Oriented Structure of the Occlusion Distortion: Is It Reliable?

Weichuan Yu, *Member*, *IEEE*, Gerald Sommer, Steven Beauchemin, *Member*, *IEEE*, and Kostas Daniilidis, *Member*, *IEEE*

Abstract—In the energy spectrum of an occlusion sequence, the distortion term has the same orientation as the velocity of the occluding signal. Recent works claimed that this oriented structure can be used to distinguish the occluding velocity from the occluded one. Here, we argue that the orientation structure of the distortion cannot always work as a reliable feature due to the rapidly decreasing energy contribution. This already weak orientation structure is further blurred by a superposition of distinct distortion components. We also indicate that the superposition principle of Shizawa and Mase for multiple motion estimation needs to be adjusted.

Index Terms—Optical flow, occlusion, motion discontinuities, spectral analysis.

1 Introduction

THE motivation of studying multiple motions (including occlusion and additive transparency) in the spectral domain is mainly due to the inadequacy of spatial motion models [3]. The spectrum of multiple motions was first analyzed by Fleet [8]. Beauchemin and Barron [2], [3], [4] formulated an explicit model of occlusion in the frequency domain. They claimed that the distortion term in the occlusion spectrum can be used to distinguish the occluding velocity from the occluded one because this distortion term has the same orientation as the occluding velocity. In this paper, we show that the orientation of the distortion term cannot be used to reliably identify an arbitrary occluding velocity. We start with the spectral analysis of a 1D occlusion sequence. The conclusion of Beauchemin and Barron [3] is proven to be a special case of our analysis in the sense that their analysis uses only a few spectral components, while the number of spectral components in real signals is arbitrary. Then, we discuss the 2D occlusion spectrum and the spectrum of the additive transparency. We further indicate that the superposition principle of multiple motions proposed by Shizawa and Mase [15] needs to be adjusted. Finally, this paper discusses other related work as well as the merits and shortcomings of frequency-based motion models.

2 The Spectral Analysis of Occlusion and Transparency

The spectrum of occlusion was first analyzed by Fleet and Langley [8], [9]. They used a characteristic function $\chi(\mathbf{x})$ and modeled the occlusion in the spatial domain as

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

where $I_1(\mathbf{x})$ is a 2D **occluding** signal moving with a velocity $\mathbf{v}_1 = (v_{1x}, v_{1y})^T$ and $I_2(\mathbf{x})$ is a 2D **occluded** signal moving with a velocity $\mathbf{v}_2 = (v_{2x}, v_{2y})^T$. This equation is the seed of all other occlusion models appeared in the later part of this paper. For simplicity, in this paper we focus on constant velocity models.

2.1 1D Occlusion Spectrum

Beauchemin and Barron [3] gave a detailed spectral analysis of a 1D occlusion sequence. They replaced the vectors in (1) with scalars and used a 1D Heaviside unit step function $u(x)$ for $\chi(x)$ yielding

$$I(x,t) = u(x-v_1 t)I_1(x-v_1 t) + [1-u(x-v_1 t)]I_2(x-v_2 t) \quad (2)$$

with

$$u(x) = \begin{cases} 1 & x \geq 0 \\ 0 & \text{otherwise.} \end{cases} \quad (3)$$

The Fourier transform of (2) reads

$$\begin{aligned} &\tilde{I}(\omega_x,\omega_t) = \\ &\tilde{u}(\omega_x)\delta(\omega_x v_1+\omega_t) * \tilde{I}_1(\omega_x)\delta(\omega_x v_1+\omega_t) + \tilde{I}_2(\omega_x)\delta(\omega_x v_2+\omega_t) \\ &\quad - \tilde{u}(\omega_x)\delta(\omega_x v_1+\omega_t) * \tilde{I}_2(\omega_x)\delta(\omega_x v_2+\omega_t), \end{aligned} \quad (4)$$

where

$$\tilde{u}(\omega_x) = \pi\delta(\omega_x) + \frac{1}{i\omega_x}. \quad (5)$$

Here, $*$ means convolution and $\tilde{}$ denotes the Fourier transform of the corresponding signal.

