Detection and Characterization of Multiple Motion Points

Weichuan Yu[1] Kostas Daniilidis[2] Steven Beauchemin[2] Gerald Sommer[1]
[1]Institute of Computer Science
Christian Albrechts University
Preusserstrasse 1-9
D-24105 Kiel
Germany
[2]GRASP Laboratory
University of Pennsylvania
3401 Walnut Street
Philadelphia, PA 19104-6228
USA
Email: wy@ks.informatik.uni-kiel.de

Abstract

The computation of optical flow is a well studied topic in biological and computational vision. However, the existence of multiple motions in dynamic imagery due to occlusion or even transparency still raises challenging questions. In this paper, we propose an approach for the detection and characterization of occlusion and transparency. We propose a theoretical framework for both types of multiple motions which explicitly shows the difference between occlusion and transparency in the frequency domain. Then, we employ an EM-algorithm for the computation of one or two image velocities and a simple test for the detection of occlusion. Our approach differs from other EM-approaches which blindly assume the superposition of two models in the spatial domain without providing with a separate formal model for occlusion. We test and compare the characterization performance in synthetic and real data.

1 Introduction

It was early recognized in the history of optical flow estimation that motion boundaries necessitate a special treatment. Due to the aperture problem, flow computation apply assumptions on the behavior of the optical flow in the neighborhood of the considered point. Such assumptions are either explicit in area-based techniques or implicit in filter-based schemes where the addressed neighborhood is the filter support. In global approaches of optical flow the assumptions are encoded in the regularization term. Regularization approaches were also the first in addressing the problem of motion boundaries by penalizing flow smoothness at hypothetical positions of motion boundaries [15]. We will not delve into this group of flow segmentation algorithms which later followed the paradigm of anisotropic diffusion [17]. A second group of approaches is based on segmenting regions where parametric models of flow can be fitted [18, 7, 12, 23]. Due to space limitations the reader is referred to [4] for a survey of flow estimation methods.

We are interested here in local techniques for the detection of points involving multiple motions and their classification in occlusion and transparency points. Most of the local approaches consider this problem as the fitting of superimposed models or as the violation of a parametric model. Approaches computing the local spatiotemporal grayvalue tensor [19, 16] test the tensor's eigenvalue to detect points of multiple motions whereas Black and Anandan [6] apply robust estimation in order to detect such points. The majority of these approaches relies on the Brightness Change Constraint Equation

$$I_x u + I_y v + I_t = 0 \qquad (1)$$

where I_x, I_y, and I_t denote the spatiotemporal partial derivatives of the image intensity and (u,v) is the optical flow vector. Looking at the case of multiple motions as a simultaneous problem of estimation and grouping researchers have recently elaborated algorithms based on the Expectation-Maximization Principle [1]. Whereas the Maximization step is the usual maximum-likelihood parameter estimation given the assignment of points to groups, the Expectation step is regrouping the points by updating membership weights. Several authors [13, 22] applied the EM-algorithm on the BCCE model (1) or on already computed flow vectors. This process is equivalent to fitting two planes through the origin in the (I_x, I_y, I_t)-space. However, this fitting does not reveal what happens in reality. In case of occlusion, it can be shown that the (I_x, I_y, I_t)-space does not contain only two planes. In case of transparency we can hardly assume that the intensity profile is differentiable. A recent approach by Fleet et al. [5] gives the best explicit model of occlusion for the spatial domain by applying the steerability

theory in the detection of the occlusion boundary as a step edge in both components of the optical flow field.

The underlying theoretical framework for the approach in this paper relies on spectral analysis and was first presented in [3] based on observations in [9]. Another framework for multiple motions formulated in the frequency domain is the superposition principle of Shizawa and Mase [21] which, however, does not discriminate between occlusion and transparency. We start with an illustration of the difference between occlusion and transparency in Fig. 1. First, by isolating a spatial window we observe that occlusion is more local than transparency. In case of transparency the entire window contains two motions. Second, the transparency is the addition of two models whereas the occlusion involves a step-function.

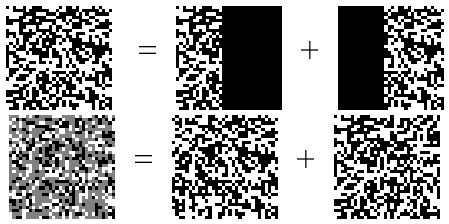

Figure 1: Although occlusion or transparency can be decomposed into multiple layers they are based on different decomposition principle. **Top:** One frame of the occlusion sequence is decomposed into two layers by a Heaviside unit step function. There is motion discontinuity only at the boundary. **Bottom:** One frame of the transparency sequence is a simple superposition of two layers. Multiple motions exist in the entire window.

