

SPATIO-TEMPORAL MOTION ESTIMATION FOR TRANSPARENCY AND OCCLUSIONS

Erhardt Barth*, Ingo Stuke^α, Til Aach^α, and Cicero Mota^β

Institute for Neuro- and Bioinformatics, University of Lübeck
Ratzeburger Allee 160, 23538 Lübeck, Germany

^α Institute for Signal Processing, University of Lübeck

^β Institute for Applied Physics, University of Frankfurt, Germany

ABSTRACT

We present a spatio-temporal analysis of motion at occluding boundaries as an extension of previous results for transparent motions. We show how these new results generalize alternative approaches derived in the Fourier domain that are limited by assuming straight occlusion boundaries. Furthermore, we derive a novel hierarchical algorithm that can deal with single, multiple-transparent, and occluded motions.

1. INTRODUCTION

Here we extend earlier results on transparent motions, first presented in [6], to the case of occluded motions. The model of occlusion that we use here is due to Fleet and Langley [4]. They also analyzed the problem of occlusion in the Fourier domain. This type of analysis was further developed in [2, 9]. Alternative spatial approaches have been developed in [5] and [3].

2. THEORETICAL RESULTS

2.1. From one to multiple transparent motions and occlusions

The well known brightness constancy constraint equation

$$\alpha(\mathbf{u})f = 0 \quad (1)$$

has been extended for the case of transparent motions by Shizawa and Mase [7]:

$$\alpha(\mathbf{u})\alpha(\mathbf{v})f = 0. \quad (2)$$

$f(\mathbf{x}, t)$ is the image-sequence intensity for coordinate $\mathbf{x} = (x, y)$ and time t . $\alpha(\mathbf{u}) = u_x \frac{\partial}{\partial x} + u_y \frac{\partial}{\partial y} + \frac{\partial}{\partial t}$ is the derivative in the direction of the vector $(u_x, u_y, 1)$. Here we will extend this motion model to include the case of occlusions.

*Work is supported by the Deutsche Forschungsgemeinschaft under Ba 1176/7-1. CM is also affiliated with the University of Amazonas, Brazil. We thank T. Martinetz who reminded us of the case of stationary background.

Suppose that two signals g_1, g_2 describe motions with velocities \mathbf{u}, \mathbf{v} respectively, i.e. $\alpha(\mathbf{u})g_1 = 0$ and $\alpha(\mathbf{v})g_2 = 0$, and that we use a blending signal h , i.e. $0 \leq h(\mathbf{x}, t) \leq 1$, with velocity \mathbf{u} to form a new signal. We write [4]

$$f = (1 - h)g_1 + hg_2. \quad (3)$$

The above equation describes a single motion \mathbf{u} for the case that $h = 0$, two transparent motions for $h = 1/2$, two translucent motions for $g_1 = 0$, and two occluded motions for $h = 1 - \chi$, where the binary mask χ defines g_1 as the occluding signal. To reveal the non-linear nature of a general algorithm for motion estimation, we note that if $\alpha(\mathbf{u})f$ never vanishes and the motions \mathbf{u} and \mathbf{v} vary only slowly, we have $\alpha(\mathbf{u})\alpha(\mathbf{v}) \log |\alpha(\mathbf{u})f| = 0$. In what follows, we will analyze the case of occlusions and refer to [6, 1, 8] for the case of transparent motions.

2.1.1. Occluded motions

The above mentioned case of occlusion defined by $h = 1 - \chi$ expands to:

$$f(\mathbf{x}, t) = \chi(\mathbf{x} - t\mathbf{u})g_1(\mathbf{x} - t\mathbf{u}) + (1 - \chi(\mathbf{x} - t\mathbf{u}))g_2(\mathbf{x} - t\mathbf{v}). \quad (4)$$

By applying the operator $\alpha(\mathbf{u})\alpha(\mathbf{v})$ to the previous equation we obtain

$$\alpha(\mathbf{u})\alpha(\mathbf{v})f = -\alpha(\mathbf{v})\chi(\mathbf{x} - t\mathbf{u})\alpha(\mathbf{u})g_2(\mathbf{x} - t\mathbf{v}). \quad (5)$$

