*MVA2002* IAPR Workshop on Machine Vision Applications, Dec. 11 - 13, 2002, Nara - ken New Public Hall, Nara, Japan

8—34

# A High-Quality Stereo Algorithm by Integrating MaskSEA and SEA-with-EVL

Takashi Noguchi* Yuichi Ohta*
Institute of Engineering Mechanics and Systems
University of Tsukuba


## Abstract

There are an especially high number of false matching results in the object boundaries obtained by stereo matching. In general, these false results are caused by occlusion. In order to obtain high quality depth information, it is necessary to account for the effects of occlusion. "MaskSEA" and "SEA-with-EVL" were produced as a way of coping with occlusion. In MaskSEA, the quality of the depth information was enhanced by excluding the stereo pairs which were influenced by occlusion. In SEA-with-EVL, however, the quality was enhanced by limiting the illegal large evaluation value which were generated by the occlusion phenomenon. These two methods can be combined. In this paper, we propose a "MaskSEA-with-EVL" with the new evaluation method EVL applied to the usual MaskSEA. The experimental results showed that the quality of the depth information at the boundary area was better than the quality obtained when each technique was used independently.

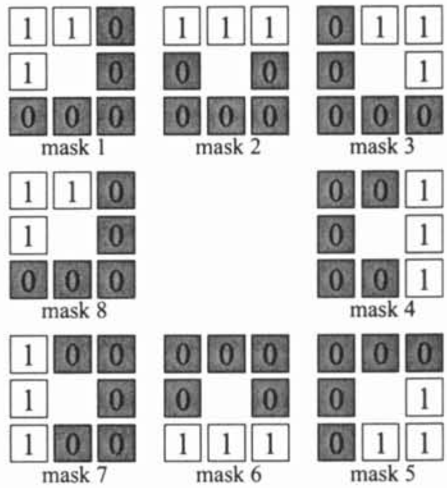




Figure 1: Occlusion Masks

## 1 Introduction

The stereo algorithm is very effective for obtaining depth information of a scene. Window matching is usually used to obtain more stable results. A large window introduces more stable results, but a large window often contains objects with different disparities. The occlusion phenomenon usually occurs in the boundary area generated by objects which have a different depth. When an occlusion area exists in the window, it is easy for false matching to occur. To avoid false matching, SEA (Stereo by Eye Array)[1][2], SEA-with-EVL (SEA with Evaluation Value Limiter)[3], and several window-shift based techniques[4][5] have been developed. SEA avoids false matching by excluding the stereo pair which is affected by occlusion using occlusion masks. We call this method "MaskSEA" in this paper. However, this method can't cope with the "appearance" type of occlusion, so the quality at the boundary area decreases when a large window is used. In addition, all occlusion patterns can't be imitated by the occlusion masks, so false matching increases in the complicated boundary area. However, SEA-with-EVL pays attention to the fact that false matching is caused by the illegal large evaluation value generated by the occlusion phenomenon. This technique increases the quality by rejecting the influences of both types of occlusion, "appearance" and "disappearance" by using the LAD (Limited Absolute Difference), the new evaluation value with the limiter, instead of the AD (Absolute Difference), the conventional evaluation. However, SEA-with-EVL doesn't consider the handling of the reference image that can't observe the target pixel, so false matching occurs often in and around the boundary. Therefore, we propose "MaskSEA-with-EVL", which is a new stereo matching method that combines these two algorithms. This technique can make up for the loss that occurs when each technique is used separately. This new technique has two advantages. The first one is that a higher quality of depth information can be obtained than when using each technique separately. The second advantage is that the increase in false matching can be restrained at the boundary, which the occlusion masks aren't consistent with real occlusion pattern. This technique is very simple, so it can be easily attached to the existent stereo algorithms, and won't disturb the real-time processing of the algorithm.

[1]Adress: 1-1-1 Tennodai, Tsukuba, Ibaraki, 305-8573, JAPAN
E-mail: {noguchi,ohta}@image.esys.tsukuba.ac.jp

## 2 Conventional method used in MaskSEA-with-EVL

### 2.1 MaskSEA

MaskSEA uses nine cameras which are arranged into a matrix-shape. It forms eight stereo pairs in total, with the center image as a base image. This method is called SEA. MaskSEA uses either three or five stereo pairs of the SEA. The combination of the stereo pairs used in stereo matching is shown in Figure 1. The square in this figure shows a reference image, and the reference image shown in white is used for the depth estimation. These are called occlusion masks. In total, eight occlusion masks were defined based on the occlusion pattern at the general boundary. MaskSEA calculated the evaluation value for each occlusion masks, and adopts the minimum value of them. In general, the evaluation values with the reference image that can't observe a

