extension to the basic algorithm realizes a quite effective algorithm to cope with occlusion. We define "occlusion masks", which are typical occlusion patterns occurring in real scenes. In the following explanation, the eight masks shown in Figure 2 are assumed as the occlusion masks.

Each mask $M_t(k, l)(t = 1, 2, \cdots, 8)$ represents a pattern of occlusion depending on the orientation of occluding boundary. The gray cell or white cell on $M_t(k, l)$ indicates whether occlusion occurs or not between the image pair $I^{0,0}$ and $I^{k,l}$; gray cell (the value is 0) indicates that occlusion occurs and white cell (the value is 1) indicates not. By omitting the images $I^{k,l}$ whose values in $M_t(k, l)$ are 0, the effect of occlusion can be eliminated. For this purpose, we redefine the penalty $e(x, y, d)$ in equation 3 by incorporating the assumption on occlusion. First, we define the value $e_t(x, y, d)$ as the penalty of the disparity d with the assumption on occlusion type t. The essential property of this penalty is that it does not count for the $e^{k,l}$ of the occluded image $I^{k,l}$.

$$e_t(x, y, d) = \frac{w_t}{n_t} \sum_{M_t(k,l)=1} e^{k,l}(x, y, d) \tag{5}$$

where n_t indicates the number of $1s$ in mask $M_t(k, l)$. w_t is a constant to give a certain bias to the selection process that the non-occluding case is preferred if occlusion does not occur. It is set to 1 when $t = 0$ and set to a constant slightly greater than 1 when $t \neq 0$.

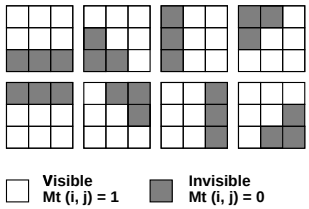
The $e(x, y, d)$ is redefined by selecting the minimum of the $e_t(x, y, d)$,

$$e(x, y, d) = \min_{t=0,1,\cdots,8} e_t(x, y, d). \tag{6}$$

Estimating $\hat{d}$ by using equation 4, the type t selected for the estimated disparity in equation 6 indicates the presence of occlusion and the direction of occluding boundary.

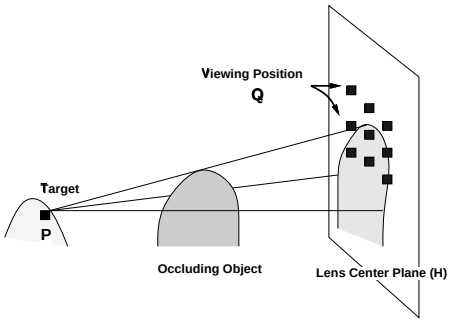
2.3 Criteria for the Determination of Occlusion Masks

By detecting occlusions, it becomes possible to estimate the correct disparities around the occluding boundaries. This enables us to acquire the disparity maps with sharp object boundaries. However, there still remains problems concerning mask selection: what kinds of and how many occlusion masks are required?

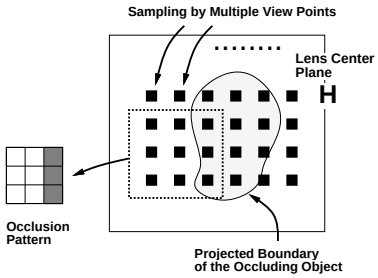


$M_t(k,l)$ is 0 for gray cells (cells labeled invisible), 1 for white cells (cells labeled visible).

Figure 2: Eight occlusion masks



(a) Geometrical relationship



(b) Projected boundary and occlusion pattern

Figure 3: Geometrical relationship among target point, occluding boundary, and viewing position.

