

# Eliminating Outliers in Motion Occlusion Analysis

Weichuan Yu<sup>1</sup>, Kostas Daniilidis<sup>2</sup>, Gerald Sommer<sup>1</sup>

<sup>1</sup>Institut für Informatik  
Christian-Albrechts-Universität  
Preußerstraße 1–9  
D-24105 Kiel, Germany  
{wy,gs}@ks.informatik.uni-kiel.de

<sup>2</sup>GRASP Laboratory  
University of Pennsylvania  
3401 Walnut Street  
Philadelphia, PA 19104-6228, USA  
kostas@grip.cis.upenn.edu

**Abstract.** *Occlusion boundaries are considered either as outliers or as noise in most optical flow algorithms. In order to treat the boundary problem, many probabilistic algorithms like maximum likelihood [6] or expectation-maximization (EM) [16, 3] are proposed to gradually decrease the weights of pixels in boundary regions during estimation iterations. However, these approaches still include the outliers in the estimation. If the number of pixels in boundary regions is comparable to the number of pixels with single motion, we will not be able to robustly estimate the motion parameters since probabilistic methods are purely based on statistics.*

*In this paper, we propose to mark the outliers **directly** using a method based on eigenvalue analysis [8]. Then we eliminate these outliers in the multiple motion estimation. Comparisons show that this method can improve the precision of estimation results. We use also the “warp-and-subtract” technique to localize and to track occlusion boundaries. The closest work has been done by Fleet et al. [2] as well as by Yu et al. [1]. These are the only approaches with an explicit model of occlusion which, however, is not sufficient to deal with outliers.*

## 1 Introduction

In the computation of optical flow the detection and tracking of occlusion boundaries are challenging problems. At occlusion boundaries, the single motion assumption and the smoothness assumption are violated. Since most optical flow algorithms are based on these two assumptions (e.g. the well known brightness change constraint equation), they can neither provide correct estimation results in boundary regions nor track the movement of boundaries.

The boundary problem was first addressed by Nagel and Enkelmann [10]. In order to estimate motion parameters robustly, they introduced a spatial regularization term to penalize motion discontinuities. Weickert and Schnörr further extended this regularization term into spatio-temporal space [15]. Black and Anandan [3] treated occlusion regions similarly. They referred to the pixels near occlusion boundaries as *outliers* of the motion constraint and set lower weights

to these pixels in the estimation. The concept of *outlier* comes from statistics. It means a small amount of data points with large deviation from the bulk of all data points. This concept represents exactly the relationship between the pixels near occlusion boundaries and the pixels with a single motion, since the spatio-temporal partial derivatives of the pixels with a single motion form a plane in the derivative space coordinated with  $(I_x, I_y, I_t)$  and the derivatives of the pixels near occlusion boundaries deviate from this plane due to motion discontinuities. Based on this concept, many probabilistic methods were proposed to model occlusion boundaries [9] and to estimate multiple motions near occlusion boundaries [6, 7, 16].

Schunck considered occlusion boundaries as noise in the constraint line clustering [12]. By noticing that motions have more components than noise, he applied a statistic method to cluster the dominant intersection of constraint lines for motion estimation. This statistic method was used in Hough transform based approaches as well ([5, 11]).

In above approaches the occlusion boundaries are modelled implicitly. There are also explicit models of occlusion boundaries in the frequency domain [1] and in the spatial domain [2, 4]. For example, in [4] an occlusion boundary in a circular mask is modelled with six parameters, i.e. four motion parameters of both occluding and occluded signals, the orientation of this boundary, and the distance between the boundary and the center of the circular mask. With this explicit model they wish to predict the locations of occlusion boundaries in the next frame exactly and therefore exclude the corresponding boundary regions in the next estimation. Moreover, by tracking the movement of boundaries they can further solve the foreground/background ambiguity [4].

However, these approaches still include the outliers in the estimation. This makes the estimation fragile, especially if the number of outliers is comparable to the number of pixels with a single motion, since probabilistic methods are purely based on statistics.

Our motivation is to improve the quality of input data before extracting motion parameters. According to our observation this is possible by combining current techniques.

This paper is constructed as follows: In section 2 we introduce the outlier detection method. In section 3 we compare motion estimation results before and after eliminating outliers and apply the “warp-and-subtract” technique to localize and to track occlusion boundaries. We show experimental results in section 4. Then we conclude the paper with some discussions.