In this paper, we give an exact model of occlusion and transparency in the frequency model. The only existing counterpart in the literature of spatial approaches is [5]. We gain an insight into the distortion due to occlusion (Sec. 2). We employ an EM-algorithm for the recovery of the two dominant motions (Sec. 3). Then, we apply a distortion existence test in the frequency domain which we verify with an additional test in the spatial domain (Sec. 4). We show the performance of the algorithm in synthetic and real sequences and compare it with the performance of a purely spatial EM-algorithm (Sec. 5).

2 Spectrum of Multiple Motions

The spectrum of multiple motions was first analyzed by Fleet and Langley [9]. Assuming that the occlusion boundary is a function $\chi(\vec{x})$ they model the occlusion in the spatial domain as follows:

$$I(\vec{x},t) = \chi(\vec{x}-\vec{v}_1 t) I_1(\vec{x}-\vec{v}_1 t) + [1-\chi(\vec{x}-\vec{v}_1 t)] I_2(\vec{x}-\vec{v}_2 t) \tag{2}$$

where $I_1(\vec{x})$ is a 2D **occluding** signal moving with velocity $\vec{v_1} = (v_{1x}, v_{1y})^T$ and $I_2(\vec{x})$ is a 2D **occluded** signal moving with velocity $\vec{v_2} = (v_{2x}, v_{2y})^T$.

Beauchemin and Barron [3] were the first who extracted an exact model in the frequency domain. They model the occlusion in the spatial domain with a Heaviside unit step function $U(\vec{x})$ for $\chi(\vec{x})$:

$$U(\vec{x}) = \begin{cases} 1 & \vec{x}^T \hat{\eta} \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

where $\vec{x}$ denotes 2D spatial Cartesian coordinates and $\hat{\eta}$ is a unit vector parallel to the gradient of the occluding boundary.

We denote the spatial frequency vector as $\vec{\kappa} = (\omega_x, \omega_y)^T$ and the temporal frequency as ω_t. Then, the Fourier transform of the image sequence reads

$$\begin{aligned} \tilde{I}(\vec{\kappa},\omega_t) = \tilde{U}(\vec{\kappa})\delta(\vec{\kappa}^T\vec{v}_1+\omega_t) * \tilde{I}_1(\vec{\kappa})\delta(\vec{\kappa}^T\vec{v}_1+\omega_t) \\ + \tilde{I}_2(\vec{\kappa})\delta(\vec{\kappa}^T\vec{v}_2+\omega_t) \\ - \tilde{U}(\vec{\kappa})\delta(\vec{\kappa}^T\vec{v}_1+\omega_t) * \tilde{I}_2(\vec{\kappa})\delta(\vec{\kappa}^T\vec{v}_2+\omega_t) \end{aligned} \tag{4}$$

where $*$ means convolution and $\tilde{}$ denotes the Fourier transform of the corresponding signal. The spectrum of a 2D step function is given by

$$\tilde{U}(\vec{\kappa}) = 2\pi[\pi\delta(|\vec{\kappa}|) + \frac{\delta(\vec{\kappa}^T\hat{\eta}_\perp)}{i\vec{\kappa}^T\hat{\eta}}] \tag{5}$$

where $\hat{\eta}_\perp$ denotes a unit vector perpendicular to the normal $\hat{\eta}$ of the occluding boundary. Taking the properties of the impulse function into account we obtain (see Appendix for detail):

$$\begin{aligned} \tilde{I}(\vec{\kappa},\omega_t) = [2\pi^2\tilde{I}_1(\vec{\kappa}) + A(\vec{\kappa})]\delta(\vec{\kappa}^T\vec{v}_1+\omega_t) \\ + (1-2\pi^2)\tilde{I}_2(\vec{\kappa})\delta(\vec{\kappa}^T\vec{v}_2+\omega_t) \\ + \quad B(\vec{\kappa}) \end{aligned} \tag{6}$$

with

$$A(\vec{\kappa}) = \frac{2\pi}{i\vec{\kappa}^T\hat{\eta}} * \tilde{I}_1(\vec{\kappa}) \tag{7}$$

$$B(\vec{\kappa}) = \frac{2\pi}{i\vec{\kappa}^T\hat{\eta}} * \tilde{I}_2(\vec{\kappa}) \tag{8}$$

The first two terms of expression (6) are two oriented planes passing through the origin of the frequency space. Their normal vectors namely $(u_1, v_1, 1)$ and $(u_2, v_2, 1)$ are the velocities of the two signals. The second is the exact spectrum of the occluded signal but the first contains an additional distortion term $A(\vec{\kappa})$

on the plane of the occluding spectrum. However, we are interested here in the orientation of the plane and the term $A(\vec{\kappa})$ does not disturb the orientation. Actually, $A(\vec{\kappa})$ strengthens this spectral plane. Therefore, we do not consider it as distortion. The main discriminating term is the third one $B(\vec{\kappa})$, which lies outside of the two motion planes. We observe that this distortion term due to the occlusion is independent of the velocities. It depends on the normal of the occluding boundary and the spectrum of the occluded signal.