(a) Scene1

(b) Scene2

(c) Scene3

Figure 4: Three scenes used in the experiments

In the followings, we investigate the optimal occlusion mask sets for the camera matrix stereo. In determining a set of occlusion masks, we have to consider several factors:

1. The number of occlusion masks should be small. In the stereo matching algorithm in SEA, the computational time is proportional to the number of masks.
2. The occlusion masks must be close to actual occlusion patterns. In other words, actual occlusion patterns have to be substituted by the occlusion masks.
3. The closeness between an occlusion mask and an actual occlusion pattern should be considered in the context of stereo matching. There is remarkable performance difference in the following two cases: (a) Regarding occluded viewing positions as visible in the penalty estimation by equation 6; (b) Regarding visible viewing positions as occluded in the penalty estimation by equation 6. The accuracy deterioration caused by (a) is much worse than that by (b). Hereafter, we call (a) as ItoV (regarding Invisible target as Visible), (b) as Vtol (regarding Visible target as Invisible).
4. To make use of the advantage of polynocular stereo, *i.e.* robustness against noise, it is better to use as many images as possible. However, if we simply increase the cells labeled visible in occlusion masks, the possibility of ItoV may increase. There exists trade-off between the robustness against noise and the applicability to actual occlusion patterns.

To determine a set of mask satisfying the above criteria we examined several statistics: concerning the second and the third criterion, the occlusion patterns in real scenes are observed; the accuracy degradation caused by ItoV or Vtol is measured concerning the third criterion; the accuracy degradation caused by reducing the number of cameras is measured concerning the forth criterion.

3 Occlusion Patterns in SEA

To clarify the characteristics of the occlusion patterns in actual scenes, let us consider the geometry of SEA. The visibility of a target point in the SEA is determined by the geometrical relation among that point, the viewing position (*i.e.* the lens center of the camera) and the boundary of the occluding object. This relation is illustrated in Figure 3(a). If the point is visible from a camera, the viewing position must be visible from the point. In the other words, if point **P** is visible from point **Q**, **Q** must be lit by a point light source located at **P**.

Since the lens center of each camera is located on a grid point on plane **H**, it is enough to consider the projection of the object boundary onto the plane **H**. Figure 3(b) shows this relationship. If the lens center of a camera is inside the shadow cast by the occluding object, the point **P** is invisible from the camera. Therefore, the occlusion patterns in SEA can be derived from the projected boundaries of objects in real scenes by coarse sampling [1].

The actual patterns and their probability of occurrence are shown in Figure 5. They are measured by using the two indoor scenes shown in Figure 4[2]. In this measure, the patterns which are different only by rotation are merged into a single pattern. We can see that some occlusion patterns have quite high probabilities compared to other patterns with the same number of occluded viewing positions.

The occluded viewing positions cause large dissimilarity values in the matching process. As mentioned before, a large dissimilarity value may also be caused by noise. In the case of noise, however, we can assume that their occurrence is random and the patterns with a same number of viewing positions with large dissimilarity should have similar probabilities. Then it will be reasonable to decide that such a pattern shown in the left side of Figure 5 is caused by occlusion, not by noise.

4 Conditions for Accuracy Degradation

4.1 Visible/Invisible Confusion

The effect of misjudgement on the visibility of view points in disparity estimation is measured. The SEA algorithm is applied only to the target points with occlusion. For this statistics, three scenes as shown in Figure 4 are used. The true depth map manually created for each scene is also used.

Each figure in Table 1 shows the ratio of target points for which SEA can estimate correct disparities under correct visibility judgement but fails under one of the following two cases.

1. Regarding an occluded viewing position as visible (ItoV for a cell in the correct occluding pattern).
2. Regarding a visible viewing position as occluded (VtoI for a cell in the correct occluding pattern).

[1] In complicated situations, more than one object may occlude a target point. So generally, the 'logical-OR' of multiple boundaries must be considered as a projected boundary

[2] Since we manually created true disparity maps for three scenes as shown in Figure 12, actual occlusion patterns can be automatically detected from them. We collected every occlusion pattern around every occluding boundary.

Table 1: Accuracy degradation by visible/invisible confusion

Each figure shows the accuracy degradation caused by VtoI or ItoV in disparity estimation in 3×3 or 5×5 stereo.