We first note that $-\alpha(\mathbf{v})\chi(\mathbf{x} - t\mathbf{u}) = (\mathbf{u} - \mathbf{v}) \cdot \nabla\chi(\mathbf{x} - t\mathbf{u})$ where the derivatives of the discontinuous mask χ were taken in the sense of distribution theory. In what follows, all derivatives will be taken in this sense. First, we evaluate the distribution defined by the right-hand side of the above equation. Let ϕ be a Schwartz test function, then

$$\langle (\mathbf{u} - \mathbf{v}) \cdot \nabla\chi, \phi \rangle = \int (\mathbf{u} - \mathbf{v}) \cdot \nabla\chi(\mathbf{x})\phi(\mathbf{x}) d\mathbf{x} = \quad (6)$$

$$\int \chi(\mathbf{x})\nabla\phi(\mathbf{x}) \cdot (\mathbf{v} - \mathbf{u}) d\mathbf{x} = \int_{\Omega} \nabla\phi(\mathbf{x}) \cdot (\mathbf{v} - \mathbf{u}) d\mathbf{x}$$

where Ω is the support of χ . To make use of Gauss' theorem in the plane, we denote by B the boundary of Ω , by \mathbf{N} the unit normal to B , and by ds the arc-length element of B . We finally obtain the following equality:

$$\langle (\mathbf{u} - \mathbf{v}) \cdot \nabla \chi, \phi \rangle = \int_B \phi(\mathbf{x})(\mathbf{v} - \mathbf{u}) \cdot \mathbf{N}(\mathbf{x}) ds. \quad (7)$$

Remember that a Dirac distribution with support on B can be defined by the line integral

$$\langle \delta_B, \phi \rangle = \int_B \phi(\mathbf{x}) ds. \quad (8)$$

By comparing Equations (7) and (8), we find

$$(\mathbf{u} - \mathbf{v}) \cdot \nabla \chi = (\mathbf{v} - \mathbf{u}) \cdot \mathbf{N} \delta_B, \quad (9)$$

and therefore Equation (5) becomes

$$\alpha(\mathbf{u})\alpha(\mathbf{v})f = (\mathbf{v} - \mathbf{u}) \cdot \mathbf{N} \delta_B(\mathbf{x} - t\mathbf{u})\alpha(\mathbf{u})g_2(\mathbf{x} - t\mathbf{v}) \quad (10)$$

Thus, motion estimation fails at occlusions because both Equations (1) and (2) are not valid at points on the occluding boundary and should be replaced by Equation (10). Therefore, to estimate two occluding motions we either (i) use Equation (2) but do not integrate at occlusion points where we have motion discontinuities, or (ii) solve equation (10) to perform the estimation.

2.2. Fourier analysis of motion at occluding boundaries

We will now use our results to analyze occlusion in the Fourier domain for the simplest case of a straight occluding boundary. To do so, we start with the simplest case of one-dimensional motion where the boundary is just one point.

2.2.1. One-dimensional motion

Suppose that the support of g_1 is the positive half-line of real numbers. In this case $\chi(x) = 1$ if $0 \leq x$ and 0 otherwise. The unit normal to the boundary is $N(x) = 1$, and equation (10) becomes

$$\alpha(u)\alpha(v)f = (v - u)\alpha(u)g_2(x - tv)\delta(x - tu) \quad (11)$$

In the Fourier domain we obtain

$$2\pi j(u\xi + \xi_t)(v\xi + \xi_t)F = (u - v)\delta(u\xi + \xi_t) * (u\xi + \xi_t)G_2(\xi)\delta(v\xi + \xi_t), \quad (12)$$

denoting the transformed functions with the corresponding capital letters and the transform variables with (ξ, ξ_t) . By evaluating the right-hand side, we find

$$2\pi j(u\xi + \xi_t)(v\xi + \xi_t)F = (u - v)(u\xi + \xi_t)G_2\left(\frac{u\xi + \xi_t}{u - v}\right) \quad (13)$$

Remembering that for a distribution T we have $\xi T = 0 \iff T = c\delta$ for some complex constant c , after a forward and backward change of coordinates, we conclude that there exist two functions $A(\xi)$ and $B(\xi)$ such that

$$F = A(\xi)\delta(u\xi + \xi_t) + B(\xi)\delta(u\xi + \xi_t) + \frac{u - v}{2\pi j(v\xi + \xi_t)}G_2\left(\frac{u\xi + \xi_t}{u - v}\right) \quad (14)$$

We therefore conclude that $F(\xi, \xi_t)$ is the superposition of two Dirac lines with orientations in the directions of the motions and an additional distortion function. A similar result had been previously obtained by analyzing the Equation (4) - see [2]. The profile of the distortion function is hyperbolic along the lines $u\xi + \xi_t = c$ and, along the lines $v\xi + \xi_t = c$, has the same profile as the occluded signal.