If the energy of the distortion term $B(\vec{\kappa})$ is very high we are not able to recognize the two planes. The critical factor in the amplitude of $B(\vec{\kappa})$ is the hyperbolic term $\frac{2\pi}{i\vec{\kappa}^T\vec{\eta}}$. If $\vec{\kappa}$ tends to zero, the amplitude of $B(\vec{\kappa})$ will be larger than that of the second term $(1-2\pi^2)\tilde{I}_2(\vec{\kappa})\delta(\vec{\kappa}^T\vec{v}_2+\omega_t)$ and we will not be able to estimate the parameters of the occluded signal. Fortunately, $\frac{2\pi}{i\vec{\kappa}^T\vec{\eta}}$ reduces very quickly with the increase of $|\vec{\kappa}|$. In most regions of the spectral domain the amplitude of the distortion is much less than that of signals, as shown in Fig. 2. Therefore, we may consider only the spectrum above a lower bound for the frequency and identify the two dominant planes.

Transparency is a special case of occlusion since we can simply substitute $\chi(\vec{x}-\vec{v}_1 t)$ with a real constant $\Phi \in (0,\ 1)$. The corresponding spectrum is then characterized by two oriented planes without any distortion (see Fig. 2):

$$\begin{aligned}\tilde{I}(\vec{\kappa},\omega_t) &= \Phi\tilde{I}_1(\vec{\kappa})\delta(\vec{\kappa}^T\vec{v}_1+\omega_t) \\ &\quad +(1-\Phi)\tilde{I}_2(\vec{\kappa})\delta(\vec{\kappa}^T\vec{v}_2+\omega_t) \qquad (9)\end{aligned}$$

We have provided with an exact model for occlusion in the frequency domain and described the terms that discriminate occlusion from transparency. The occlusion distortion term is independent of the velocities and depends only on the spatial orientation of the motion boundary and the spectrum of the occluded signal. Due to the hyperbolic nature of this term the main energy proportion is on the two motion planes even in the case of occlusion. This model can be viewed as a generalization of the spatiotemporal energy model of single motion [2, 11]. The remaining task is to estimate the parameters of the motion planes and decide whether it is the case of an occlusion or not.

3 Estimation of Multiple Velocities

We are going to apply the EM-algorithm in the frequency domain. In order to avoid the block effect of the discrete Fourier transform we perform a Fourier-Transform windowed first by a Gaussian function which results in nothing else than a Gabor function with decoupled support and central frequency.

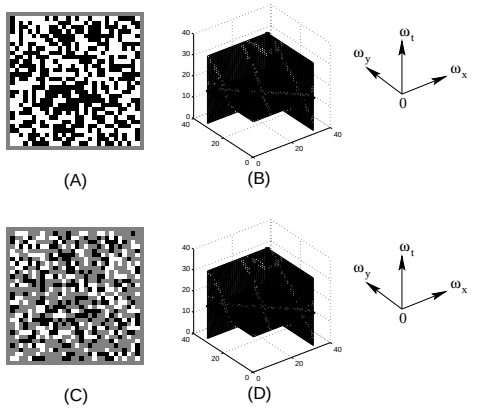

Figure 2: The spectra of occlusion and transparency have similar properties. **(A):** A frame of the random dot occlusion sequence. The occluding signal is moving with the speed $(1,1)$ and the speed of the occluded signal is $(2,-1)$. **(B):** The energy 3D-volume of the spectrum of occlusion. The origin lies in the middle of the drawing. Compared to the dominant planes the distortion is almost negligible. **(C):** A frame of the random dot transparency sequence. The motion parameters are the same as those in (A). **(D):** The energy spectrum of transparency. The high frequency artifact at one side of the spectrum is due to the periodic property of the DFT.

In order to alleviate the effects of the hyperbolic distortion term at low frequencies we multiply the energy spectrum with a 3D low-stop function:

$$LS(\vec{\kappa},\omega_t) = \frac{1}{\alpha+G(0,\frac{\pi}{16},\vec{\kappa},\omega_t)} - \frac{1}{\alpha+G(0,\frac{\pi}{16},0,0,0)}$$

where $G(0,\frac{\pi}{16},\vec{\kappa},\omega_t)$ denotes a 3D Gaussian function in the spectral domain with mean value of zero and variance of $\frac{\pi}{16}$. The parameter α is set to 0.1. We may reduce α to further amplify high frequency components.