Equation (4) (the same as (7) in [3]) is valid for an arbitrary signal which satisfies the Dirichlet conditions. Substituting (5) into (4) and using the product property of the impulse function yield

$$\begin{aligned} &\tilde{I}(\omega_x,\omega_t) = \\ &\left[\pi\tilde{I}_1(\omega_x) + \frac{1}{i\omega_x} * \tilde{I}_1(\omega_x)\right]\delta(\omega_x v_1+\omega_t) + (1-\pi)\tilde{I}_2(\omega_x)\delta(\omega_x v_2+\omega_t) \\ &\quad + \frac{i}{\omega_x}\delta(\omega_x v_1+\omega_t) * \tilde{I}_2(\omega_x)\delta(\omega_x v_2+\omega_t). \end{aligned} \quad (6)$$

The first two terms in (6) are two oriented lines (for generality, we assume $v_1 \neq v_2$) passing through the origin of the spectral space. Note that the additional convolution $\frac{1}{i\omega_x} * \tilde{I}_1(\omega_x)$ in the first term does not disturb the orientation. Instead, it strengthens the corresponding spectral line. Thus, we do not consider it as distortion.

The distortion comes from the third term in (6). This term is a convolution of two spectral lines which indicate the occluding velocity v_1 and the occluded velocity v_2, respectively. To obtain a clear geometric interpretation of this convolution, we decompose the static spectrum $\tilde{I}_2(\omega_x)$ of the occluded signal into Dirac components

$$\tilde{I}_2(\omega_x) = \sum_m c_m \delta(\omega_x - \omega_m), \quad (7)$$

where m indexes the different frequency components and c_m denotes the corresponding coefficients. After this reformulation, the distortion term then reads

- *W. Yu is with the Department of Diagnostic Radiology, Yale University, BML 332, PO Box 208042, New Haven, CT 06520-8042. E-mail: weichuan@noodle.med.yale.edu.*
- *G. Sommer is with the Institute of Computer Science, University Kiel, Preusserstrasse 1-9, D-24105 Kiel, Germany. E-mail: gs@ks.informatik.uni-kiel.de.*
- *S. Beauchemin is with the Department of Computer Science, The University of Western Ontario, London, ON, Canada. E-mail: beau@csd.uwo.ca.*
- *K. Daniilidis is with the GRASP Laboratory, University of Pennsylvania, 3401 Walnut Street, Philadelphia, PA 19104-6228. E-mail: kostas@grasp.cis.upenn.edu.*

Manuscript received 20 June 2001; revised 6 Dec. 2001; accepted 4 Feb. 2002. Recommended for acceptance by Z. Zhang.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number 114388.

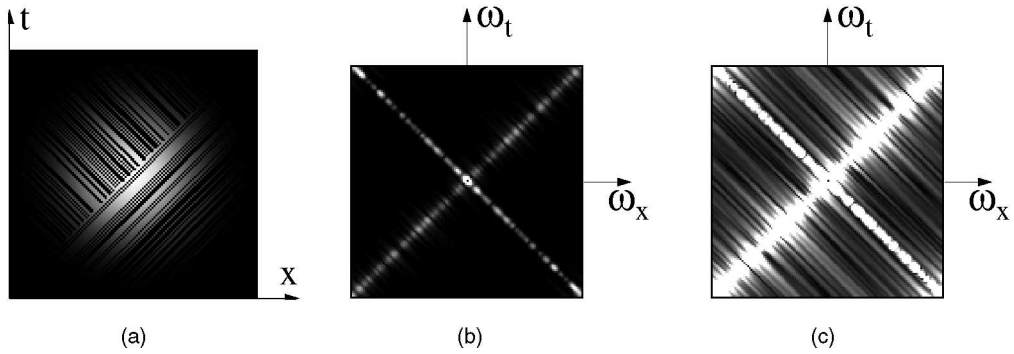

Fig. 1. (a) One-dimensional random dot occlusion sequence windowed by a Gaussian. The occluding velocity is 1 [pixel/frame] and the occluded velocity is -1 [pixel/frame]. (b) The energy spectrum of the occlusion sequence. For display purposes, we delete the DC component. The occluding and the occluded signal are clearly characterized by two oriented lines passing through the origin. The zipper-like structure along the occluded line is caused by the distortion. Each distortion line has the same orientation as the occluding signal. Its amplitude decreases hyperbolically after leaving the occluded spectral line. (c) In order to display the superposition effect of distortion lines, we set a threshold equal to 1 percent of the maximal value of the occlusion spectrum. All values above this threshold are reduced to be equal to the threshold. The more distortion lines we have, the less distinct the distortion structure is.