2.2.2. Two-dimensional motion

Assume that Ω is an half-plane. In this case $\chi(\mathbf{x}) = 1$ if $0 \leq \mathbf{N} \cdot \mathbf{x}$ and 0 otherwise. We transform Equation (10) to the frequency domain, now $\xi = (\xi_x, \xi_y)$, and obtain

$$2\pi j(\mathbf{u} \cdot \xi + \xi_t)(\mathbf{v} \cdot \xi + \xi_t)F = (\mathbf{u} - \mathbf{v}) \cdot \mathbf{N} G_2(\theta)(\mathbf{u} \cdot \xi + \xi_t) \quad (15)$$

Note that \mathbf{N} has not transformed since it is constant due to the assumption of an extended straight boundary. $\theta = \theta(\xi, \xi_t)$ is the solution of $(\mathbf{u} - \mathbf{v}) \cdot \theta = \mathbf{u} \cdot \xi + \xi_t$, $\mathbf{N}^\perp \cdot \theta = \mathbf{N}^\perp \cdot \xi$ and therefore

$$F = A(\xi)\delta(\mathbf{u} \cdot \xi + \xi_t) + B(\xi)\delta(\mathbf{v} \cdot \xi + \xi_t) + \frac{(\mathbf{u} - \mathbf{v}) \cdot \mathbf{N}}{2\pi j(\mathbf{v} \cdot \xi + \xi_t)}G_2(\theta) \quad (16)$$

where

$$\theta = \frac{(\mathbf{u} \cdot \xi + \xi_t) \cdot \mathbf{N} + \xi \cdot \mathbf{N}^\perp (\mathbf{u} - \mathbf{v})^\perp}{(\mathbf{u} - \mathbf{v}) \cdot \mathbf{N}} \quad (17)$$

Thus, in a way similar to the one-dimensional case, we conclude that $F(\xi, \xi_t)$ is a superposition of two Dirac planes with orientations in the directions of the motions and an additional distortion function. A similar result has been obtained in [9]. Again, the profile of the distortion function is hyperbolic along lines with orientations $\mathbf{N} + \mathbf{u} \cdot \mathbf{N}e_t$ and, along the planes $\mathbf{v} \cdot \xi + \xi_t = c$, has the same profile as the occluded signal. e_t is the ξ_t axis defined as $e_t = (0, 0, 1)$.

2.3. Hierarchical algorithm

We now summarize previous results [6] that we need to define our algorithm. We expand Equation (2) to $\sum_I c_I f_I = 0$

where $I \in \{xx, yy, xy, xt, yt, tt\}$, f_I are the partial derivatives of f , and c_I are the mixed-motion parameters. We obtain $\mathbf{L}\mathbf{V} = \mathbf{0}$ where $\mathbf{L} = (f_I)$ and $\mathbf{V} = (c_I)^T$. For the case of n motions, this leads to a system of equations that can be solved by finding the eigenvector to the zero eigenvalue of $\mathbf{J}_n = \int \mathbf{L}(\mathbf{x})^T \mathbf{L}(\mathbf{x}) \omega(\mathbf{x}) d\mathbf{x}$, the *generalized structure tensor for n motions* of order $m = (n + 2)(n + 1)/2$. The eigenvector defines the mixed-motion parameters that are then separated by interpreting the motion vectors as complex numbers that are the roots of $Q_n(z) = z^n - A_{n-1}z^{n-1} + \dots + (-1)^n A_0$. To compute the coefficients, we just note that A_i are homogeneous symmetric functions of degree $n - i$ of $\mathbf{v}_1, \dots, \mathbf{v}_n$. For example, the coefficients of $Q_n(z)$ for two motions are [6] $A_1 = c_{xt} + jc_{yt}$ and $A_0 = c_{xx} - c_{yy} + jc_{xy}$. The confidence measure is based on the sum S of the diagonal minors of \mathbf{J}_n and the determinant K of \mathbf{J}_n . Our hierar-