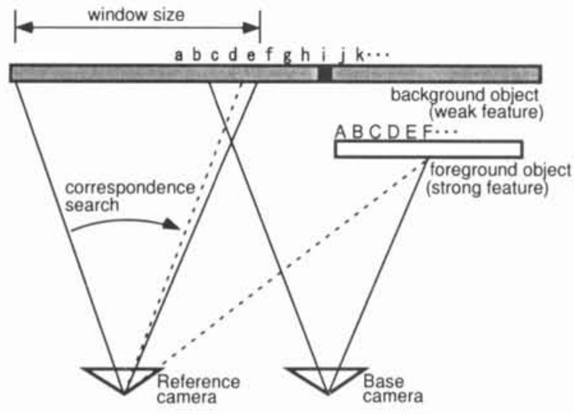



Figure 2: Stereo imaging.

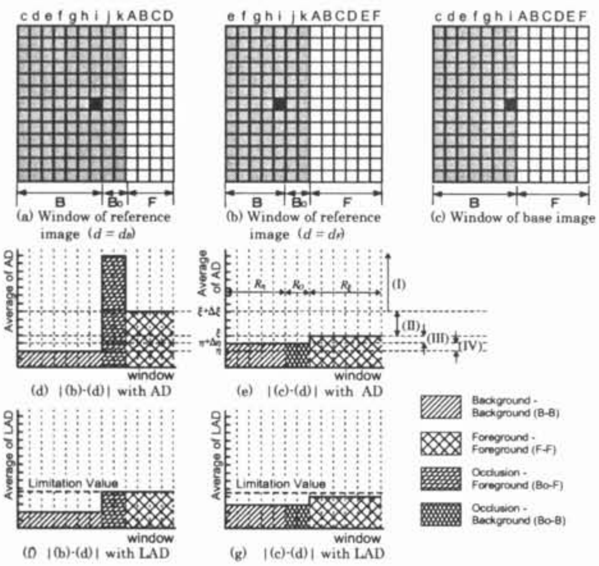



Figure 3: Position of window and their AD and LAD.

target pixel increases illegally large. In other words, this technique copes with the disappearance type of occlusion by choosing the minimum evaluation value, but it can't cope with the appearance type of occlusion, so if the window size is large, false matching increases. In addition, there are a small number of reference images in each mask, so the quality of the non-boundary area is degraded than usual SEA that uses 8 stereo pairs.

### 2.2 SEA-with-EVL

This method uses LAD with an upper limit set against the conventional evaluation value AD. Figures 2 and 3 show the mechanism of false matching and the avoidance of false matching by setting the upper limit. The window shown in figures 3 (a), (b), and (c) could be obtained from the stereo imaging shown in Figure 2. (a) (b) are the reference windows, and (c) is the base window. (a) is the true matching position for the target pixel. (b) is the true matching position for the foreground objects. The window evaluation values such as (d) and (e) could be obtained by the absolute difference calculation. (d) and (e) show the window evaluation value averaged by the column direction. In this case, the evaluation value must records the minimum at position (a), but in general, there are illegally large evaluation values when the occlusion area exists in the window. Therefore, the evaluation value of position (a) doesn't records the minimum, but (b) does. As a result, false matching is shown as the expansion of front objects. To avoid such false matching, the limitation value is set as shown in figures (f) and (g) by a broken line. The limitation value can restrain the illegally large evaluation value in the window, and can avoid false matching by setting a proper limitation value; in this way it is possible to cope with both types of occlusion phenomenon. Because of this, even if a comparatively large window is used, it is possible to obtain higher quality depth information than usual. However, this method doesn't consider the handling of the reference image that can't observe the target pixel, so false matching occurs a great deal in and around the boundary.

## 3 MaskSEA-with-EVL

MaskSEA and SEA-with-EVL can cope with the occlusion phenomenon by different techniques. However, each technique presents a different problem, and the techniques are not efficient. However, the problem presented by one technique can be compensated for by the other technique. Therefore, we propose MaskSEA-with-EVL, a new stereo matching method combining these two algorithms. This method uses the LAD used in SEA-with-EVL instead of the AD used in the conventional MaskSEA. As a result, the weakness of MaskSEA, which is that it can't cope with the appearance type of occlusion, can be compensated for by the SEA-with-EVL technique. The MaskSEA-with-EVL technique can be constructed by the following formulas.

