Computing Multiple Image Motions

S. S. Beauchemin, K. Daniilidis and R. Bajcsy
GRASP Laboratory
Department of Computer and Information Science
University of Pennsylvania
Philadelphia PA 19104-6228 USA

Abstract

The computation of image motion for the purposes of determining egomotion is a challenging task as image motion includes discontinuities and multiple values mostly due to scene geometry, surface translucency and various photometric effects such as surface reflectance. We present algorithms for computing multiple image motions arising from occlusion and translucency which are capable of extracting the information-content of occlusion boundaries and distinguish between those and additive translucency phenomena. Sets of experimental results obtained on synthetic images are presented. These algorithms are based on recent theoretical results on occlusion and translucency in Fourier space.

1 Introduction

The importance of motion in image processing cannot be understated: in particular, approximations to image motion may be used to estimate 3D scene properties and motion parameters from a moving visual sensor, to perform motion segmentation, to compute the focus of expansion and time-to-collision, to perform motion-compensated image encoding, to compute stereo disparity and to measure biological parameters in medical imagery [1]

Based on recent theoretical developments in discontinuous motion, we devise multiple motion algorithms. We consider 1D and 2D signals, adopt a constant model of velocity and use a robust statistical procedure to extract multiple motions from local frequency spectra. The motion information provided by the algorithms includes single velocity, multiple (2) velocities, assessment of transparency versus occlusion, and upon occlusion events, the orientation of the occlusion boundary and the identification of the occluding signal.

1.1 Literature Survey

Many phenomena may cause multiple image motions. Occlusions, translucencies and various photometric effects such as specularities are among probable causes. In addition, occlusions contain valuable information concerning the geometry of the scene and may be used to decouple optical flow fields into their rotational and translational components, identify depth discontinuities, segment the scene with respect to motion and so on.

Computing multiple motions is a complex and rarely undertaken task. Indeed, most of the existing optical flow methods that have appeared in the literature make an explicit use of the optical flow constraint equation

$$\nabla \mathbf{I}^T \mathbf{v} - \mathbf{I}_t = 0, \tag{1.1}$$

where $\nabla \mathbf{I} = (\mathbf{I}_x, \mathbf{I}_y)^T$ is the spatial intensity gradient and $\mathbf{v} = (u, v)^T$ is the image velocity. At motion discontinuities, where the information content of a signal mostly resides, the use of (1.1) becomes problematic as the single motion hypothesis is violated. Area-based and feature-based correlation techniques are equally sensitive to occlusion as local image structures and features appear and disappear from one image to the next. To further complicate matters, regularization techniques which impose a degree of continuity to optical flow are also clearly inadequate over occlusion boundaries. However, in the more recent research in optical flow, the non-linear, discontinuous and multiple-valued nature of image motion in the coordinates of the image plane has been recognized [1].

In order to allow multiple motion events in optical flow estimation processes, a number of strategies have been devised, such as strong intensity gradients acting as inhibitors of flow coherence [8] and robust estimators designed to capture dominant motions [3]. Other techniques such as clustering [9], superposed motion layers and distributions [10], parametric models of motion with discontinuous functions [4] and mixtures of probability densities [7] have appeared.

Our approach emanates from recent theoretical results [2] describing the Fourier structure of occlusion and translucency phenomena for constant and linear models of optical flow.

1.2 Models of Optical Flow

The optical flow function may be expressed as an order n function of the image coordinates. Generally, we may write the Taylor series expansion for a i^{th} velocity as:

$$\mathbf{v}_i(\mathbf{x},t) = \sum_{j=0}^{p}\sum_{k=0}^{q}\sum_{l=0}^{r}\frac{\partial^{j+k+l}\mathbf{v}_i}{j!k!l!\partial x^j\partial y^k\partial t^l}x^j y^k t^l, \quad (1.2)$$

where $p+q+r \leq n$. For instance, the first-order expansion is written as $\mathbf{v}_i^{(1)}(\mathbf{x},t) = J_i\mathbf{x} + \mathbf{a}_i t$, where

$$J_i = \begin{pmatrix} a_{i1} & a_{i2} \\ b_{i1} & b_{i2} \end{pmatrix}$$

is the Jacobi matrix and $\mathbf{a}_i^T = -(a_{i3}, b_{i3})$ is translation[1]. We adopt in what follows a constant model of optical flow[2].

2 Structure of Occlusion

We proceed to describe the structure of occlusion events in the frequency domain for 1D and 2D signals composed of an arbitrary number of distinct frequencies.

Let $\mathbf{I}_1(x)$ and $\mathbf{I}_2(x)$ be 1D functions satisfying Dirichlet conditions such that they may be expressed as complex exponential series expansions:

$$\begin{aligned}\mathbf{I}_1(x) &= \sum_{n=-\infty}^{\infty} c_{1n}e^{ink_1x} \\ \mathbf{I}_2(x) &= \sum_{n=-\infty}^{\infty} c_{2n}e^{ink_2x}, \end{aligned} \quad (2.3)$$

where n is integer, c_{1n} and c_{2n} are complex coefficients and k_1 and k_2 are the fundamental frequencies of both signals.

Let $\mathbf{I}_1(x,t) = \mathbf{I}_1(v_1^{(0)}(x,t))$ and $\mathbf{I}_2(x,t) = \mathbf{I}_2(v_2^{(0)}(x,t))$. The frequency spectrum of the occlusion is:

$$\begin{aligned}\hat{\mathbf{I}}(k,\omega) &= \pi\sum_{n=i-\infty}^{\infty} c_{1n}\delta(k-nk_1, \omega+nk_1a_1) \\ &+ (1-\pi)\sum_{n=-\infty}^{\infty} c_{2n}\delta(k-nk_2, \omega+nk_2a_2) \\ &+ i\sum_{n=-\infty}^{\infty}\left(\frac{c_{2n}\delta(ka_1+\omega-nk_2\Delta a)}{(k-nk_2)}\right. \\ &- \left.\frac{c_{1n}\delta(ka_1+\omega)}{(k-nk_1)}\right), \end{aligned} \quad (2.4)$$

where $\Delta a = a_1 - a_2$.