$$\begin{aligned}&\frac{i}{\omega_x}\delta(\omega_x v_1+\omega_t)*\tilde{I}_2(\omega_x)\delta(\omega_x v_2+\omega_t)\\&=\sum_m \frac{i\,c_m}{\omega_x-\omega_m}\delta(\omega_x v_1+\omega_t-\omega_m(v_1-v_2)).\end{aligned}\tag{8}$$

It is clear that each Dirac component $\delta(\omega_x-\omega_m)$ in $\tilde{I}_2(\omega_x)$ will cause an oriented distortion line $\delta(\omega_x v_1+\omega_t-\omega_m(v_1-v_2))$ after the convolution.

The orientation structure formed by the distortion lines characterizes the occlusion spectrum and many approaches (e.g., [9], [3]) have tried to use this structure in occlusion analysis. But unfortunately, this structure is not a robust feature. For example, if $\tilde{I}_2(\omega_x)$ has many Dirac components with significant energy contribution, which is common for typical real signals, this kind of orientation may disappear due to the superposition (see Fig. 1). More importantly, after leaving the intersection point with the occluded spectral line, the distortion line decreases rapidly in amplitude because its weight is a hyperbolic term $\frac{1}{\omega_x-\omega_m}$. In most spectral regions, the distortion is too weak to be useful.

Theorem 1 in [3] (cf. (11) in [3]) is correct with respect to the orientation of the distortion term using two different cosine functions as occluding and occluded signal. However, the property of Theorem 1 cannot reliably hold for an arbitrary signal because the specific orientation of the distortion may vanish after the superposition of cosine/sine signals. Theorem 2 in [3] did not account for this superposition effect and turned out to be a specific conclusion only. We may explain this point by demonstrating that (11) in [3] is only a special case of (6). Using the same cosine functions as in Theorem 1 in [3] for the occluding and occluded signal:

$$\begin{cases} I_1(x-v_1t) &= c_1 cos(\omega_1(x-v_1t)) \\ I_2(x-v_2t) &= c_2 cos(\omega_2(x-v_2t)) \end{cases},$$

we obtain the corresponding spectra

$$\begin{cases}\tilde{I}_1(\omega_x)\delta(\omega_x v_1+\omega_t)\\ \quad=\frac{c_1}{2}[\delta(\omega_x-\omega_1)+\delta(\omega_x+\omega_1)]\delta(\omega_t+\omega_x v_1)\\ \quad=\frac{c_1}{2}\delta(\omega_x-\omega_1,\omega_t+\omega_1 v_1)+\frac{c_1}{2}\delta(\omega_x+\omega_1,\omega_t-\omega_1 v_1)\\ \tilde{I}_2(\omega_x)\delta(\omega_x v_2+\omega_t)\\ \quad=\frac{c_2}{2}[\delta(\omega_x-\omega_2)+\delta(\omega_x+\omega_2)]\delta(\omega_t+\omega_x v_2)\\ \quad=\frac{c_2}{2}\delta(\omega_x-\omega_2,\omega_t+\omega_2 v_2)+\frac{c_2}{2}\delta(\omega_x+\omega_2,\omega_t-\omega_2 v_2).\end{cases}\tag{9}$$

Substituting (9) into (6) yields

$$\begin{aligned}\tilde{I}(\omega_x,\omega_t)=&\frac{\pi}{2}c_1[\delta(\omega_x-\omega_1,\omega_t+\omega_1 v_1)+\delta(\omega_x+\omega_1,\omega_t-\omega_1 v_1)]\\&-\frac{i}{2}c_1\left[\frac{1}{\omega_x-\omega_1}\delta(\omega_x v_1+\omega_t)+\frac{1}{\omega_x+\omega_1}\delta(\omega_x v_1+\omega_t)\right]\\&+\frac{1-\pi}{2}c_2[\delta(\omega_x-\omega_2,\omega_t+\omega_2 v_2)+\delta(\omega_x+\omega_2,\omega_t-\omega_2 v_2)]\\&+\frac{i}{2}c_2\left[\frac{1}{\omega_x-\omega_2}\delta(\omega_x v_1+\omega_t-\omega_2(v_1-v_2))\right.\\&\left.+\frac{1}{\omega_x+\omega_2}\delta(\omega_x v_1+\omega_t+\omega_2(v_1-v_2))\right].\end{aligned}\tag{10}$$

This equation is exactly the same as (11) in [3]. Here, we do not consider the second term in (10) as distortion because it strengthens the spectral line of the occluding signal. Note that the distortion lines also partially contribute to the occluded spectral line (just evaluate the fourth term in (10) by setting $\omega_x=\omega_2$ or $\omega_x=-\omega_2$).