	Scene1 VtoI/ItoV	Scene2 VtoI/ItoV	Scene3 VtoI/ItoV
3×3	3.5/32	6.4/45	8.2/41 (%)
5×5	0.74/11	1.4/15	1.4/16 (%)

We can observe that the effect by ItoV is much serious than that by VtoI. It should be avoided in the actual stereo matching. Therefore, at most one ItoV is allowed in the successive experiments. On the other hand, the accuracy degradation caused by VtoI is small enough to be neglected when the VtoI cells are few. We can also observe that the ratio of degradation by ItoV and VtoI is roughly 10:1.

4.2 Number of Views

The statistics in the previous section show that ItoV should be kept as small as possible. However, an occlusion mask with small number of cells labeled visible makes the stereo matching less robust in the sense of noise tolerance.

To investigate this trade-off, we examined the relation between the matching accuracy and the number of cameras. For this purpose, only the target points which are visible at all the cameras are processed to avoid the effect of occlusion.

The number of cameras and their arrangements are binocular, collinear trinocular[9], orthogonal trinocular[8], pentanocular arranged in "+", pentanocular arranged in "$\times$", 3×3 camera matrix, and 5×5 camera matrix. The result is shown in Table 2. The improvement of accuracy by increasing the number of cameras almost saturates around five cameras. This implies that when there is five cells labeled visible in an occlusion mask, it has no merit to try to increase the number of visible cells under the risk of ItoV confusion.

5 Occlusion Mask

5.1 Base Mask Pattern

The base mask pattern means the mask pattern with normalized orientation and size. Actual occlusion masks can be generated by rotating and resampling a base mask pattern. The following two criteria will be reasonable to determine the base mask patterns.

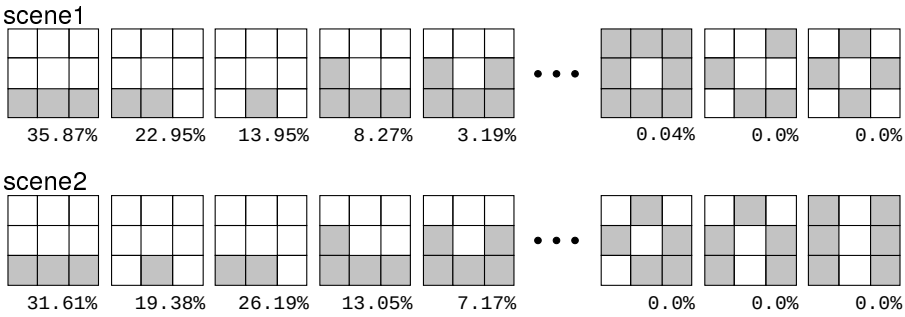


Figure 5: Probabilities of occlusion patterns.

Table 2: Relation between the number of cameras and the matching accuracy

cameras	2	3 —	3 L	5 +	5 x	9 3×3	25 5×5
Scene1	67	72	84	87	88	88	96 (%)
Scene2	59	64	74	79	80	83	84 (%)
Scene3	65	75	74	82	81	84	85 (%)

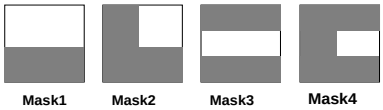


Figure 6: Base mask patterns

- According to the claim in Section 3, mask patterns should fit the occluding boundaries projected onto lens center plane. Therefore, the spatial frequency is usually low for occlusion patterns to be considered in actual indoor scenes.
- The invisible cells can replace the visible cells as far as a mask has a certain number of visible cells. More than five visible cells are almost redundant for the false target reduction. Too many visible cell may increase the danger of ItoV confusion.

Considering the above criteria, we composed four base mask patterns as shown in Figure 6. These patterns have spatial frequency (1, 0), (1, 1), (3/2, 0), and (3/2, 1) in vertical and horizontal directions. While the base masks 1 and 3 seems to be similar to the base pattern of DCT or WH-Transform, their size is odd and the center of the mask is always labeled visible.