In the 1D case, equation (2.4) reveals that the frequency spectra of both signals are preserved to within scaling factors. In addition, the Dirac delta functions $\delta(ka_1+\omega)$ and $\delta(ka_1+\omega\pm k_2\Delta a)$ constitute linear spectra, intersecting the frequencies of both the occluding and occluded signals, and are oriented in the direction of the constraint line pertaining to the occluding signal. Figure 2.1 shows a typical example with 1D translating sinusoids in an occlusion scene.

In the 2D case, equation (2.7) shows similarities with (2.4). The frequency spectra are planar and preserved to within scaling factors under occlusion and the distortions cast by the occlusion boundary fit oriented planes parallel to the plane containing the spectrum of the occluding signal.

Let $\mathbf{I}_1(\mathbf{x})$ and $\mathbf{I}_2(\mathbf{x})$ be 2D functions satisfying Dirichlet conditions such that they may be expressed as complex exponential series expansions:

$$\begin{aligned}\mathbf{I}_1(\mathbf{x}) &= \sum_{\mathbf{n}=-\vec{\infty}}^{\vec{\infty}} c_{1\mathbf{n}}e^{i\mathbf{x}^T N\mathbf{k}_1} \\ \mathbf{I}_2(\mathbf{x}) &= \sum_{\mathbf{n}=-\vec{\infty}}^{\vec{\infty}} c_{2\mathbf{n}}e^{i\mathbf{x}^T N\mathbf{k}_2}, \end{aligned} \quad (2.5)$$

where $\mathbf{n} = (n_x, n_y)^T$ and $N = \mathbf{n}^T I$ are integers, $\mathbf{x}$ are spatial coordinates, $\mathbf{k}_1 = (k_{1x}, k_{1y})^T$ and $\mathbf{k}_2 = (k_{2x}, k_{2y})^T$ are fundamental frequencies and $c_{1\mathbf{n}}$ and $c_{2\mathbf{n}}$ are complex coefficients. Let $\mathbf{I}_1(\mathbf{x},t) = \mathbf{I}_1(\mathbf{v}_1^{(0)}(\mathbf{x},t))$, $\mathbf{I}_2(\mathbf{x},t) = \mathbf{I}_2(\mathbf{v}_2^{(0)}(\mathbf{x},t))$ and the occluding boundary be locally represented by:

$$\mathbf{U}(\mathbf{x}) = \begin{cases} 1 & \text{if } \mathbf{x}^T\mathbf{n}_1 \geq 0 \\ 0 & \text{otherwise,} \end{cases} \quad (2.6)$$

where $\mathbf{n}_1$ is a vector normal to the occluding boundary at $\mathbf{x}$. The frequency spectrum of the occlusion is:

$$\begin{aligned}&\hat{\mathbf{I}}(\mathbf{k},\omega) = \\ &\pi\sum_{\mathbf{n}=-\vec{\infty}}^{\vec{\infty}} c_{1\mathbf{n}}\delta(\mathbf{k}-N\mathbf{k}_1, \omega+\mathbf{a}_1^T N\mathbf{k}_1) \\ &+(1-\pi)\sum_{\mathbf{n}=-\vec{\infty}}^{\vec{\infty}} c_{2\mathbf{n}}\delta(\mathbf{k}-N\mathbf{k}_2, \omega+\mathbf{a}_2^T N\mathbf{k}_2) \\ &-i\sum_{\mathbf{n}=-\vec{\infty}}^{\vec{\infty},}\left(\frac{c_{1\mathbf{n}}\delta((\mathbf{k}-N\mathbf{k}_1)^T\mathbf{n}_1^{\perp}, \mathbf{k}^T\mathbf{a}_1+\omega)}{(\mathbf{k}-N\mathbf{k}_1)^T\mathbf{n}_1}\right. \end{aligned}$$

[1] We use a negative translational rate without loss of generality and for mere mathematical convenience.

[2] The constant model may be simply defined as a linear model with $J_i = I$, yielding $\mathbf{v}_i^{(0)}(\mathbf{x},t) = \mathbf{x} - \mathbf{a}_i t$.

$$+ \frac{c_{2\mathbf{n}}\delta((\mathbf{k}-N\mathbf{k}_2)^T\mathbf{n}_1^\perp, \mathbf{k}^T\mathbf{a}_1+\omega-\Delta\mathbf{a}^T N\mathbf{k}_2)}{(\mathbf{k}-N\mathbf{k}_2)^T\mathbf{n}_1}\Bigg) \tag{2.7}$$