In Fig. 1, we display a 1D random dot occlusion sequence smoothed by a Gaussian window and the corresponding energy spectrum. The spectrum is characterized by two dominant spectral lines with distortions crossing the occluded spectral line. The amplitude of distortion decreases rapidly after leaving the occluded spectral line. In order to display the superposition effect of the distortion, we set a threshold equal to 1 percent of the maximal amplitude in the occlusion spectrum. All values above the threshold are reduced to be equal to the threshold. Though the orientation of a single distortion line still can be recognized in the right image, there is no dominant orientation structure due to the superposition of many distortion lines. Note that the gray-value difference among different distortion lines is below 1 percent of the maximal value of the spectrum. If we raise the threshold to 10 percent, the energy contribution of the oriented distortion structure is hardly observable and we get an energy spectrum similar to that in the middle image. Moreover, this structure will be further disturbed by noise. Thus, the distortion cannot be used to reliably identify the occluding velocity.

Fleet and Langley [9] also stated that the orientation of the distortion is only dominant when there is a small number of frequencies with significant power either in the occluding or in the occluded signal. They followed the theta motion model [17] and assumed that the occlusion window moves independently of both occluding and occluded signal (cf. (15) in [9]). In this paper, we follow the idea of Beauchemin and Barron [3] and assume that the occlusion boundary moves consistently with the occluding signal (cf. (2)). This difference is why the orientation of the distortion in

(a) (b) (c)

Fig. 2. (a) and (b) The first and 16th frame of the occlusion sequence. The white window in the 16th frame is centered at $(122, 137)$. Centered at this point, we cut out a cube with $32 \times 32 \times 32$ pixels from the sequence. (c) The epipolar slice of the sequence along row 122. The first frame is at the top of the slice. The occlusion is characterized as two overlapping structures. Both motions are nearly constant (about $(1, 0)$[pixel/frame] for occluding signal and (-1,0)[pixel/frame] for occluded signal).

our model does not depend on the static spectrum of the occluding signal (i.e., $\tilde{I}_1(\omega_x)$). In addition, the shape of the characteristic function $\chi(\mathbf{x})$ is not explicitly described in [17], [9] (cf. (12) in [9]), while Beauchemin and Barron modified this function into a step function (cf. (4) in [3]). We believe that our specification describes the occlusion boundary more explicitly and provides an easier geometric interpretation. It is worth mentioning that Fleet did develop the case of a step function in his earlier work (see (3.17) in [8]). But there is a derivation error in his equation (cf. (13) below).

2.2 2D Occlusion Spectrum

In this section, we extend the above analysis to a 2D occlusion sequence [16]. We only need to replace the $\chi(\mathbf{x})$ in (1) with a 2D Heaviside unit step function $U(\mathbf{x})$

$$U(\mathbf{x}) = \begin{cases} 1 & \mathbf{x}^T\hat{\boldsymbol{\eta}} \geq 0 \\ 0 & \text{otherwise,} \end{cases} \tag{11}$$

where $\mathbf{x}$ denotes 2D spatial Cartesian coordinates and $\hat{\boldsymbol{\eta}}$ is a unit vector normal to the occluding boundary.

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

$$\begin{aligned}\tilde{I}(\mathbf{k}, \omega_t) = \tilde{U}(\mathbf{k})\delta\big(\mathbf{k}^T\mathbf{v}_1 + \omega_t\big) * \tilde{I}_1(\mathbf{k})\delta\big(\mathbf{k}^T\mathbf{v}_1 + \omega_t\big) + \tilde{I}_2(\mathbf{k})\delta\big(\mathbf{k}^T\mathbf{v}_2 + \omega_t\big) \\ - \tilde{U}(\mathbf{k})\delta\big(\mathbf{k}^T\mathbf{v}_1 + \omega_t\big) * \tilde{I}_2(\mathbf{k})\delta\big(\mathbf{k}^T\mathbf{v}_2 + \omega_t\big)\end{aligned} \tag{12}$$

with

$$\tilde{U}(\mathbf{k}) = 2\pi\left[\pi\delta(|\mathbf{k}|) + \frac{\delta\big(\mathbf{k}^T\hat{\boldsymbol{\eta}}_\perp\big)}{i\mathbf{k}^T\hat{\boldsymbol{\eta}}}\right]. \tag{13}$$