## 2 Detection of Outliers

We assume that the motions in image sequences are piecewise-smooth with possible occlusion. In the spatio-temporal derivative space coordinated with  $(I_x, I_y, I_t)$  we observe the following distributions [8]

- For a single constant translational motion, we have a plane whose normal vector is parallel to  $(u, v, 1)$ , where  $(u, v)$  denotes the optical flow vector.

The eigenvalues of this plane satisfy

$$\sigma_1 \geq \sigma_2 > \sigma_3 = 0. \quad (1)$$

- For a single constant motion having aperture problem, the plane above degenerates into a line whose corresponding eigenvalues satisfy

$$\sigma_1 > \sigma_2 = \sigma_3 = 0. \quad (2)$$

- For occlusion we observe multiple planes plus distortions [1] with three positive eigenvalues

$$\sigma_1 \geq \sigma_2 \geq \sigma_3 > 0. \quad (3)$$

Thus, we can judge if there are multiple motions from different combinations of eigenvalues, even without knowing motion parameters. In case of occlusion, if we can purify multiple planes from distortions (i.e. outliers), we may improve the precision of estimation results. The remaining question is how to detect these outliers. We observe that if we have occlusion in a window, the occlusion boundaries should locate in this window as well, though we do not know their exact positions. Based on this observation we use a multi-window strategy to eliminate outliers before estimation. We detect occlusion regions using eigenvalue analysis with small windows and mark these regions as outliers. In a large window containing these small windows, the pixels outside outlier regions are guaranteed to be “normal” pixels. Using only these “normal” pixels for estimation we avoid the disturbance of outliers and improve therefore the precision of estimation results in the large window.

It should be noticed that we abandon also some “normal” pixels by marking outliers with small windows. Therefore, we prefer to reduce the size of the small window so that this loss is as small as possible. On the other side, in order to provide robust eigenvalue analysis, we must have adequate number of pixels in the small window. In order to solve this conflict, we limit the spatial size of the small window, but extend its temporal size to include pixels from other frames (e.g. from frame  $(t_0 - 1)$  and  $(t_0 + 1)$ , where  $t_0$  denotes the current frame).

In the practice the eigenvalues may deviate from their standard values due to noise or derivative approximation error. Therefore, instead of checking if  $\sigma_3 = 0$ , we set a threshold  $\lambda_{31}$  for outlier detection. If  $\sigma_3 > \lambda_{31}\sigma_1$ , we conclude that there are multiple motions. In addition, we may check the aperture problem by defining another threshold  $\lambda_{21}$ . In this paper we set  $\lambda_{31} = \lambda_{21} = 0.2$ . The results of detections are shown in figure 1-3.

### 3 Estimation of Multiple Motions and Tracking Motion Boundaries

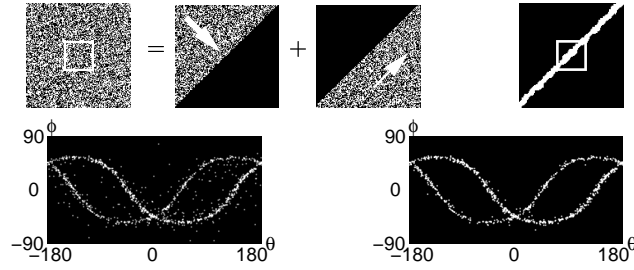
Before applying the EM algorithm [16] for motion estimation, we must verify that there are still sufficient pixels remaining. We define a reliable measure which

is a ratio between the number of pixels remaining and the total number of pixels in the window

$$r = \frac{\mathcal{N}_i}{\mathcal{N}_{all}}, \quad (i = 1, 2) \quad (4)$$

where  $\mathcal{N}_1/\mathcal{N}_2$  denotes the number of remaining pixels of the occluding/occluded signal. If either of these two ratios is below a threshold, we have to enlarge the window to include more pixels for estimation.

The precision improvement of estimation results after eliminating outliers is shown in figure 1 and table 1. The occluding signal moves with a speed of (1,1) pixel/frame and the occluded signal with a speed of (1,-1) pixel/frame. For the clarity of displaying we project the 3D data onto the orientation space with variables  $\theta$  and  $\phi$ , where  $\theta$  and  $\phi$  are horizontal and vertical angles in the spherical coordinates. We can see that after eliminating outliers the curves in the  $(\theta, \phi)$  space are more clearly. Consequently, we obtain better estimation results (see table 1). In order to analyze the effect of window size in the estimation, we reduce the window size from  $33 \times 33$  to  $17 \times 17$ . In the  $17 \times 17$  window, the number of outliers is easier to be comparable to the number of “normal” pixels. As a result, the disturbance of outliers increases strongly. In contrast, if we eliminate outliers before estimation we can still obtain reasonable results.