where $\Delta\mathbf{a} = \mathbf{a}_1 - \mathbf{a}_2$. Equation (2.7) is a generalization of equation (2.4) from 1D to 2D signals and its geometric interpretation is similar. For instance, frequencies $(N\mathbf{k}_1, -\mathbf{a}_1^T N\mathbf{k}_1)$ and $(N\mathbf{k}_2, -\mathbf{a}_2^T N\mathbf{k}_2)$ fit the constraint planes of the occluding and occluded signals, defined as $\mathbf{k}_1^T\mathbf{a}_1 + \omega = 0$ and $\mathbf{k}^T\mathbf{a}_2 + \omega = 0$. In the distortion term, the Dirac δ function with arguments $(\mathbf{k}-N\mathbf{k}_2)^T\mathbf{n}_1^\perp$ and $\mathbf{k}^T\mathbf{a}_1 + \omega - \Delta\mathbf{a}^T N\mathbf{k}_2$ represent a set of lines parallel to the constraint plane of the occluding signal $\mathbf{k}^T\mathbf{a}_1 + \omega = 0$ and, for every discrete frequency $N\mathbf{k}_1$ and $N\mathbf{k}_2$ exhibited by both signals, there is a frequency spectrum fitting the lines given by the intersection of planes $\mathbf{k}^T\mathbf{a}_1 + \omega - \Delta\mathbf{a}^T N\mathbf{k}_2 = 0$ and $(\mathbf{k}-N\mathbf{k}_1)^T\mathbf{n}_1^\perp = 0$. The magnitudes of these spectra are determined by their corresponding scaling functions $c_{1\mathbf{n}}[(\mathbf{k}-N\mathbf{k}_1)^T\mathbf{n}_1\,]^{-1}$ and $c_{2\mathbf{n}}[(\mathbf{k}-N\mathbf{k}_2)^T\mathbf{n}_1\,]^{-1}$.

3 Estimation of Multiple Image Motion

Equations (2.4) and (2.7) provide a model of the Fourier spectrum at an occlusion boundary. We devise several algorithms operating on local Fourier transforms which are capable of extracting multiple velocity measurements along with the information-content of occlusion boundaries.

3.1 1D Algorithm

Given a frequency measurement $(\hat{k}_j, \hat{\omega}_j)$, its corresponding velocity estimate is given by $\hat{\mathbf{v}}_i = (-\hat{\omega}_j/\hat{k}_j, 1)^T$. In optimal conditions, the non-zero spectrum of a purely translating image signal should be entirely consistent with its velocity. That is to say, every frequency measurement $(\hat{k}_j, \hat{\omega}_j)$ should be consistent with $\mathbf{v}_i$, the true signal velocity. However, owing to multiple factors such as acquisition noise, signal deformations and deviations from the locally constant velocity model, it may be that some variablility in the degree of agreement between measurements and the true velocity exists.

In light of this, an error metric, corresponding to the angular deviation between a measurement $\hat{\mathbf{m}}_j = (\hat{k}_j, \hat{\omega}_j)$ and an estimate of the i^{th} velocity $\hat{\mathbf{v}}_i$ may be defined as [7]

$$\theta(\hat{\mathbf{m}}_j, \hat{\mathbf{v}}_i) = \sin^{-1}\left(\frac{\hat{\mathbf{m}}_j^T\hat{\mathbf{v}}_i}{\|\hat{\mathbf{m}}_j\|_2\|\hat{\mathbf{v}}_i\|_2}\right). \tag{3.8}$$

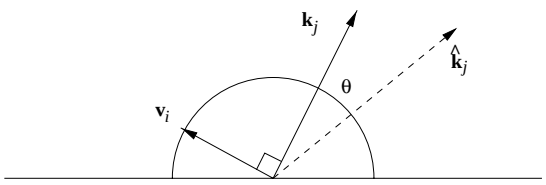

Figure 3.2: *The geometry of the angular error measure. The line defined by $\hat{\mathbf{k}}_j = (\hat{k}_j, \hat{\omega}_j)^T$ should be perpendicular to the line parallel to $\mathbf{v}_i = (-\omega_j/k_j, 1)^T$, as depicted by vector $\hat{\mathbf{k}}_j = (k_j, \omega_j)^T$.*

In addition, it is mathematically convenient to simplify the error metric and use the sine of the angle as the amount of deviation:

$$\xi(\hat{\mathbf{m}}_j, \hat{\mathbf{v}}_i) = \frac{\hat{\mathbf{m}}_j^T\hat{\mathbf{v}}_i}{\|\hat{\mathbf{m}}_j\|_2\|\hat{\mathbf{v}}_i\|_2}. \tag{3.9}$$

Under the assumption that the angular error is normally distributed, we define a mixture model of normal distributions to account for multiple motions. Consider G to be the set of measurements $\hat{\mathbf{m}}_j$, $j = 1, \ldots, n$. The probability density function for $\hat{\mathbf{m}}_j \in G$ is represented by the mixture of g normal distributions:

$$f(\hat{\mathbf{m}}_j, \psi) = \sum_{i=1}^{g} \pi_i f_i(\hat{\mathbf{m}}_j, \phi), \tag{3.10}$$

where $f_i(\hat{\mathbf{m}}_j, \phi)$ is a normal probability density function, $\psi = (\pi_1, \ldots, \pi_g, \mathbf{v}_1, \ldots, \mathbf{v}_g)^T$ is the vector of the mixture parameters and $\phi = (\mathbf{v}_1, \ldots, \mathbf{v}_g)^T$ is the vector of normal distribution parameters. π_i is the is the probability of $\hat{\mathbf{m}}_j$ to be from normal distribution f_i. The π_i are mixture probabilities and thus must satisfy

$$\sum_{i=1}^{g} \pi_i = 1 \tag{3.11}$$

In addition, the mixture parameters ϕ must satisfy the likelihood equation

$$\sum_{i=1}^{g}\sum_{j=1}^{n} \frac{\tau_{ij}\partial \ln f_i(\hat{\mathbf{m}}_j; \phi)}{\partial\phi} = 0 \tag{3.12}$$

which yields the constraints

$$\pi_i = \textstyle\sum_{j=1}^{n} \frac{\tau_{ij}}{n} \quad \text{and} \quad \mathbf{v}_i = \sum_{j=1}^{n} \frac{\tau_{ij}\hat{\mathbf{m}}_j^\perp}{n\pi_i}, \tag{3.13}$$

where τ_{ij} is the posterior probability that $\hat{\mathbf{m}}_j$ belongs to f_i and $\hat{\mathbf{m}}_j^\perp = (-\hat{\omega}_j/\hat{k}_j, 1)^T$ [7]. In this mixture model,