We obtain, thus, a set of data points in the spectral domain associated with the amplitude of the Fourier-Transform. Their amplitudes can be viewed as a mass density in fitting the plane to these points. We denote with A_i the amplitude of the i-th point in the data set. We assume the existence of two unknown motions (u_1,v_1) and (u_2,v_2). In case of one motion, we will observe that the EM algorithm converges to one solution.

We choose arbitrary initial values (u_{10},v_{10}) and (u_{20},v_{20}). In the Expectation-step we assign the weights W_{i1} and W_{i2} to the i-th point as following

according to the corresponding squares of the residuals [14]:

$$W_{i1} = \frac{1}{1 + e^{\perp \frac{R_{i2} - R_{i1}}{\sigma^2}}} \tag{10}$$

$$W_{i2} = \frac{1}{1 + e^{\perp \frac{R_{i1} - R_{i2}}{\sigma^2}}} \tag{11}$$

where

$$R_{i1} = A_i^2(\omega_{ix}u_1 + \omega_{iy}v_1 + \omega_{it})^2 \tag{12}$$

$$R_{i2} = A_i^2(\omega_{ix}u_2 + \omega_{iy}v_2 + \omega_{it})^2 \tag{13}$$

where σ is a parameter to adjust the tolerant level of the residual. The weights are simple applications of the Bayes rule which give the ownership probability of every point.

In the Maximization step we solve the following two linear systems in order to update (u_1, v_1) and (u_2, v_2) where the indices $i1$ and $i2$ run over all points:

$$\begin{pmatrix} W_{i1}A_i\omega_{ix} & W_{i1}A_i\omega_{iy} & W_{i1}A_i\omega_{it} \end{pmatrix} \begin{pmatrix} u_1 \\ v_1 \\ 1 \end{pmatrix} = 0 \tag{14}$$

$$\begin{pmatrix} W_{i2}A_i\omega_{ix} & W_{i2}A_i\omega_{iy} & W_{i2}A_i\omega_{it} \end{pmatrix} \begin{pmatrix} u_2 \\ v_2 \\ 1 \end{pmatrix} = 0 \tag{15}$$

The EM-algorithm consists of subsequent iterations of the E- and M-step until there is no significant difference in the parameter estimates.

4 Localizing Occluding Boundary

After estimating the parameters of the two motions we would like to know which kind of multiple motion it is. If it is occlusion, we want to further localize the occluding boundary.

For all points $(\omega_{ix}, \omega_{iy}, \omega_{it})$ outside the two motion planes which satisfy

$$|\omega_{ix}u_j + \omega_{iy}v_j + \omega_{it}| > \epsilon \qquad j = 1, 2 \tag{16}$$

we count the number N_d of points with their amplitudes above a threshold. Since the distortion depends on the spectrum of the occluded signal $\tilde{I}_2(\vec{\kappa})$ the number N_d varies dramatically with $\tilde{I}_2(\vec{\kappa})$ as well. For example, if I_2 is a cosine image sequence then $\tilde{I}_2(\vec{\kappa})$ has only two spectral components and we have also a very small N_d. In order to solve this problem we count the number N_p on the planes as well and choose the relative ratio $R_a = \frac{N_d}{N_p}$ as our criterion. R_a will be much larger in case of occlusion than in case of transparency since all energy of the transparency lie on the dominant planes.

It is difficult to choose a suitable threshold without preknowledge. But we can round this problem by setting a series of thresholds and observing the variation of corresponding R_a. With the increase of threshold R_a will decrease in case of occlusion whereas it remains almost the same in case of transparency.

To enhance the robustness of a decision for occlusion we perform also a test on the spatial coherence based on observations in [8, 22, 12]. In the spatial domain we consider three successive frames from the image sequence and denote them with $I_{i\perp 1}$, I_i, and I_{i+1}. Then, we calculate the difference between two frames $I_{i\perp 1}$ and I_i using estimated speeds:

$$\Delta I_1 = I_i(x, y) - I_{i\perp 1}(x + u_1\delta t, y + v_1\delta t)$$

$$\Delta I_2 = I_i(x, y) - I_{i\perp 1}(x + u_2\delta t, y + v_2\delta t)$$

If the multiple motion is occlusion, we will observe one region with zero intensity in each one of ΔI_1 and ΔI_2 and these two regions are complementary in coordinates. Their intersection localizes the occluding boundary B_i. Repeating the same process on frames I_i and I_{i+1} we obtain the shifted boundary B_{i+1}. Using B_i and B_{i+1} we can determine the moving speed of the occluding boundary and thus segregate occluding and occluded signal (Fig. 3). If the multiple motion is transparency, we can not observe any region without error in ΔI_1 and ΔI_2 (Fig. 4). This fact can be used to distinguish occlusion and transparency. The reader may ask why this cumbersome multiple criteria. The answer lies in the limited resolution of the spectral domain. Although the definition of occlusion in the spectral domain is exact we can not rely on the decision for occlusion without a subsequent check in the spatial domain which fortunately is not computationally expensive.