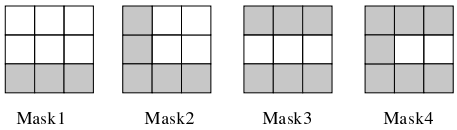


Figure 7: Base masks for 3×3 camera matrix

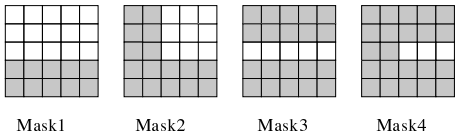


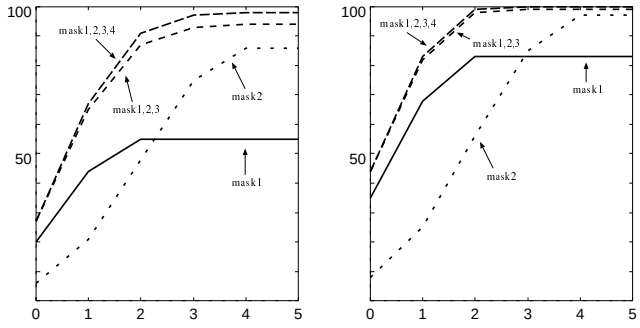
Figure 8: Base masks for 5×5 camera matrix

5.2 Actual Mask Set

The base mask patterns projected onto 3×3 mesh are shown in Figure 7. Only the base mask1 has more than five cells labeled visible. However, if we assume that the disparity estimation using four visible cells is accurate enough, the base mask1 may be substituted by the base mask2. That is because all cells labeled invisible in base mask1 are included in base mask2, base mask1 is substituted by base mask2 with two Vtol allowing small increase of error. The effectiveness of the rest two masks is doubtful, since base mask3 has only three cells labeled visible, and base mask4 has only two. The effectiveness will be examined in the experiments in Section 6.

All of the rotational variations of a base mask are used in the set of occlusion masks when the base mask is incorporated in the SEA algorithm.

The base mask patterns projected onto 5×5 mesh are shown in Figure 8. Base masks through 1 to 3 have more than five cells labeled visible. So, base mask1 may be substituted by base mask2 with small degradation of matching. Base mask4, however, does not have enough cells labeled visible as in the 3×3 case. There are 16 rotational variations for each of the base masks 1, 2, and 4, and eight variations for base mask3.



(a) $\mathsf{ItoV} = 0$ (b) $\mathsf{ItoV} = 1$

Figure 9: Coverage of base masks

5.3 Coverage of Base Mask

The coverage of the base mask set for the actual occlusion patterns should be examined. For this purpose, we compared the base masks and the occlusion patterns gathered in the indoor scenes shown in Figure 4.

The method of comparison is as follows:

- For each size of camera matrix, base masks and actual occlusion patterns are compared.
- The distance between a mask and an occlusion pattern is counted by the number of ItoV and Vtol cells.
- The number of ItoV is allowed for at most one cell, since the degradation by ItoV is serious.
- The percentage of occlusion patterns within a certain distance from at least one of the base masks is measured.

The result for 3×3 camera matrix is shown in Figure 9. As shown in the figure, the base masks can cover the majority of the occlusion patterns in real scenes. If base masks 1–3 are used, the occlusion patterns within the distance $(\mathsf{ItoV}, \mathsf{Vtol}) = (0, 3)$ from the base mask set occupy about 90% out of all, and the patterns within the distance $(1, 2)$ occupy 97%.

By adding base mask4, the rate increases by a few percent. However, the effectiveness of base mask4 should be carefully examined, since the number of cells labeled visible is not large enough both in 3×3 and 5×5 camera matrices. The answer to this question will be given by the experiments in Section 6.