**Fig. 1. Up Left:** One frame from a random dot occlusion sequence. The white box shows us the window across the occlusion boundary and the white arrows show us moving directions. The thick arrow in the second image denotes the occluding signal and the thin arrow in the third image denotes the occluded signal. **Up Right:** Marked outliers after eigenvalue analysis using a  $5 \times 5 \times 3$  window. For clarity we show the white box here again. **Down Left:** Spherical representation of 3D data in the  $(I_x, I_y, I_t)$  space before eliminating outliers. **Down Right:** Spherical representation after eliminating outliers. Two curves are more clearly to see. See table 1 for estimation results.

After obtaining multiple motion parameters in the boundary regions we further localize occlusion boundaries in one frame and track the movement of boundaries using the “warp-and-subtract” technique. This technique is based on the observation of spatial coherence of the image sequence. The reader is referred to [7, 16, 1] for details about this technique due to the space limitation. The results are shown in figures 2 and 3.

window size		occluding speed	occluded speed
33 × 33	before	(0.9895, 1.0009)	(0.9803, -0.9840)
	after	(0.9988, 0.9997)	(0.9886, -0.9940)
17 × 17	before	(0.8801, 0.9710)	(0.8589, -0.8685)
	after	(0.9876, 1.0132)	(0.9932, -0.9980)

**Table 1.** Estimation results before and after eliminating outliers. For comparison we apply the EM algorithm with same parameters and same initial values before and after eliminating outliers.

## 4 Experiments

In this section we show some real examples. Figure 2 shows the estimation results of an occlusion sequence. For performance comparison we apply the EM algorithm vertically along the vertical occlusion boundary. We do not have the exact ground truth, but we observe that there is almost no depth difference among pixels on each side of the vertical boundary. Therefore, we use the estimation results with a larger window as ground truth. We observe the improvement after eliminating outliers clearly.

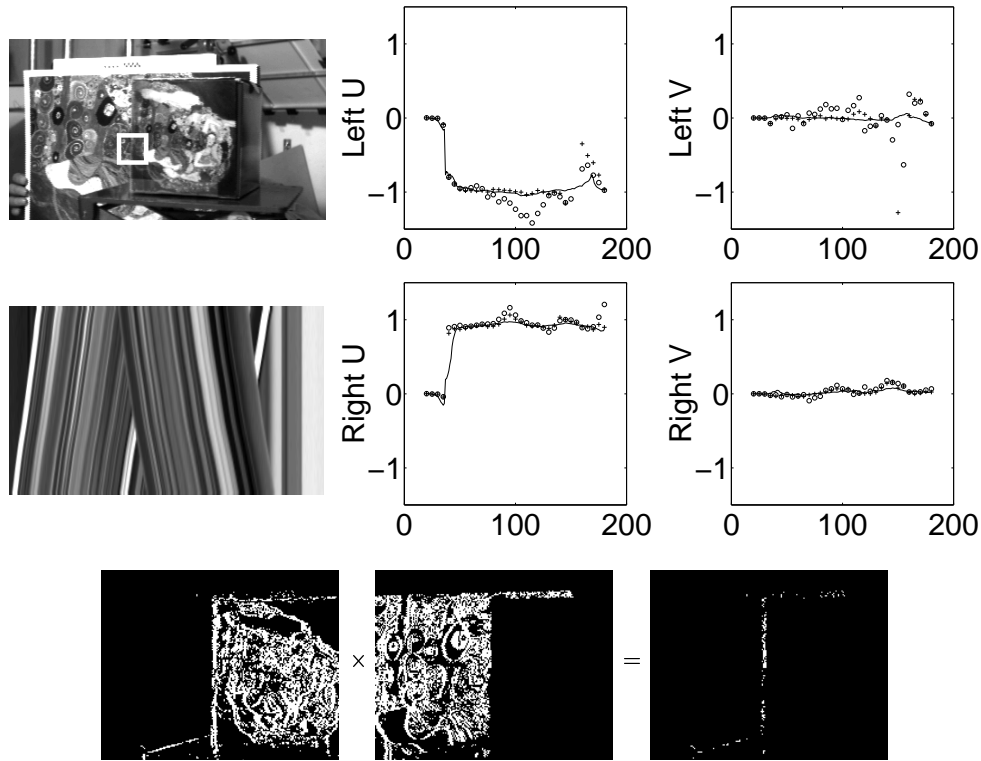
Figure 3 shows us another real sequence with ground truth. This block world sequence is very difficult for the constant motion model used here because the real motions are affine and the occlusion regions have also the aperture problem. But since we know the ground truth, we can still compare the performances of the EM algorithms before and after eliminating outliers.