5 Experiments

In this section we present experimental results for our algorithm on occlusion detection. Without exception the spatiotemporal support in all the experiments is a cube of $32 \times 32 \times 32$ pixels. For comparison we also apply the spatial EM algorithm based on equation (1). In order to obtain I_x, I_y, and I_t we convolve the cube with the first spatiotemporal derivatives of a 3D Gaussian function. Then we build the squared residuals R'_{i1} and R'_{i2} as following:

$$R'_{i1} = (I_{ix}u_1 + I_{iy}v_1 + I_{it})^2 \tag{17}$$

$$R'_{i2} = (I_{ix}u_2 + I_{iy}v_2 + I_{it})^2 \tag{18}$$

Replacing R_{i2} and R_{i2} in equation (12) and (13) with R'_{i1} R'_{i2} we build the weights W'_{i1} and W'_{i2} in the spatial domain. Correspondingly, the linear system turns

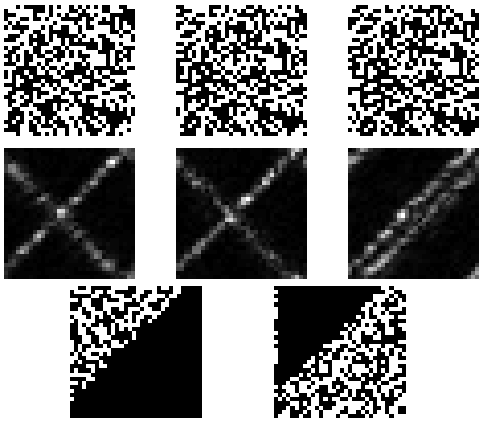

Figure 3: Application of spatial and spectral EM-algorithm on random dot occlusion sequence. **Top:** Three consecutive frames of the sequence. The occluding and the occluded signals moving with $(1,1)$ and $(1,-1)$ (pixel/frame), respectively. **Middle:** The spectrum of the cube shown with three sections. The origin lies in the middle of the image. From left to right: ω_y–ω_t, ω_y–ω_x, and ω_x–ω_t section. The sections give indications for the existence of two planes in the spectral domain. We apply the EM-algorithm with $\sigma = 0.1$ and initial speeds $(1.2,-0.1)$ and $(0.8,0.3)$. The results using the spatial EM are $(0.980,-0.996)$ and $(0.999,0.962)$ and the speeds estimated using the spectral EM are $(0.999,1.001)$ and $(0.996,-1.001)$. **Bottom:** ΔI_1 and ΔI_2, the two complementary regions with zero value are bounded by the occluding boundary.

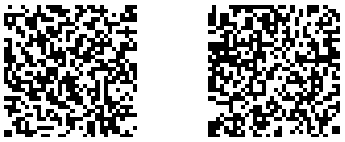

Figure 4: In the case of transparency with the same motion parameters there is no region with zero value in ΔI_1 and ΔI_2. **Left:** ΔI_1. **Right:** ΔI_2.

out to be:

$$\begin{pmatrix} W_{i1}^{'} I_{ix} & W_{i1}^{'} I_{iy} & W_{i1}^{'} I_{it} \end{pmatrix} \begin{pmatrix} u_1 \\ v_1 \\ 1 \end{pmatrix} = 0 \qquad (19)$$

$$\begin{pmatrix} W_{i2}^{'} I_{ix} & W_{i2}^{'} I_{iy} & W_{i2}^{'} I_{it} \end{pmatrix} \begin{pmatrix} u_2 \\ v_2 \\ 1 \end{pmatrix} = 0 \qquad (20)$$

Fig. 3 shows an example of applying both the spectral and the spatial EM-algorithms to segment different motions in a random dot occlusion sequence. The occluding signal is moving with $(1,1)$ (pixels/frame) and the occluded signal has a speed of $(1,-1)$. Both spatial EM and spectral EM converge after 6 iterations. The occluding boundary is displayed as intersection of the zeros regions in ΔI_1 and ΔI_2. The spectral EM-algorithm can handle transparency as well whereas the spatial EM can not. We prove this conclusion using a transparency random dot sequence with the same motion parameters. The spatial EM will not recognize the transparency as multiple motions: Both speeds converge to $(0.999,1.000)$. On the contrary, the spectral EM converges to $(0.998,0.999)$ and $(0.997,-0.998)$ after 5 iterations. In Fig. 4 are the corresponding ΔI_1 and ΔI_2 for transparency. There is no region without error. To test the existence of occlusion in the spectral domain we set $\epsilon = \frac{\pi}{4}$ and a series of thresholds to 0.1% and 1% of the maximal amplitude in the spectral domain. The corresponding R_a are 1.89 and 0.28 in case of occlusion and 0.29 and 0.25 in case of transparency, just like we expected.