6 Experiments

The performance in disparity estimation of the occlusion detectable algorithm in Section 2.2 is measured to verify the effectiveness of the mask set obtained in the previous section. The percentage is measured that the obtained disparity is within one pixel of the true disparity manually given. The results are summarized in Table 3. For all cases with occlusion masks, we can observe the performance improvement compared to the case without occlusion masks. Since scene3 is the most complex scene, in which leaves of potted plants are occluding other leaves with similar colors, the performance is lower than the others.

The obtained depth maps for the two scenes are shown in Figure 10 and Figure 11, and the true depth maps manually created are shown in Figure 12 for comparison. Improvements in the depth estimation by using occlusion masks can be observed anywhere. Sharp depth discontinuities are obtained at occluding boundaries around the desk lamp and the face mask in scene1, and around the contour of the doll in scene2. Of course, the improvements are different among the results obtained by using different sets of occlusion masks. For example, at the arm of the desk lamp in scene1, the result by base mask2 is better than that by base mask1. This tendency can also be seen in the performance rate in Table 3.

Generally speaking, the use of the occlusion masks derived from base mask2 without any additional masks brings us the best or the near best performance. The reason can be conjectured that the ratio of cells labeled visible or invisible in base mask2 is well balanced. As a result, they are robust against noise as well as they cover wide variety of occlusion patterns. On the contrary, occlusion masks derived from base mask3 and base mask4 are effective only in such a case as scene3 where quite complicated occlusion occurs. As we can see in Table 3, the performance improvement for scene1 and scene2 is small and even degradation is observed. The reason is clear. Since their cells labeled visible are not enough for noise elimination, the false targets have increased. Then not only the performance around occluding boundaries but also the performance at the part without occlusion is affected.

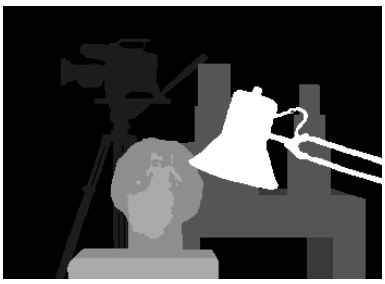
7 Conclusion

In this paper, a method for detecting and eliminating occlusion in polynocular stereo was presented. Base masks for the camera matrix stereo are designed as they can substitute the actual occlusion patterns in real scenes. For this purpose, several statistics are examined to clarify the criteria for determining the occlusion masks. By applying the occlusion masks to disparity estimation in SEA, drastic improvements were observed around occluding boundaries. The experiments showed that the occlusion masks derived from base mask2 always brings us the best or the near best performance for the typical indoor scenes.

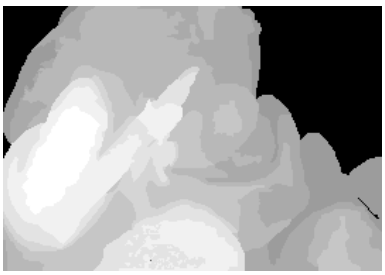
Table 3: Score by using automatic occlusion detection

Each value shows the rate of correct disparity obtained for each combination of a mask set and scene. The three values are express the rates for "whole image / targets without occlusion / targets with occlusion". For comparison, the performance is measured in the ideal case where the actual occlusion patterns are directly given as the occlusion masks. It is given at the row labeled "Actual".