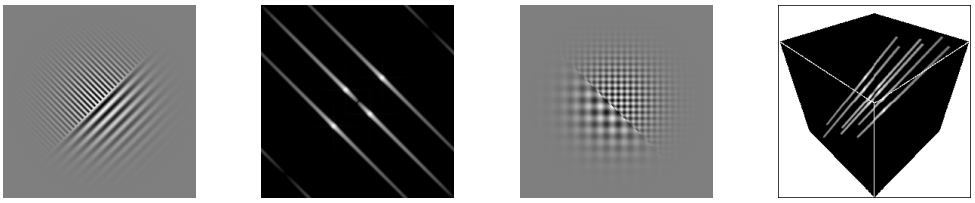

Figure 2.1: **(from left to right): a)** *Gaussian-windowed 1D signal with sinusoidals acting as occluding and occluded surfaces. The occluding signal has spatial frequency* $k_1 = \frac{2\pi}{16}$ *and velocity* $\mathbf{v}_1 = (1,1)$*. The occluded signal has frequency* $k_2 = \frac{2\pi}{8}$ *and velocity* $\mathbf{v}_2 = (-1,1)$*.* **b)** *Fourier spectrum of* **a)**. **c)** *Gaussian-windowed 2D signal with sinusoidals acting as occluding and occluded surfaces. The occluding signal has spatial frequency* $\mathbf{k}_1 = (\frac{2\pi}{16}, \frac{2\pi}{16})$ *and velocity* $\mathbf{v}_1 = (1,1,1)$*. The occluded signal has frequency* $\mathbf{k}_2 = (\frac{2\pi}{8}, \frac{2\pi}{8})$ *and velocity* $\mathbf{v}_2 = (-1,-1,1)$*.* **d)** *Fourier spectrum of* **c)**.

we hypothesize homoscedasticity, that is to say, the normal distributions within the mixture share the same standard deviation, which we consider as a constant. We also use an outlier detection mechanism based on Jepson and Black's model. We first pose the hypothesis that outlying measurements are uniformly distributed over the parameter space of the mixture, and thus we use a constant measure for the outlier probability of a measurement. We only update a mixture proportion for those. Constraints $\hat{\mathbf{m}}_j$ at a predetermined distance from other distributions should be considered as noisy measurements and not enter the velocity estimation process. The constant probability of observing a noisy measurement can be expressed as

$$\frac{1}{\sqrt{2\pi}\sigma_v} e^{\frac{-\lambda^2}{2\sigma_n^2}}, \tag{3.14}$$

from which we note that measurements at λ standard deviations from the means of the normal distributions are considered as corrupted by noise.

Further, the magnitude of measurements $\hat{\mathbf{m}}_j$ are relevant as the frequencies composing the distortion terms are typically smaller in magnitude than the frequencies of the signals from which they originate. In light of this, we incorporate the magnitude as the strength of measurements by replacing n, the number of measurements by $\sum_{j=0}^{n} \kappa(\hat{\mathbf{m}}_j)$, where κ is a measure of the magnitude of the local Fourier transform at $\hat{\mathbf{m}}_j$.

With the hypothesis of homoscedasticity, constant standard deviation and uniform distribution of noisy measurements, we establish the iterative equations for the Expectation-Maximization algorithm. The expectation step is the computation of posterior probabilities for the normal distributions, which we write as

$$\hat{\tau}_{ij}^{(k)} = \frac{\hat{\pi}_i^{(k)} e^{\frac{-1}{2\sigma_v^2}\xi^2(\hat{\mathbf{m}}_j, \hat{\mathbf{v}}_i)}}{\sum_{t=1}^{g} \hat{\pi}_t^{(k)} e^{\frac{-1}{2\sigma_v^2}\xi^2(\hat{\mathbf{m}}_j, \hat{\mathbf{v}}_t)} + \hat{\pi}_0^{(k)} e^{\frac{-\lambda^2}{2\sigma_n^2}}} \tag{3.15}$$

for $i = 1, \ldots, g$ and $j = 1, \ldots, n$ and, for the uniform distribution of noisy measurements, we write

$$\hat{\tau}_{0j}^{(k)} = \frac{\hat{\pi}_0^{(k)} e^{\frac{-\lambda^2}{2\sigma_v^2}}}{\sum_{t=1}^{g} \hat{\pi}_t^{(k)} e^{\frac{-1}{2\sigma_v^2}\xi^2(\hat{\mathbf{m}}_j, \hat{\mathbf{v}}_t)} + \hat{\pi}_0^{(k)} e^{\frac{-\lambda^2}{2\sigma_n^2}}} \tag{3.16}$$

for $j = 1, \ldots, n$. The equations for the maximization step, in which the parameters of the distributions are updated, are written as follows for the means

$$\hat{\mathbf{v}}_i^{(k+1)} = \frac{\sum_{j=1}^{n} \hat{\tau}_{ij}^{(k)} \kappa(\hat{\mathbf{m}}_j) \hat{\mathbf{m}}_j^{\perp}}{\hat{\pi}_i^{(k)} \sum_{j=1}^{n} \kappa(\hat{\mathbf{m}}_j)} \tag{3.17}$$

for $i = 1, \ldots, g$, and the mixture proportions are updated as

$$\hat{\pi}_i^{(k+1)} = \frac{\sum_{j=1}^{n} \kappa(\hat{\mathbf{m}}_j) \hat{\tau}_{ij}^{(k)}}{\sum_{j=1}^{n} \kappa(\hat{\mathbf{m}}_j)} \tag{3.18}$$

for $i = 0, \ldots, g$.