In order to test the performance of both algorithms on estimating the number of models automatically, we propose an example of one moving signal. In Fig. 5 we show a random dot sequence with a single motion with velocity $(1,-1)$. In order to avoid the over-fitting, both EM-algorithms should converge to one speed. Actually this is the case. The spatial EM-algorithm converges to $(1.000,-1.000)$ after 2 iterations and the spectral EM-algorithm converges to $(0.996,-1.002)$ after 5 iterations.

In Fig. 6 we show a real example. The image sequence is composed of one occluding signal moving from left to right and one occluded signal moving from right to left. Both spatial and spectral EM-algorithms present satisfactory results with a better performance of the spectral algorithm. With $\epsilon = \frac{\pi}{4}$ we obtain R_a of 0.242 and 0.017 for thresholds of 0.1% and 1%. By using the spatial coherence we can further localize the occluding boundary which is displayed as intersection of zero regions in ΔI_1 and ΔI_2. Unfortunately, we have not been able to find a real sequence with transparency in order to further strengthen the arguments for the spectral treatment.

6 Discussion

We presented a formal model for the occlusion in the spectral domain. In contrast to ad hoc EM-algorithms in the spatial domain we can exactly describe the distortion to occlusion and employ an algorithm on the discrimination between occlusion and transparency. The superior performance of the spectral EM-algorithm relies on the correct model of mul-

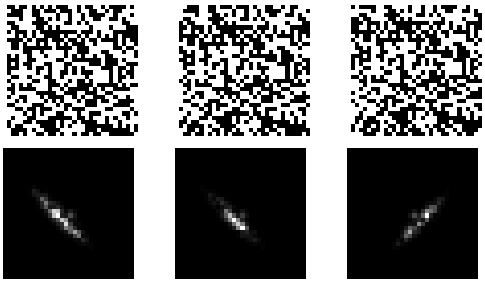

Figure 5: The over-fitting problem is avoided successfully in applying EM algorithms. **Top:** Three consecutive frames of a random dot sequence moving with a single speed $(1, -1)$. **Bottom:** The spectrum sections. The origin lies in the middle of the image. From left to right: ω_y–ω_t, ω_y–ω_x, and ω_x–ω_t section. Both spatial and spectral EM converge to a single speed.

tiple motion. Current spatial EM algorithms assume without a proof that both occlusion and transparency can be modeled by two instances of the brightness change constraint equation. Unfortunately, the spectral domain suffers under limited resolution so that the final verification test for occlusion is performed in the spatial domain. The assumed spatial coherence is the final cue that enables us to distinguish occlusion and transparency.

We have to study the behavior of the EM-algorithm both in the spectral and the spatial domain when two velocities of the occluding and the occluded signal converge to one speed. Alternative methods to the EM principle like the recent formalization of the grouping process [20] using normalized cuts have to be investigated with respect to their appropriateness to the given problem.

References

[1] A. P. Dempster , N. M. Laird , and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *J. R. Statist. Soc. B*, 39:1–38, 1977.

[2] E. H. Adelson and J. R. Bergen. Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America*, 1(2):284–299, 1985.

[3] S. S. Beauchemin and J. L. Barron. A Theory of Occlusion in the Context of Optical Flow. In *Advances in Computer Vision, Springer Wien New-York, F. Solina, W. Kropatsch, R. Klette and R. Bajcsy Eds., pp. 191-200, Nov.*, 1997.

[4] S.S. Beauchemin and J.L. Barron. The computation of optical flow. *ACM Computing Surveys*, 27:433–467, 1995.

[5] 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.