	Scene1(3x3)	Scene2(3x3)	Scene3(3x3)	Scene1(5x5)	Scene2(5x5)	Scene3(5x5)
No Mask	92.7/98.3/44.8	88.7/95.3/28.1	72.9/94.3/36.1	92.2/99.6/61.6	90.1/98.0/53.6	71.3/95.6/42.3
Mask1	95.2/98.4/67.8	93.2/95.3/73.6	76.3/94.7/44.8	95.1/99.5/77.1	95.2/97.9/82.6	73.9/95.8/47.9
Mask2	95.7/98.4/72.7	93.9/95.2/82.0	78.5/94.6/51.0	96.1/99.5/81.9	96.6/97.8/90.8	74.9/95.8/50.0
Mask2, 3	95.7/98.3/72.7	93.9/95.2/82.0	80.5/94.3/56.8	96.2/99.6/82.0	96.6/97.8/90.8	76.1/95.3/53.3
Mask1–3	95.7/98.3/73.3	93.9/95.2/82.1	80.8/94.4/57.3	96.1/99.5/81.9	96.6/97.8/91.1	76.4/95.5/53.7
Mask1–4	95.7/98.3/73.3	93.8/95.1/82.2	83.6/94.3/65.1	96.1/99.5/82.0	96.5/97.7/90.8	82.5/95.3/67.4
Actual	96.6/98.3/82.3	94.1/95.4/81.3	84.0/94.3/66.5	98.2/99.6/92.3	97.2/98.0/93.3	84.3/95.6/70.9



(a) True depth map for scene1



(c) True depth map for scene2

Figure 12: True depth maps for scene1 and scene2

References

[1] Ayache, N., and Lustman, F., "Trinocular Stereo Vision for Robotics," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol.13, No.1, pp.73-85, 1991.

[2] Barnard, S., T., and Fischler, M., A., "Computational stereo," *ACM Computing Surveys*, Vol.14, No.4, pp.553-572, 1982.

[3] Belhumeur, P., N., "A Binocular Stereo Algorithm for Reconstructing Sloping, Creased, and Broken Surfaces in the Presence of Half-Occlusion," *Proc. Fourth Int. Conf. on Computer Vision*, pp.431-438, 1993.

[4] Bolles, R., C., Baker, H., H., and Marimont, D., H., "Epipolar-Plane Image Analysis: An Approach to Determining Structure from Motion," *Int. J. of Computer Vision*, Vol.1, No.1, pp.7-55, 1987.

[5] Dhond, U., R., and Aggarwal, J., K., "Structure from Stereo - A Review," *IEEE Trans. on System, Man and Cybernetics*, Vol.19, No.6, pp.1489-1510, 1989.

[6] Ens, J., and Li, Z., N., "Real-time Motion Stereo," *Proc. IEEE Conf. on Computer Vision and Pattern Recognition '93*, pp.130-135, 1993.

[7] Koschan, A., "What is New in Computational Stereo Since 1989: A Survey on Current Stereo Papers," Technischer Bericht 93-22, Technical University Berlin, 1993.

[8] Ohta, Y., Watanabe, M., and Ikeda, K., "Improving Depth Map by Right-Angled Trinocular Stereo," *Proc. 8th Int. Conf. on Pattern Recognition*, pp.519-521, 1986.

[9] Ohta, Y., Yamamoto, T., and Ikeda, K., "Collinear trinocular stereo using two-level dynamic programming," *Proc. 9th Int. Conf. on Pattern Recognition*, pp.658-662, 1988.

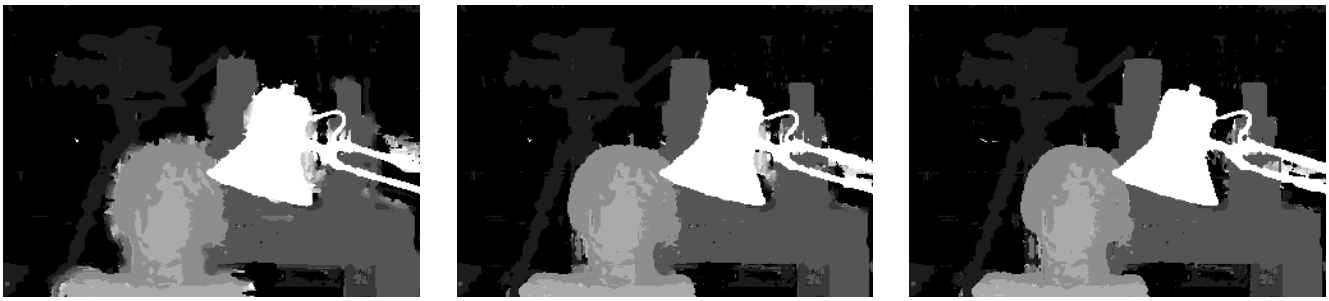
[10] Okutomi, M., and Kanade, T., "A Multiple-Baseline Stereo," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol.15, No.4, pp.353-363, 1993.