Figure 3.3a shows an example of observations randomly chosen from a superpopulation composed of 2 angular normals and a uniform distribution. The mixture parameters are $\psi = (\pi_1 = 0.4, \pi_2 = 0.4, v_1 = 1, v_2 = -1)^T$ for the normals and $\psi = (\pi_0 = 0.2)$ for the uniform distribution. An EM algorithm with angular error measure (3.9) was applied to this set of observations. After 15 iterations the algorithm converged to $\pi_0 = 0.182$, $\pi_1 = 0.411$, $\pi_2 = 0.406$, $v_1 = 1.017$ and $v_2 = -1.022$. Figures 3.3b and c show observations for which the final posterior probabilities τ_{ij} are above 0.95

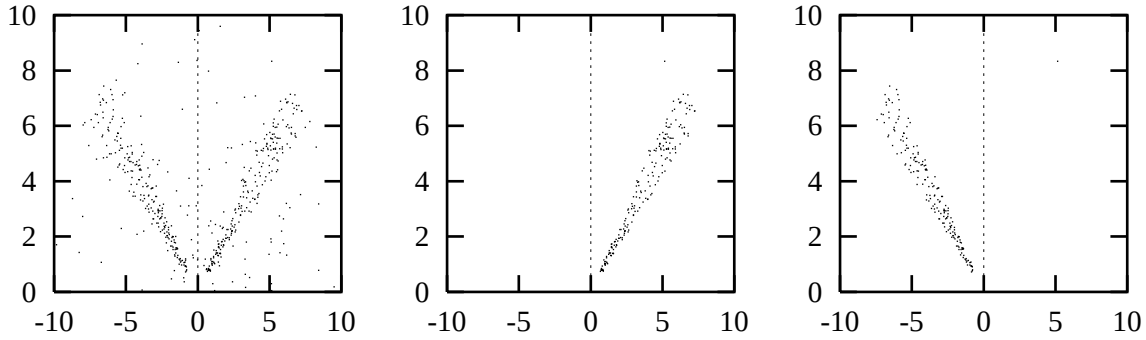

Figure 3.3: **a (left):** *Mixture of two angular normals randomly generated with* $\pi_0 = 0.20$, $\pi_1 = 0.40$, $\pi_2 = 0.40$, $\mathbf{v}_1 = (1, 1)$, $\mathbf{v}_2 = (-1, 1)$ *and* $\sigma_v = 0.075$ *radians.* **b) (center):** *EM results for observations with* $\tau_{1j} > 0.95$ *and* **c) (left):** *with* $\tau_{2j} > 0.95$. *The algorithm converged to* $\pi_0 = 0.182$, $\pi_1 = 0.411$, $\pi_2 = 0.406$, $\mathbf{v}_1 = (1.017, 1)$ *and* $\mathbf{v}_2 = (-1.022, 1)$ *in 15 iterations. Measurements at a distance* $\lambda = 2.5$ *standard deviations are considered as outliers.*

for $i = 1, 2$. Thresholding on the posterior probabilities allows to associate the observations with the various probability density functions composing the mixture.

In order to identify the spectra associated with occluding boundaries, we first find peak frequency measurements for both signals. That is to say, we find for signal t, the frequency $\hat{\mathbf{m}}_t$ such that $\tau_{tk} > \tau_{ik}$ for $t \neq i$ and $\kappa(\hat{\mathbf{m}}_t)$ is maximal and determine the strength of measurements $\hat{\mathbf{m}}_j$ along the direction perpendicular to the hypothesized occluding velocity at the peak frequency of the hypothesized occluded signal. Therefore, two tests are performed in order to verify which of both hypotheses is the correct one.

To test for the signal corresponding to velocity $\mathbf{v}_i$ as occluding, the procedure is to first consider only those measurements $\hat{\mathbf{m}}_j$ belonging to the uniform distribution of the mixture: $\tau_{0j} > \tau_{ij}$, for $i = 1, 2$ and $j = 1 \ldots n$, as determined by the EM algorithm and the peak frequency of the signal corresponding to velocity $\mathbf{v}_t$, where $t \neq i$. We then proceed with the computation of the strengths of measurements confirming this hypothesis.

Among measurements belonging to the uniform noise distribution, we compute their posterior probability of being part of the distortion spectra cast by the hypothesized occlusion as:

$$\hat{\tau}_{ij}(\hat{\mathbf{m}}_t, \hat{\mathbf{v}}_i) = \frac{e^{\frac{-1}{2\sigma_v^2}\xi^2((\hat{\mathbf{m}}_j - \hat{\mathbf{m}}_t), \hat{\mathbf{v}}_i)}}{e^{\frac{-1}{2\sigma_v^2}\xi^2((\hat{\mathbf{m}}_j - \hat{\mathbf{m}}_t), \hat{\mathbf{v}}_i)} + e^{\frac{-\lambda^2}{2\sigma_n^2}}}. \quad (3.19)$$

We also determine the posterior probabilities of the measurements to be from the uniform noise distribution to the exclusion of the spectra of the occlusion as:

$$\hat{\tau}_{0j}(\hat{\mathbf{m}}_t, \hat{\mathbf{v}}_i) = \frac{e^{\frac{-\lambda^2}{2\sigma_n^2}}}{e^{\frac{-1}{2\sigma_v^2}\xi^2((\hat{\mathbf{m}}_j - \hat{\mathbf{m}}_t), \hat{\mathbf{v}}_i)} + e^{\frac{-\lambda^2}{2\sigma_n^2}}}. \quad (3.20)$$

Mixture proportions may be obtained from these posterior probabilities that assess the hypothesis under test. These proportions are computed as:

$$\hat{\pi}_i(\hat{\mathbf{v}}_i) = \frac{\sum_{j=1}^{n} \kappa(\hat{\mathbf{m}}_j)\hat{\tau}_{ij}(\hat{\mathbf{m}}_t, \mathbf{v}_i)}{\sum_{j=1}^{n} \kappa(\hat{\mathbf{m}}_j)} \quad (3.21)$$

for $i = 0, 1$. Thus, if velocity $\hat{\mathbf{v}}_i$ is occluding, then the strengths of measurements confirming this hypothesis outnumber those pertaining to its contrary and thus

$$\frac{\pi_i(\hat{\mathbf{v}}_i)}{\pi_0(\hat{\mathbf{v}}_i)} > \frac{\pi_i(\hat{\mathbf{v}}_t)}{\pi_0(\hat{\mathbf{v}}_t)}. \quad (3.22)$$

Hence, various hypothesis-testing methods may be applied to determine the image events giving rise to multiple velocities.