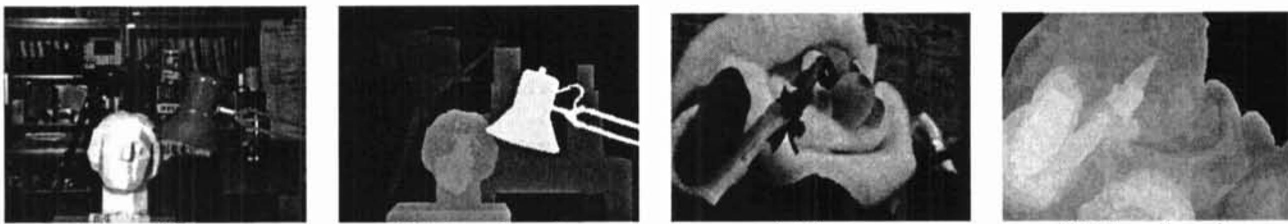
$$Q_{\mathrm{LAD}r}(i,j,d) = min\{|I_r(i,j,d) - I_{\mathrm{B}}(i,j)|, L\} \quad (1)$$

$$Q_{\mathrm{SLAD}r}(i,j,d) = \sum_{p=-N}^{N} \sum_{q=-M}^{M} Q_{\mathrm{LAD}r}(i+p, j+q, d) \quad (2)$$

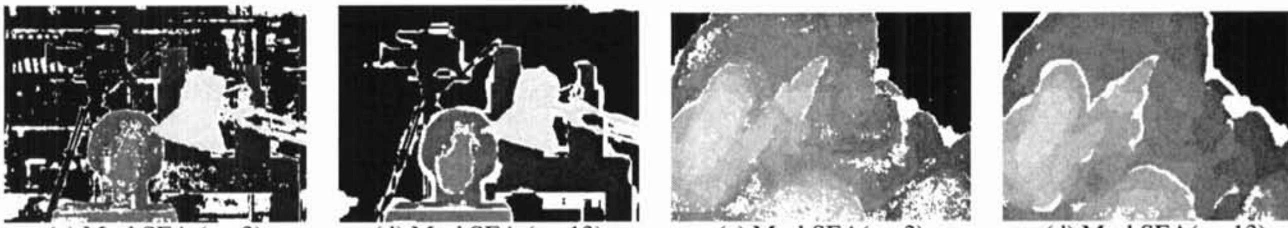
$$Q_{\mathrm{SSLAD}m}(i,j,d) = \sum_{m(r)=1} Q_{\mathrm{SLAD}r}(i,j,d) \quad (3)$$

$$\tilde{d}(i,j) = \underset{d}{arg\,min} Q_{\mathrm{SSLAD}m}(i,j,d) \quad (4)$$

Here, $L$ is the limitation value, $M$ and $N$ are the size of the window, $r$ is the number of reference images, and $m$ is the mask number. Equation (1) was used to calculate the evaluation value of the pixels. If the absolute difference is larger than the limitation value $L$, $L$ is adopted as the evaluation value. Equation (2) was used to calculate the SLAD (Sum of LAD), the evaluation value of the window, obtained by one stereo pair. Equation (3) was used to calculate the SSLAD (Sum of SSLAD). The stereo pairs that were used in the summation were specified by each occlusion mask. In total, eight SSLAD$m$ ($m$=1 to 8) were calculated here. Equation (4) selects $d$, which records the minimum evaluation value as a disparity at (i,j). These equations are similar to those used for the conventional MaskSEA, but the use of the LAD instead of the AD in equation (1) is the critical difference. This technique is expected to have two advantages. The first advantage is that a higher quality of depth information can be obtained with this technique than with each technique separately, because this new technique can cope with the appearance type of occlusion that causes false matching in the conventional MaskSEA. The other advantage is that the increase of false matching can be restrained at the boundary, where the occlusion masks aren't consistent with the real occlusion pattern. In the conventional MaskSEA, basic occlusion patterns are set as occlusion masks based on the general occlusion pattern, as