Algorithm 1 Hierarchical motion estimation

- 1: Compute \mathbf{J}_n
 - 2: **if** $K^{1/m} < \epsilon_n S^{1/m-1}$ (high confidence) **then**
 - 3: Compute the mixed motion parameters based on \mathbf{J}_n
 - 4: **if** $n = 1$ **then**
 - 5: $\mathbf{v} = (V_x, V_y)$
 - 6: **else**
 - 7: \mathbf{u}, \mathbf{v} are the roots of $Q_2(z)$
 - 8: Append treated pixel \mathbf{x}_0 to list L
 - 9: **for all** $\mathbf{x}_0 \notin L$ **do**
 - 10: Repeat steps 1 to 11 with $\omega(\mathbf{x} - \mathbf{x}_0) = 0, \forall \mathbf{x} \notin L$.
 In addition, the size of the kernel ω is increased such as to include an number of M locations that are not in L .
-

chical algorithm first evaluates the confidence in one motion and estimates that one motion in case of high confidence. Otherwise, the confidence for two motions is evaluated and two motions are estimated. Moreover, motion at locations with low confidence is recomputed with a convolution kernel ω that integrates only locations with high confidence.

3. EXPERIMENTAL RESULTS

The results have been obtained for natural textures by placing and translating a square of one texture over a background with a second texture. In Figure (1), (a) shows the middle frame of the occlusion test sequence. The background is moving with velocity $\mathbf{v} = (0, 1)$ and the occluding square with $\mathbf{u} = (1, 0)$. The estimated motion vectors after the first iteration of Algorithm 1 are depicted in (b). Evidently, in a large area around the occluding boundary the estimation of the motion vectors did not fulfill the confidence criterion. The occluding boundary has been marked by a rectangle for convenience. After completing the iterations, we obtain the results shown in image (c). The number of

iterations varies from pixel to pixel and depends on the criterion M and the confidence measure for that location. The size of the kernel ω was $7 \times 7 \times 3$ and it was increased in space up to a criterion of $M = 15$. Note that the correct motion vectors are found on both sides of the boundary. Also note, that at some locations on the boundary we obtain the correct two motion vectors at that one location. This, however, becomes a problem only at the corners where one should use additional criteria for choosing the appropriate motion vector out of the two.

Figure (2) depicts results obtained for a stationary background. The same textures as in the previous example are used but the square is now moving with velocity $\mathbf{u} = (1, 1)$. In (a) we show results obtained by estimating only one motion without a confidence criterion. Note that in the boundary region all motion vectors are wrong. Image (b) depicts the results obtained after the first iteration of Algorithm 1 and image (c) the final result. The results are similar to the previous example. We first find a large area around the bounding box where the the confidence criterion is not matched but the initial result is much improved by iteration. In this case we used a criterion $M = 150$ (this number seems large but it corresponds roughly to a full window of size $7 \times 7 \times 3$). Note that we therefore obtained a somewhat stronger blurring of the occluding boundary, i.e. a small but more significant area where we obtained two motions that are, however, correctly estimated to be the zero motion of the background and the motion of the square respectively. In all cases first-order derivatives have been estimated using Gaussian kernels with a sigma of 1 pixel in all three directions (x, y, t) and second-order derivatives by iterating the first-order operations accordingly. The confidence criterion was in all cases defined by the values $\epsilon_1 = 0.1$ and $\epsilon_2 = 0.2$.

4. SUMMARY AND CONCLUSIONS

We have presented a general framework for estimation of single and multiple motions for both cases of transparency and occlusion. By linearizing a problem that is nonlinear in the spatial domain, and can therefore not be transferred to the Fourier domain, our approach is more general and includes Fourier-based approaches as a special case. Moreover, we have presented new results for the estimation of motions at occlusions and a hierarchical algorithm that can deal with both transparent and occluded motions.

5. REFERENCES

- [1] E. Barth, I. Stuke, and C. Mota. Analysis of motion and curvature in image sequences. In *Proc. IEEE Southwest Symp. Image Analysis and Interpretation*, pages 206–10, Santa Fe, NM, Apr. 7–9, 2002. IEEE Computer Press.
- [2] S. S. Beauchemin and J. L. Barron. The frequency struc-