3.2 2D Algorithm

The algorithm for 2D signals is essentially similar to the 1D algorithm we described. The measurements $\hat{\mathbf{m}}_j = (k_{xj}, k_{yj}, \omega_j)^T$ and velocity estimates $\hat{\mathbf{v}}_i = (v_x, v_y, v_t)^T$ are used in the error metric (3.9) to deterimine the posterior probabilities τ_{ij}, as is the case with the 1D algorithms. However, the choice of velocity estimates differs substantially. In the case of 2D signals, the velocity estimates at each EM iteration must maximize

the numerator exponential of (3.15). In this case, we follow the approach adopted by Jepson and Black [7], and consider the square of the error metric (3.9) as the equation for which the solutions yield velocity estimates. We observe that $\xi^2(\hat{\mathbf{m}}_j, \hat{\mathbf{v}}_i)$ may be written in matrix form as

$$(\mathbf{m}_j^T \mathbf{v}_i)^2 = \mathbf{v}_i^T M_j \mathbf{v}_i \tag{3.23}$$

where $M_j = \hat{\mathbf{m}}_j \hat{\mathbf{m}}_j^T$. By selecting the eigenvector corresponding to the minimum eigenvalue of M_j for $\mathbf{v}_i$, we minimize (3.23). Since M_j is real and symmetric, its eigenvalues are real and non-degenerate and the eigenvectors form an orthogonal basis in the space of measurements. In light of these observations, we define

$$\Sigma_i^{(k+1)} = \frac{\sum_{j=1}^n \tau_{ij}^{(k)} \kappa(\hat{\mathbf{m}}_j) M_j}{\sum_{j=1}^n \kappa(\hat{\mathbf{m}}_j)} \tag{3.24}$$

as the matrix from which the velocity estimate $\mathbf{v}_i^{(k+1)}$ is to be obtained in the form of the eigenvector $\mathbf{e}_i^{(k+1)}$ corresponding to the minimum eigenvalue $e_i^{(k+1)}$ of Σ_i. The minimum eigenvalue holds information about the velocity estimate obtained from its corresponding eigenvector. A zero value for e_i indicates that the velocity measurement is normal, whereas a non zero value indicates a full velocity measurement [5]. To see this, consider a set of observations consisting of collinear measurements, consistent with a normal velocity. It is observed that in such circumstances, the lines of matrix Σ_i are linearly dependent, leading to a minimum eigenvalue of value zero. Thus, the final eigenvalues e_i contain information on the nature of the measured velocities that is very relevant in most uses of image velocity.

Under the hypothesis of a straight-edged occlusion boundary, its normal may be estimated from the frequency structure of the occlusion. To perform this estimation, the algorithm must recover the orientation of the spectrum cast by the occlusion about the maximum frequency of the occluded signal, within a plane parallel to that of the occluding signal. To perform this estimation, it is necessary to include an EM iteration which converges to this linear orientation within the specified constraint plane.

We consider only those measurements which are consistent with the plane containing the peak frequency $\hat{\mathbf{m}}_t$ of the occluded signal and perpendicular to the occluding velocity $\mathbf{v}_i$, that is to say, we find $\hat{\mathbf{m}}_j - \hat{\mathbf{m}}_t$ such that $\tau_{ik} > \tau_{tk}$, for $t \neq k$. We proceed with the computation of posterior probabilities given an initial estimate $\hat{\mathbf{n}}^{(0)}$ of the orientation of the linear spectra cast by the occlusion:

$$\hat{\tau}_{ij}^{(k)} = \frac{\hat{\pi}_i^{(k)} e^{\frac{-1}{2\sigma_v^2}\theta^2((\hat{\mathbf{m}}_j - \hat{\mathbf{m}}_t), \hat{\mathbf{n}})}}{\sum_{t=1}^g \hat{\pi}_t^{(k)} e^{\frac{-1}{2\sigma_v^2}\theta^2((\hat{\mathbf{m}}_j - \hat{\mathbf{m}}_t), \hat{\mathbf{n}})} + \hat{\pi}_0^{(k)} e^{\frac{-\lambda^2}{2\sigma_n^2}}}. \tag{3.25}$$

$$\hat{\tau}_{0j}^{(k)} = \frac{\hat{\pi}_0^{(k)} e^{\frac{-\lambda^2}{2\sigma_v^2}}}{\sum_{t=1}^g \hat{\pi}_t^{(k)} e^{\frac{-1}{2\sigma_v^2}\theta^2((\hat{\mathbf{m}}_j - \hat{\mathbf{m}}_t), \hat{\mathbf{n}})} + \hat{\pi}_0^{(k)} e^{\frac{-\lambda^2}{2\sigma_n^2}}} \tag{3.26}$$

where θ is the error measure (3.8). The estimate of the spectral orientation and the mixture proportions are updated as:

$$\hat{\mathbf{n}}^{(k+1)} = \frac{\sum_{j=1}^n \hat{\tau}_{ij}^{(k)} \kappa(\hat{\mathbf{m}}_j)(\hat{\mathbf{m}}_j - \hat{\mathbf{m}}_t)}{\hat{\pi}_i^{(k)} \sum_{j=1}^n \kappa(\hat{\mathbf{m}}_j)} \tag{3.27}$$

$$\hat{\pi}_i^{(k+1)} = \frac{\sum_{j=1}^n \hat{\tau}_{ij}^{(k)} \kappa(\hat{\mathbf{m}}_j)}{\sum_{j=1}^n \kappa(\hat{\mathbf{m}}_j)} \tag{3.28}$$