[11] Satoh, K., and Ohta, Y., "Passive Depth Acquisition for 3D Image Displays," *IEICE Trans. on Information and Systems*, Vol.E77-D, No.9, pp.949-957, 1994.

[12] Satoh, K., and Ohta, Y., "Occlusion Detectable Stereo Using A Camera Matrix," *Proc. 2nd Asian Conf. on Computer Vision*, 1995 (to be appeared).

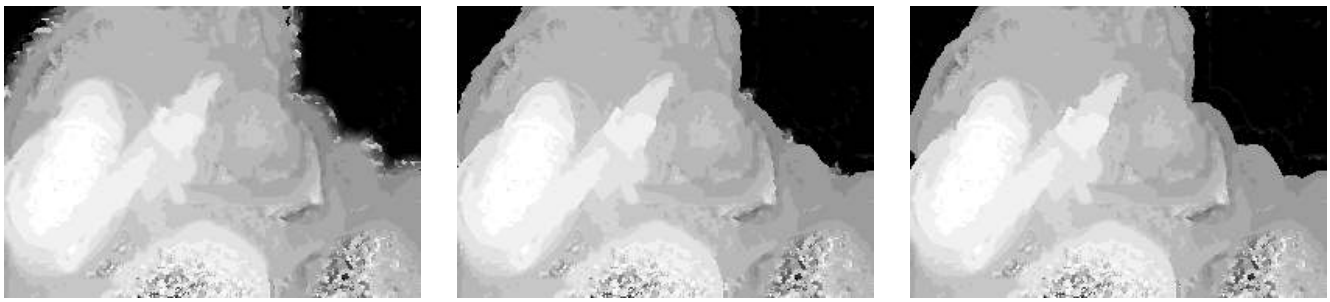
[13] Shah, J., "A Nonlinear Diffusion Model for Discontinuous Disparity and Half-Occlusions in Stereo," *Proc. IEEE Conf. on Computer Vision and Pattern Recognition '93*, pp.34-40, 1993.

(a) Without occlusion mask (3x3) (b) With base mask1 (3x3) (c) With base mask2 (3x3)



(d) Without occlusion mask (5x5) (e) With base mask1 (5x5) (f) With base mask2 (5x5)

Figure 10: Obtained depth map (Scene1)



(a) Without occlusion mask (3x3) (b) With base mask1 (3x3) (c) With base mask2 (3x3)

Figure 11: Obtained depth map (Scene2)

[14] Stewart, C., V., and Dyer, C., R., "The Trinocular General Support Algorithm: A Three-Camera Stereo Algorithm for Overcoming Binocular Matching Errors," *Proc. Second Int. Conf. on Computer Vision*, pp.134-138, 1988.

[15] Tsai, R., Y., "Multiframe Image Point Matching and 3-D Surface Reconstruction," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol.5, No.2, pp.159-174, 1983.

[16] Yachida, M., Kitamura, Y., and Kimachi, M., "Trinocular Vision: New Approach for Correspondence Problem," *Proc. 8th Int. Conf. on Pattern Recognition*, pp.1041-1044, 1986.

[17] Yoshida, K., and Hirose, S., "Real-Time Stereo Vision with Multiple Arrayed Camera," *Proc. 1992 Int. Conf. on Robotics and Automation*, pp.1765-1770, 1992.

[18] Nakahara, T., and Kanade T., "Experiments in Multiple-Baseline Stereo," Technical Report CMU-CS-93-102, School of Computer Science, Carnegie Mellon University, 1993.