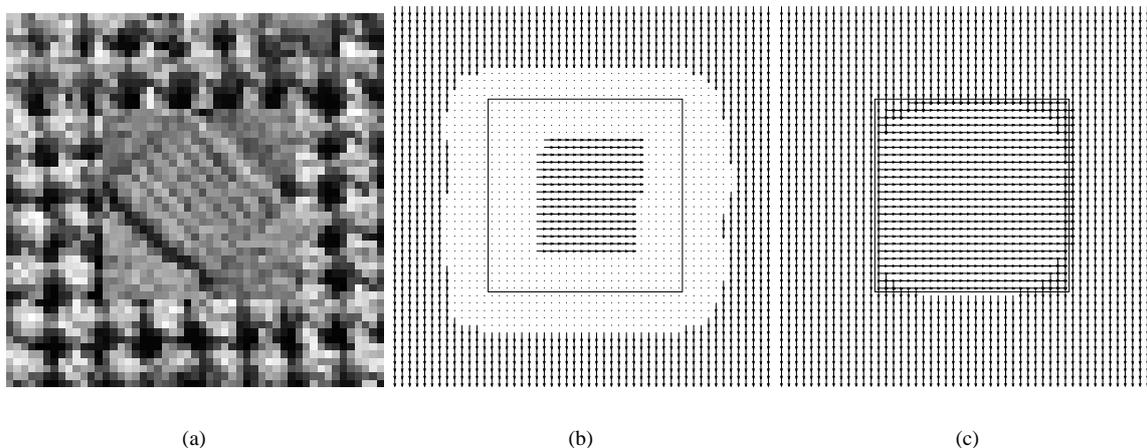


Fig. 1. Two occluding motions - see text for details.

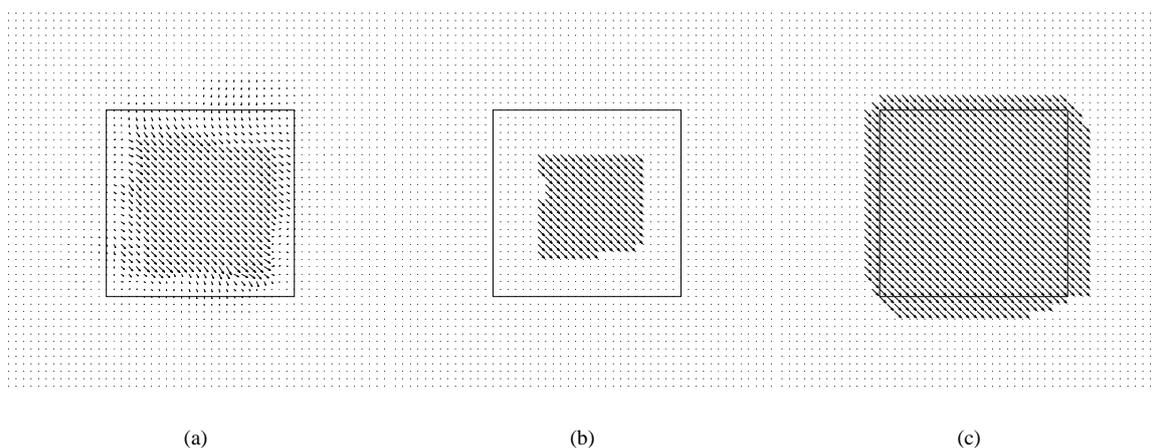


Fig. 2. Stationary background - see text for details.

ture of 1D occluding image sequences. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 22(2):200–6, Feb. 2000.

- [3] T. Darrell and E. Simoncelli. Nulling filters and the separation of transparent motions. In *IEEE Conf. Computer Vision and Pattern Recognition*, pages 738–9, New York, June 14–17, 1993. IEEE Computer Press.
- [4] D. J. Fleet and K. Langley. Computational analysis of non-Fourier motion. *Vision Research*, 34(22):3057–79, Nov. 1994.
- [5] M. Irani, B. Rousso, and S. Peleg. Computing occluding and transparent motions. *International Journal of Computer Vision*, 12(1):5–16, Feb. 94.
- [6] C. Mota, I. Stuke, and E. Barth. Analytic solutions for multiple motions. In *Proc. IEEE Int. Conf. Image Pro-*

cessing, volume II, pages 917–20, Thessaloniki, Greece, Oct. 7–10, 2001. IEEE Signal Processing Soc.

- [7] M. Shizawa and K. Mase. Simultaneous multiple optical flow estimation. In *IEEE Conf. Computer Vision and Pattern Recognition*, volume I, pages 274–8, Atlantic City, NJ, June 1990. IEEE Computer Press.
- [8] I. Stuke, T. Aach, C. Mota, and E. Barth. Estimation of multiple motions: regularization and performance evaluation. In B. Vasudev, T. R. Hsing, A. G. Tescher, and T. Ebrahimi, editors, *Image and Video Communications and Processing 2003*, volume 5022 of *Proceedings of SPIE*, pages 75–86, May 2003.
- [9] W. Yu, G. Sommer, S. Beauchemin, and K. Daniilidis. Oriented structure of the occlusion distortion: is it reliable? *IEEE Trans. Pattern Analysis and Machine Intelligence*, 24(9):1286–90, Sept. 2002.