4 Experiments

We performed numerical experiments on synthetic sinusoidal imagery composed of four 1D occlusion scenes and one 2D occlusion sequence, as described by Figure 4.4. The images used in these experiments are virtually free from noise. Local frequency measurements are obtained for an image location by computing a local Fast Fourier Transform within a region of side size 32. We observed that 30 iterations were sufficient for the EM algorithm to converge. The initial estimates for velocities and mixture proportions may be chosen randomly, but we prefer to have initial velocity estimates set as apart as possible to avoid convergence of both estimates to a single peak. When the EM iterations begin, we set σ_v to 0.2618 radians, or 15 degrees. At each step, we decrease σ_v to obtain a final value of 0.01745, or 1 degree. It is observed that a larger value for the standard deviation during the first iterations brings the initial velocity estimates in the neighborhood of the true parameters while a smaller value for the last iterations improves the accuracy of the final estimates. A value of 2.5 for λ and 1.0 for σ_n are chosen for the uniform distributions. It was experimentally determined that in order to assess the presence of multiple motions, the mixture probabilities must satisfy

$$\frac{\pi_i}{\pi_0} > \epsilon_1, \tag{4.29}$$

where $\epsilon_1 = 0.3$. In addition, to assess velocity $\hat{\mathbf{v}}_i$ as occluding, we required that

$$\left| \frac{\pi_i(\hat{\mathbf{v}}_i)}{\pi_0(\hat{\mathbf{v}}_i)} - \frac{\pi_i(\hat{\mathbf{v}}_t)}{\pi_0(\hat{\mathbf{v}}_t)} \right| > \epsilon_2, \tag{4.30}$$