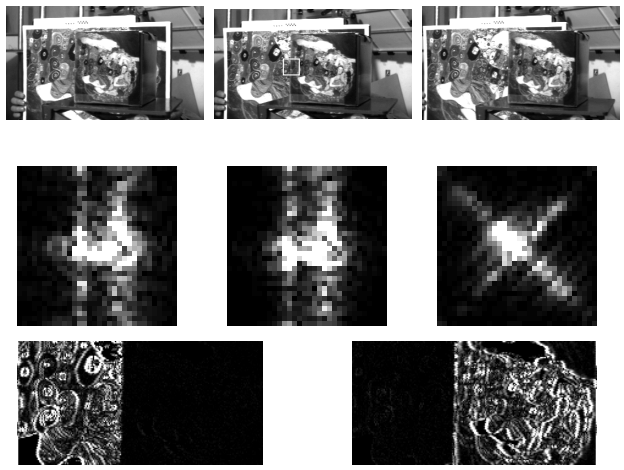

Figure 6: **Top:** The first, 16-th and 32-th frames of the image sequence. A cube around the occluding boundary is marked with a white box in the 16-th frame. **Middle:** The spectrum of the cube shown with three sections. From left to right: ω_y–ω_t, ω_y–ω_x, and ω_x–ω_t section. We apply the EM-algorithm with $\sigma = 0.1$ and initial speeds $(1.2, -0.1)$ and $(0.8, 0.3)$. After 12 iterations the estimated speeds are stabile. The results using spatial EM are $(1.239, 0.332)$ and $(-1.299, -0.051)$ and the results using spectral EM are $(0.985, 0.002)$ and $(-1.137, -0.029)$. Since both signals move almost horizontally the vertical motion components should be near zero. The spatial EM shows a higher error. **Bottom Left:** ΔI_1. **Bottom Right:** ΔI_2. For clarity we zoom in the region near the occluding boundary. We can see that two complementary regions with zero value are bounded by the occluding boundary.

[6] M. J. Black and P. Anandan. The robust estimation of multiple motions: parametricand piecewise-smooth flow fields. *Computer Vision and Image Understanding*, 63(1):75–104, 1996.

[7] P. Bouthemy. A maximum likelihood framework for determining moving edges. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 11:499–511, 1989.

[8] G. T. Chou. A model of figure-ground segregation from kinetic occlusion. In *Proc. Int. Conf. on Computer Vision*, pages 1050–1057, Boston, MA, June 20-23, 1995.

[9] D.J. Fleet and K. Langley. Computational analysis of non-fourier motion. *Vision Research*, 34:3057–3079, 1994.

[10] G. H. Granlund and H. Knutsson. *Signal Processing for Computer Vision*. Kluwer Academic Publishers, 1995.

[11] D. J. Heeger. Optical flow using spatiotemporal filters. *International Journal of Computer Vision*, 1(4):279–302, 1987.

[12] J. R. Bergen , P. J. Burt , R. Hingorani, and S. Peleg. A three-frame algorithm for estimating two-component im-

age motion. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 14(9):886–895, 1992.

[13] A. Jepson and M. Black. Mixture models for image representation. *Technical Report ARK96-PUB-54*, 1996.

[14] 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.

[15] 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.

[16] H.H. Nagel and A. Gehrke. Spatiotemporally adaptive estimation and segmentation of of-fields. In *Proc. Fifth European Conference on Computer Vision*, volume II, pages 86–102, Freiburg, Germany, June 2-6, H. Burkhardt (Ed.), Springer LNCS 1406, 1998.

[17] M. Proesmans, L.J. VanGool, E. Pauwels, and A. Oosterlinck. Determination of optical flow and its discontinuities using non-linear diffusion. In *Proc. Third European Conference on Computer Vision*, pages 295–304, Stockholm, Sweden, May 2-6, J.O. Eklundh (Ed.), Springer LNCS 801, 1994.

[18] M. Irani, B. Rousso, and S. Peleg. Computing occluding and transparent motions. *International Journal of Computer Vision*, 12:5–16, 1994.

[19] B. Jaehne, H. Haussecker, H. Scharr, H. Spies, D. Schmundt, and U. Schurr. Study of dynamical processes with tensor-based spatiotemporal image processing techniques. In *Proc. Fifth European Conference on Computer Vision*, volume II, pages 322–336, Freiburg, Germany, June 2-6, H. Burkhardt (Ed.), Springer LNCS 1406, 1998.

[20] 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.

[21] M. Shizawa and K. Mase. A unified computational theory for motion transparency and motion boundaries based on eigenenergy analysis. In *IEEE Conf. Computer Vision and Pattern Recognition*, pages 289–295, Maui, Hawaii, June 3-6, 1991.

[22] 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 Fransisco, CA, June 18-20, 1996.

[23] S.F. Wu and J. Kittler. A gradient-based method for general motion estimation and segmentation. *Journal of Visual Communication and Image Representation*, 4:25–38, 1993.