## 5 Conclusion

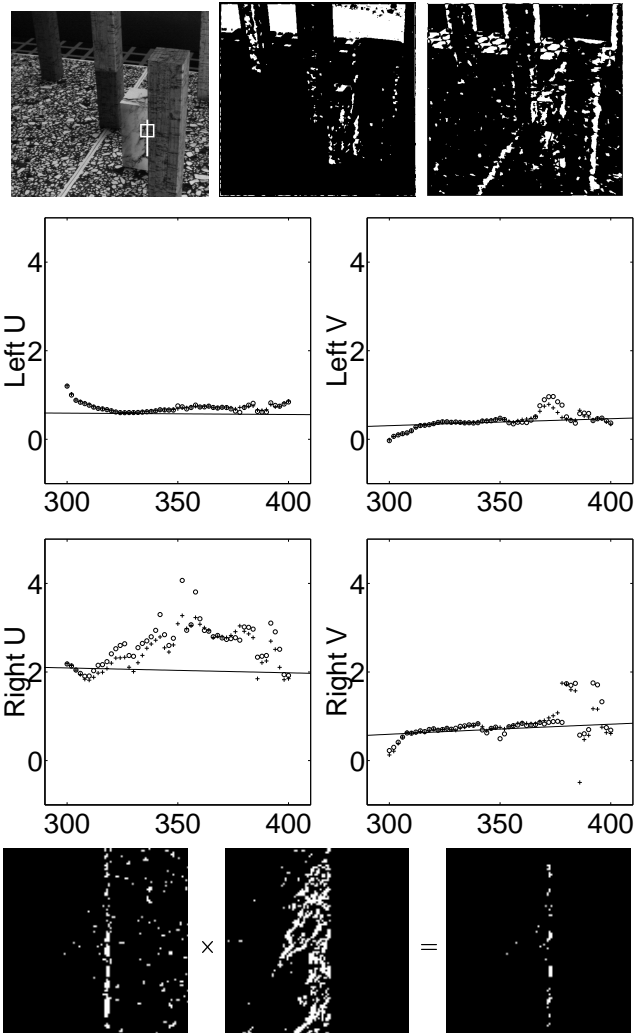
In this paper we proposed to eliminate outliers in multiple motion estimation. Comparing with current probabilistic approaches, which include the outliers in the estimation, our method improves the *quality* of input data and therefore provides more exact results. Moreover, observing the spatial coherence on each side of occlusion boundaries we applied the “warp-and-subtract” technique to localize and to track occlusion boundaries. We do not use an explicit local model of the boundary region. But we can still obtain the desired information about the occlusion boundaries after localizing them.

The techniques in our algorithm have already been used in previous related works. The meaning of our work is that we propose a multi-window strategy for occlusion analysis. This strategy is very simple and it works well.

Recently, Shi and Malik proposed to segment images without motion estimation. They introduced a concept of normalized cut and minimized it for segmentation. This normalized cut is a connection measure between one pixel and its neighbors with respect to brightness, color, texture [13] or even motion correlation information [14]. This approach is based on the same observation in the “warp-and-subtract” technique. This fact reminds us again that the spatial coherence information is very useful for image segmentation.



**Fig. 2. Row 1 left:** One frame of an occlusion sequence with  $200 \times 350$  pixels. In this sequence a right moving box occludes a left moving picture. The white box centered at  $(122, 137)$  contains the vertical occlusion boundary. **Row 2 Left:** The slice of the sequence along row 122. The first frame is at the top of the slice. **Rows 1 and 2 Middle and Right:** Estimation results along column 137 using a  $15 \times 15$  window. We use the results with a  $31 \times 31$  window as ground truth and draw them with solid lines. We draw the results before eliminating outliers with circles and draw the results after eliminating outliers with crosses. For comparison we display different speed components separately. For clarity of displaying we sample the results with an interval of 5 pixels along column 137. In the window centered at  $(160, 137)$  the results are not reasonable, since there are only *four* pixels of the occluded signal remaining after eliminating outliers. This example demonstrates the necessity of introducing reliable measure (equation (4)). **Row 3:** The segmentation result after “warp-and-subtract”. For clarity we enlarge the region containing occlusion boundaries. After each warping we observe one region with zero intensity. In the right image we see the localized boundaries. The “warp-and-subtract” technique works also for boundaries with complex contours like the corner of the box. We may further track the movement of boundaries to solve the foreground/background ambiguity [4], since occlusion boundaries move consistently with the occluding signal.