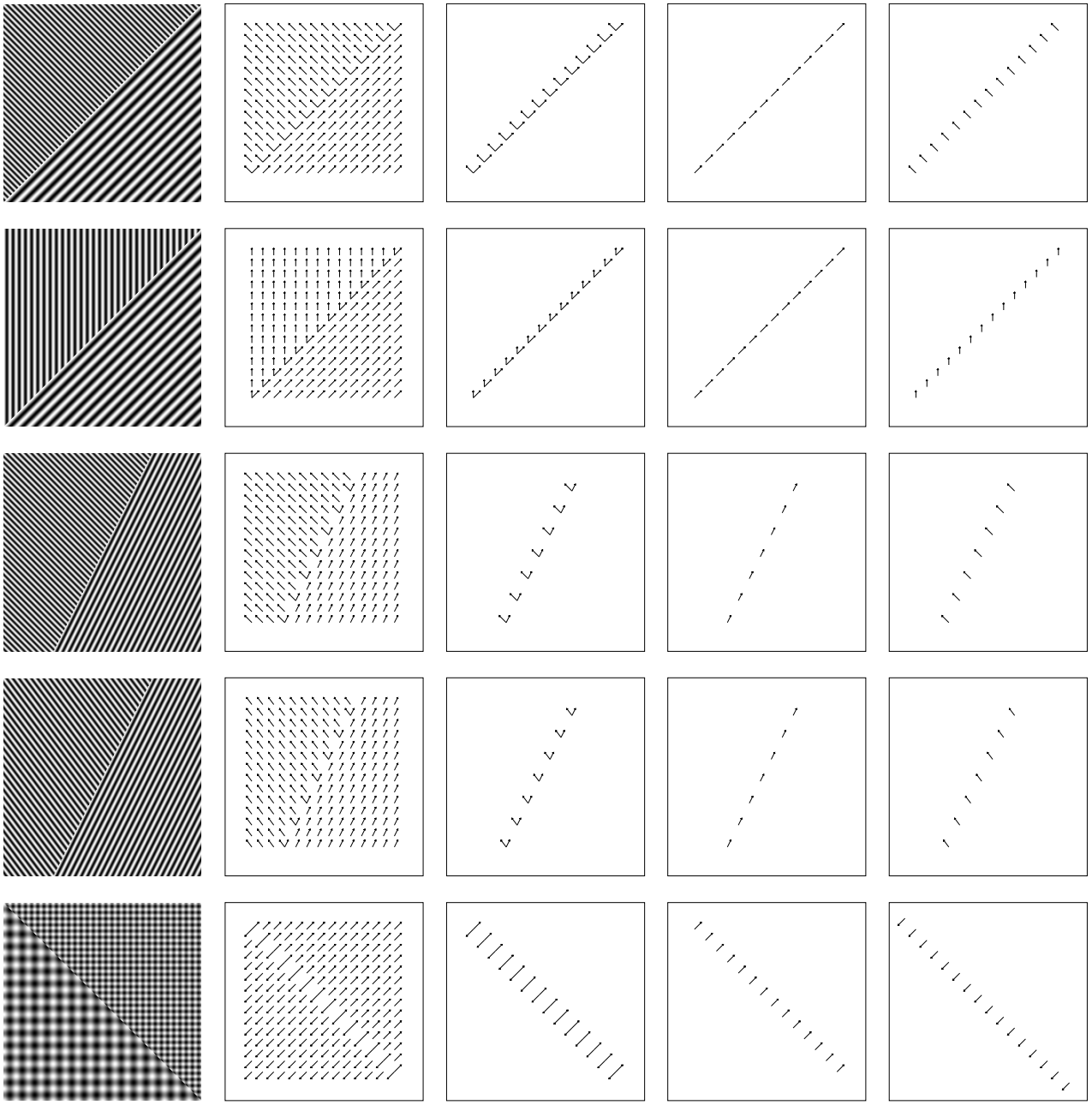

Figure 4.4: *Synthetic imagery and results with* $\mathbf{k}_1$ *and* $\mathbf{v}_1$ *occluding.* **(top to bottom):** *1D imagery* **a)** $k_1 = \frac{2\pi}{16}$, $k_2 = \frac{2\pi}{8}$, $\mathbf{v}_1 = (1,1)$ *and* $\mathbf{v}_2 = (-1,1)$. **b)** $k_1 = \frac{2\pi}{16}$, $k_2 = \frac{2\pi}{8}$, $\mathbf{v}_1 = (1,1)$ *and* $\mathbf{v}_2 = (0,1)$. **c)** $k_1 = k_2 = \frac{2\pi}{16}$, $\mathbf{v}_1 = (0.5,1)$ *and* $\mathbf{v}_2 = (-1,1)$. **d)** $k_1 = k_2 = \frac{2\pi}{16}$, $\mathbf{v}_1 = (0.5,1)$ *and* $\mathbf{v}_2 = (-0.75,1)$. **e)** *2D imagery* $\mathbf{k}_1 = (\frac{2\pi}{16}, \frac{2\pi}{16})$, $\mathbf{k}_2 = (\frac{2\pi}{8}, \frac{2\pi}{8})$ $\mathbf{v}_1 = (1,1,1)$ *and* $\mathbf{v}_2 = (-1,-1,1)$. **(left to right): a)** *Synthetic image.* **b)** *Optical flow.* **c)** *Multiple velocities.* **d)** *Occluding velocities.* **e)** *Occluded velocities.*

where $\epsilon_2 = 1.0 \times 10^{-3}$. Figure 4.4 shows the results obtained on the occlusion scenes. These optical flow fields are virtually free from error, due to the perfect nature of the synthetic imagery. However, we have observed that the degree to which these algorithms are capable of identifying multiple velocities that are rela-

tively similar in their orientation is not very satisfying. The velocities must be at least 15 to 20 degrees apart in orientation for the algorithms to yield a positive assessment of multiple velocities. Issues such as the values of the various standard deviations for the mixture and the orientations of the initial estimates have a definite influence on this phenomenon. One potential solution to obtain better orientational resolution would be to perform several EM iterations in parallel with different values for their initial estimates and then proceed with an analysis of the final convergence values.

5 Conclusion

The nature of discontinuous image motions in Fourier space has long been unclear. The algorithms proposed in this contribution are based on a firm theoretical framework which describes the coherent behavior of occlusion events in Fourier space. However, open questions abound: Theoretically, the structure of planar motion, quadratic in the imaging plane, remains to be established in Fourier space. In addition, the algorithms proposed herein may serve as a first stage into the perceptual grouping of velocities, allowing to identify the occluding and occluded signals not only at occluding boundaries but within regions exhibiting coherent motions, and therefore leading to performing motion-based image segmentation. Further, Occlusion detection operators could also be developed within the context of this theoretical framework and unreported experiments conducted with occlusion-tuned Gabor filters on a 1D pair of translating signals show this possibility.

Experimentally, the limiting conditions under which the current techniques fail must be established. For instance, the degree of multiple velocity resolution and the factors influencing it must be identified. In its current state, the experimental evaluation only confirms that noise-free imagery under optimal conditions yield noise-free results. However, it has been clear for some time that a number of vision algorithms fail to meet this fundamental criterion [1].

To conclude, we have demonstrated the feasibility of computing discontinuous motions and other measurements such as the local identification of occluding velocities and occlusion boundary normals, translucency phenomena and the disambiguation of occluding signals suffering from the aperture problem. The theoretical framework under which these algorithms have been devised constitutes a foundation for further research in motion analysis. Indeed, we strongly believe that further developments in the field of optical flow and motion analysis ought to be based on firmly established theoretical backgrounds rather than incidental evidence [6].

References

[1] J. L. Barron, D. J. Fleet, and S. S. Beauchemin. Performance of optical flow techniques. *IJCV*, 12(1):43–77, 1994.

[2] S. S. Beauchemin, A. Chalifour, and J. L. Barron. Discontinuous optical flow: Recent theoretical results. In *Vision Interface*, pages 57–64, Kelowna, Canada, May 1997.

[3] M. J. Black. A robust gradient-method for determining optical flow. Technical Report YALEU/DCS/RR-891, Yale University, New-Haven, CT, 1991.

[4] M. J. Black and A. Jepson. Estimating optical flow in segmented images using variable-order parametric models with local deformations. Technical Report SPL-94-053, Xerox Systems and Practices Laboratory, Palo Alto, California, 1994.

[5] B. Jahne. Motion determination in space-time images. In *Proceedings of ECCV*, pages 161–173, Antibes, France, April 1990.

[6] R. C. Jain and T. O. Binford. Ignorance, myopia and naivete in computer vision systems. *CVGIP:IU*, 53:112–117, 1991.

[7] A. D. Jepson and M. Black. Mixture models for optical flow computation. In *IEEE Proceedings of CVPR*, pages 760–761, New York, New York, June 1993.

[8] H.-H. Nagel. On the estimation of optical flow: Relations between different approaches and some new results. *Artificial Intelligence*, 33:299–324, 1987.

[9] B. G. Schunck. Image flow segmentation and estimation by constraint line clustering. *IEEE PAMI*, 11(10):1010–1027, 1989.

[10] J. Y. A. Wang and E. H. Adelson. Layered representation for motion analysis. In *Proceedings of CVPR*, 1993.