Appendix: The Spectrum of Occlusion

In this section we derive the spectrum of occlusion. Substituting equation (5) into (4) and utilizing the product property of impulse function we have:

$$\begin{aligned}\tilde{I}(\vec{\kappa},\omega_t) &= 2\pi^2\tilde{I}_1(\vec{\kappa})\delta(\vec{\kappa}^T\vec{v}_1+\omega_t)\\ &+(1-2\pi^2)\tilde{I}_2(\vec{\kappa})\delta(\vec{\kappa}^T\vec{v}_2+\omega_t)\\ &+\tfrac{2\pi}{i\vec{\kappa}^T\hat{\eta}}\delta(\vec{\kappa}^T\hat{\eta}_\perp,\vec{\kappa}^T\vec{v}_1+\omega_t)*\tilde{I}_1(\vec{\kappa})\delta(\vec{\kappa}^T\vec{v}_1+\omega_t)\\ &-\tfrac{2\pi}{i\vec{\kappa}^T\hat{\eta}}\delta(\vec{\kappa}^T\hat{\eta}_\perp,\vec{\kappa}^T\vec{v}_1+\omega_t)*\tilde{I}_2(\vec{\kappa})\delta(\vec{\kappa}^T\vec{v}_2+\omega_t)\end{aligned} \tag{21}$$

For simplification we define the third part as $A(\vec{\kappa},\omega_t)$ and the fourth $B(\vec{\kappa},\omega_t)$. Denoting with $\mathcal{F}$ and $\mathcal{F}^{\perp 1}$ the forward and inverse Fourier transformation we have:

$$\begin{aligned}A(\vec{\kappa},\omega_t) &= \tfrac{2\pi}{i\vec{\kappa}^T\hat{\eta}}\delta(\vec{\kappa}^T\hat{\eta}_\perp,\vec{\kappa}^T\vec{v}_1+\omega_t)*\tilde{I}_1(\vec{\kappa})\delta(\vec{\kappa}^T\vec{v}_1+\omega_t)\\ &= \mathcal{F}\{[\mathcal{F}^{\perp 1}(\tfrac{2\pi}{i\vec{\kappa}^T\hat{\eta}}\delta(\vec{\kappa}^T\hat{\eta}_\perp,\vec{\kappa}^T\vec{v}_1+\omega_t))]\\ &\quad\cdot[\mathcal{F}^{\perp 1}(\tilde{I}_1(\vec{\kappa})\delta(\vec{\kappa}^T\vec{v}_1+\omega_t))]\}\end{aligned}$$

The first part of the convolution is a line in 3D frequency space and the second part is a plane containing the line in the first part. Due to the relative symmetry of $\mathcal{F}$ and $\mathcal{F}^{\perp 1}$ [10] the first part turns out to be an impulse plane in spatial domain after inverse Fourier transformation and the second part is then an impulse line on this plane. Their multiplication results in an impulse line in 3D spatial domain. We take the forward Fourier transformation again and obtain an impulse plane with the orientation of $\vec{\kappa}^T\vec{v}_1+\omega_t=0$. Taking the coefficients into our account we have:

$$A(\vec{\kappa},\omega_t)=[\frac{2\pi}{i\vec{\kappa}^T\hat{\eta}}*\tilde{I}_1(\vec{\kappa})]\delta(\vec{\kappa}^T\vec{v}_1+\omega_t) \tag{22}$$

Taking into account that $A(\vec{\kappa},\omega_t)$ has the same orientation as occluding signal we draw a conclusion that this part of distortion just does nothing else as strengthening the spectrum of occluding signal.

The term $B(\vec{\kappa},\omega_t)$ can be computed similarly:

$$\begin{aligned}B(\vec{\kappa},\omega_t) &= \tfrac{2\pi}{i\vec{\kappa}^T\hat{\eta}}\delta(\vec{\kappa}^T\hat{\eta}_\perp,\vec{\kappa}^T\vec{v}_1+\omega_t)*\tilde{I}_2(\vec{\kappa})\delta(\vec{\kappa}^T\vec{v}_2+\omega_t)\\ &= \mathcal{F}\{[\mathcal{F}^{\perp 1}(\tfrac{2\pi}{i\vec{\kappa}^T\hat{\eta}}\delta(\vec{\kappa}^T\hat{\eta}_\perp,\vec{\kappa}^T\vec{v}_1+\omega_t))]\\ &\quad\cdot[\mathcal{F}^{\perp 1}(\tilde{I}_2(\vec{\kappa})\delta(\vec{\kappa}^T\vec{v}_2+\omega_t))]\}\end{aligned}$$

The difference here is that the two parts of the convolution do not have the same orientation ($\vec{v}_1\neq\vec{v}_2$). Therefore, we obtain *only* one point in 3D spatial domain after multiplication. Correspondingly, the distortion is overall in the frequency space. Moreover, the coefficient of $B(\vec{\kappa},\omega_t)$ is not dependent on ω_t:

$$B(\vec{\kappa},\omega_t)=[\frac{2\pi}{i\vec{\kappa}^T\hat{\eta}}*\tilde{I}_2(\vec{\kappa})]\delta(\vec{\kappa}^T\vec{v}_1+\omega_t) \tag{23}$$

Thus, we obtain the equation (6).