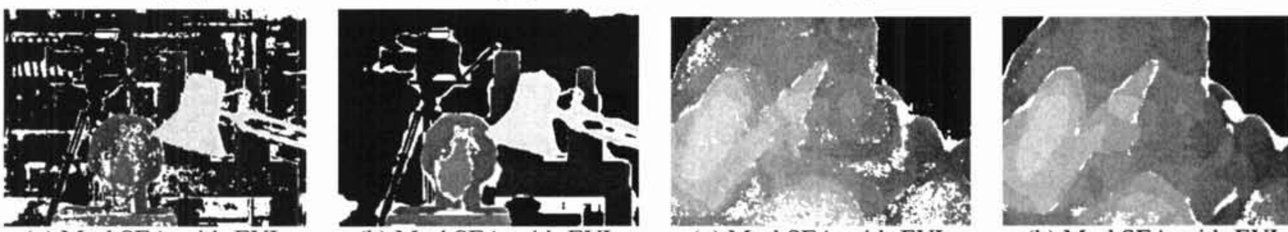


(a) Head & Lamp | (b) True disparity | (a) Santa | (b) True disparity

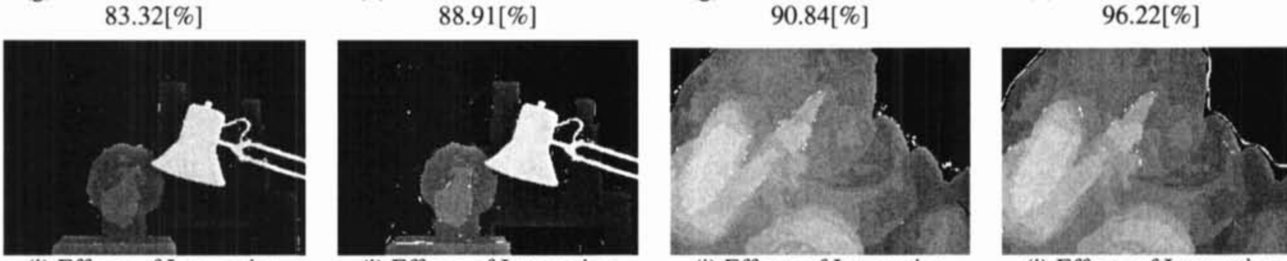


(c) MaskSEA (w=3) 83.30[%] | (d) MaskSEA (w=13) 86.92[%] | (c) MaskSEA(w=3) 90.01[%] | (d) MaskSEA(w=13) 92.87[%]

(e) SEA-with-EVL (w=3) 86.15[%] | (f) SEA-with-EVL (w=13) 88.71[%] | (e) SEA-with-EVL (w=3) 92.41[%] | (f) SEA-with-EVL (w=13) 95.51[%]



(g) MaskSEA-with-EVL 83.32[%] | (h) MaskSEA-with-EVL 88.91[%] | (g) MaskSEA-with-EVL 90.84[%] | (h) MaskSEA-with-EVL 96.22[%]



(i) Effects of Integration | (j) Effects of Integration | (i) Effects of Integration | (j) Effects of Integration

Figure 4: Effectiveness for large window size. (a) base image, (b) true disparity, (c) (d) results by MaskSEA, (e) (f) results by SEA-with-EVL, (g) (h) results by proposed method. In (c) to (h), false matching is shown by white. In (i) and (j), pixels where correct disparities were obtained only by the proposed method are indicated by white.

shown in Figure 1. But the conventional masks can't imitate all types of occlusion patterns, and therefore the occlusion mask that doesn't imitate real occlusion patterns would be used at complicated boundaries. In such cases, a great deal of false matching would occur. The illegal large evaluation value also causes false matching, which would be restrained by the advantages of SEA-with-EVL.

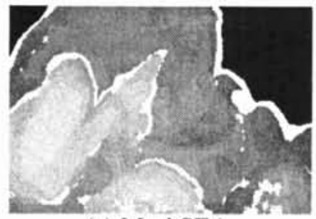
## 4 Experiments

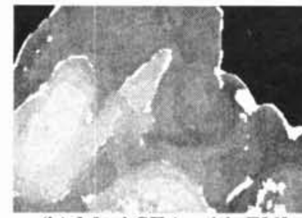
### 4.1 Effectiveness for large window size

Figure 4 shows the results of the experiments. (a) is the base image, and (b) is the true disparity image. (c) and (d) are the results of MaskSEA, (e) and (f) are the results of SEA-with-EVL. (g) and (h) are the results of MaskSEA-with-EVL. Images (c), (e), (g) and (i) are the results obtained by using a 3x3 window, and (d), (f), (h) and (j) are the results obtained by using a 13x13 window. In these images, the pixels with false matching are indicated by white. In these experiments, the limitation value for SEA-with-EVL and MaskSEA-with-EVL were set to 11, based on our experience[3]. The quality at the non-boundary area of (d) was more stable than (c), but it decreased at the boundary area, because of the bad effects of the large window. (e) and (f) show the same phenomenon. The quality of (f) is better than (d), because SEA-with-EVL can cope with both types of occlusion, but MaskSEA can't. However, there are a few false matching areas at the boundary area. In (g) and (h), the