**Fig. 3. Row 1 Left:** One frame from the block world sequence with  $512 \times 512$  pixels. The white box shows us the window across the vertical boundary and the white line shows the column along which we apply the EM algorithms for comparison. **Row 1 Middle:** Marked outliers. **Row 1 Right:** Regions with the aperture problem. **Row 2 and 3:** Estimation results before and after eliminating outliers v.s. row index (from row 300 to row 400) using a  $31 \times 31$  window. We draw the ground truth with solid lines, the results before eliminating with circles, and the results after eliminating with crosses. In fact, the constant motion model can not treat such a difficult sequence due to complicated motions and the aperture problem. But we can still see that the errors are generally reduced after eliminating outliers, specially for the occluding signal at the right side of the boundary. **Row 4:** We further test the “warp-and-subtract” technique with the ground truth. For clarity we enlarge the boundary region from row 300 to row 400.

## References

1. W. Yu, K. Daniilidis, S. Beauchemin, and G. Sommer. Detection and characterization of multiple motion points. In *IEEE Conf. Computer Vision and Pattern Recognition*, volume I, pages 171–177, Fort Collins, CO, June 23–25, 1999.
2. D.J. Fleet, M.J. Black, and A.D. Jepson. Motion feature detection using steerable flow fields. In *IEEE Conf. Computer Vision and Pattern Recognition*, pages 274–281, Santa Barbara, CA, June 23–25, 1998.
3. M. J. Black and P. Anandan. The robust estimation of multiple motions: parametric and piecewise-smooth flow fields. *Computer Vision and Image Understanding*, 63(1):75–104, 1996.
4. M.J. Black and D.J. Fleet. Probabilistic detection and tracking of motion discontinuities. In *Proc. Int. Conf. on Computer Vision*, volume I, pages 551–558, Kerkyra, Greece, Sep. 20–27, 1999.
5. M. Bober and J. Kittler. Estimation of complex multimodal motion: an approach based on robust statistics and hough transform. *Image and Vision Computing*, 12(10):661–668, 1994.
6. P. Bouthemy. A maximum likelihood framework for determining moving edges. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 11:499–511, 1989.
7. J. R. Bergen, P. J. Burt, R. Hingorani, and S. Peleg. A three-frame algorithm for estimating two-component image motion. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 14(9):886–895, 1992.
8. B. Jaehne. *Spatio-Temporal Image Processing*. Springer-Verlag, 1993.
9. A. Jepson and M. J. Black. Mixture models for optical flow computation. In *IEEE Conf. Computer Vision and Pattern Recognition*, pages 760–761, New York, NY, June 15–17, 1993.
10. H.H. Nagel and W. Enkelmann. An investigation of smoothness constraints for the estimation of displacement vector fields from image sequences. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 8:565–593, 1986.
11. J. M. Nash, J. N. Carter, M. S. Nixon. Dynamic feature extraction via the velocity hough transform. *Pattern Recognition Letters*, 18:1035–1047, 1997.
12. B. G. Schunck. Image flow segmentation and estimation by constraint line clustering. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 11(10):1010–1027, 1989.
13. J. Shi and J. Malik. Normalized cuts and image segmentation. In *IEEE Conf. Computer Vision and Pattern Recognition*, 1997.
14. J. Shi and J. Malik. Motion segmentation using normalized cuts. In *Proc. Int. Conf. on Computer Vision*, pages 1154–1160, Bombay, India, Jan. 4–7, 1998.
15. J. Weickert and C. Schnörr. Räumlich-zeitliche Berechnung des optischen Flusses mit nichtlinearen flußabhängigen Glattheitstermen. In *DAGM Symposium Mustererkennung*, pages 317–324, Bonn, Germany, Sep. 15 - 17, 1999.
16. Y. Weiss and E. H. Adelson. A unified mixture framework for motion segmentation: Incorporating spatial coherence and estimating the number of models. In *IEEE Conf. Computer Vision and Pattern Recognition*, pages 321–326, San Francisco, CA, June 18–20, 1996.