(a) MaskSEA
3pairs 86.67[%]


(b) MaskSEA with EVL
3pairs 88.88[%]

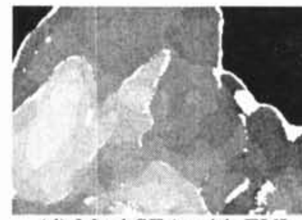

(a) MaskSEA
3pairs 92.91[%]

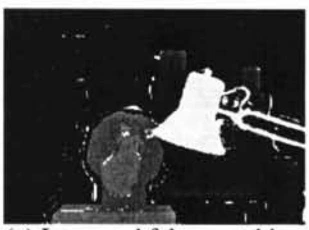

(b) MaskSEA with EVL
3pairs 95.05[%]

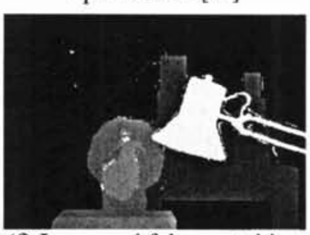

(c) MaskSEA
7pairs 86.19[%]


(d) MaskSEA with EVL
7pairs 88.82[%]


(c) MaskSEA
7pairs 91.49[%]


(d) MaskSEA with EVL
7pairs 95.76[%]


(e) Increased false matching area from (a) to (c)


(f) Increased false matching area from (b) to (d)


(e) Increased false matching area from (a) to (c)

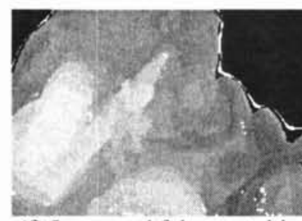

(f) Increased false matching area from (b) to (d)

Figure 5: Robustness for mask pattern inconsistency. (a) (c) results by MaskSEA, (b)(d) results by proposed method. (a) and (b) are obtained by masks with 3 stereo pairs, (c) and (d) are obtained by masks with 7 stereo pairs. (e) and (f) indicate the increased false matching. Increase in (f) is less than that in (e).

results of MaskSEA-with-EVL, those false matching areas at the boundary are greatly reduced. The false matching areas that are reduced by the proposed technique are shown in (i) and (j). The effective area can be clearly confirmed in the form which is extremely near to a portion of the boundary. In other words, this technique can improve the quality of depth estimation.

### 4.2 Robustness for mask pattern inconsistency

Figure 5 (a) and (c) are the results obtained by using MaskSEA, (b) and (d) are the results obtained by using MaskSEA-with-EVL. (a) and (b) are the results obtained by 3 stereo pairs, and (c) and (d) are the results obtained by 7stereo pairs. (e) and (f) show the pixels with false matching caused by the increase of the stereo pairs. In general boundaries, the number of reference images that can observe the target pixel was 3 to 5, and it was rare that the target pixel could be observed by seven reference images. The false matching will increase when reference images that couldn't observe the target pixel were used, as shown in (e) and (f). But the increased false matching area of (f) obtained by the proposed method is smaller than that of (e). This indicates that the increase of false matching could be restrained when an occlusion mask which aren't consistent with real collusion pattern is used. In other words, this method can cope with more complicated boundaries than the usual MaskSEA.

## 5 Conclusion

By integrating MaskSEA and SEA-with-EVL, we were able to use larger window sizes and a larger number of stereo pairs than with the conventional MaskSEA. This integrated technique is effective for improving the matching quality at area without occlusion, and the side effect at occluding boundaries can be suppressed. It is very simple to implement, and the computational costs are almost equivalent to that of the conventional MaskSEA.

## References


[1] K.Satoh and Y. Ohta: "Passive Depth Acquisition for 3D Image Displays", IEICE Trans, E77-D, 9, pp.949-957 (1994).

[2] Y.Nakamura, T.Matsuura, K.Satoh, Y.Ohta: "Occlusion Detectable Stereo", Proc. IEEE CVPR, pp.371-378 (1996).

[3] T.Noguchi and Y.Ohta "A Simple but High-Quality Stereo Algorithm" In Proc. ICPR, pp.351-354 (2002).

[4] H.Tao, H.S. Sawhhney, and R.Kumar. "A global matching framework for stereo computation" In Proc. ICCV, vol.I, pp.532-539 (2001).

[5] S.B.Kang, R.Szeliski and J.Chai "Handling Occlusion in Dense Multi-View Stereo" In Proc. CVPR, vol.I, pp.103-110 (2001).