Here, $\hat{\boldsymbol{\eta}}_\perp$ denotes a unit vector perpendicular to $\hat{\boldsymbol{\eta}}$. Note that this equation is different from (3.17) in [8], where the coefficient 2π and the Dirac term $\delta(\mathbf{k}^T\hat{\boldsymbol{\eta}}_\perp)$ are missing due to a derivation error.

Substituting (13) into (12) yields

$$\begin{aligned}\tilde{I}(\mathbf{k}, \omega_t) = [2\pi^2\tilde{I}_1(\mathbf{k}) + A(\mathbf{k})]\delta\big(\mathbf{k}^T\mathbf{v}_1 + \omega_t\big) \\ + \big(1 - 2\pi^2\big)\tilde{I}_2(\mathbf{k})\delta\big(\mathbf{k}^T\mathbf{v}_2 + \omega_t\big) + B(\mathbf{k}, \omega_t)\end{aligned} \tag{14}$$

with

$$\begin{cases} A(\mathbf{k}) & = \frac{2\pi}{i\mathbf{k}^T\hat{\boldsymbol{\eta}}}\delta\big(\mathbf{k}^T\hat{\boldsymbol{\eta}}_\perp\big) * \tilde{I}_1(\mathbf{k}) \\ B(\mathbf{k}, \omega_t) & = \frac{i2\pi}{\mathbf{k}^T\hat{\boldsymbol{\eta}}}\delta\big(\mathbf{k}^T\hat{\boldsymbol{\eta}}_\perp\big)\delta\big(\mathbf{k}^T\mathbf{v}_1 + \omega_t\big) * \tilde{I}_2(\mathbf{k})\delta\big(\mathbf{k}^T\mathbf{v}_2 + \omega_t\big). \end{cases} \tag{15}$$

The first two terms in (14) are two oriented spectral planes passing through the origin. Their normal vectors $(u_1, v_1, 1)$ and $(u_2, v_2, 1)$ denote the occluding and the occluded velocity, respectively. The occluding velocity plane is additionally strengthened by the term $A(\mathbf{k})$. The distortion term $B(\mathbf{k}, \omega_t)$ is a convolution between a 3D line and a 3D plane. To get a manifest interpretation of the distortion term, we extend the 1D decomposition used in (7) to 2D space

$$\tilde{I}_2(\mathbf{k}) = \sum_m c_m\delta(\mathbf{k} - \mathbf{k}_m). \tag{16}$$

The distortion term is then reformulated as

$$\begin{aligned}&B(\mathbf{k}, \omega_t) \\ &= \frac{i2\pi}{\mathbf{k}^T\hat{\boldsymbol{\eta}}}\delta\big(\mathbf{k}^T\hat{\boldsymbol{\eta}}_\perp\big)\delta\big(\mathbf{k}^T\mathbf{v}_1 + \omega_t\big) * \sum_m c_m\delta\big(\mathbf{k} - \mathbf{k}_m, \mathbf{k}_m^T\mathbf{v}_2 + \omega_t\big) \\ &= i2\pi\sum_m \frac{c_m}{(\mathbf{k} - \mathbf{k}_m)^T\hat{\boldsymbol{\eta}}}\delta\Big((\mathbf{k} - \mathbf{k}_m)^T\hat{\boldsymbol{\eta}}_\perp, \mathbf{k}^T\mathbf{v}_1 + \omega_t - \mathbf{k}_m^T(\mathbf{v}_1 - \mathbf{v}_2)\Big).\end{aligned} \tag{17}$$

Now, it is clear to see that $B(\mathbf{k}, \omega_t)$ consists of a set of 3D distortion lines with the same orientation formed by $\delta(\mathbf{k}^T\hat{\boldsymbol{\eta}}_\perp)\delta(\mathbf{k}^T\mathbf{v}_1 + \omega_t)$. By setting $\mathbf{k} = \mathbf{k}_m$ in (17), we can prove that each distortion line intersects the occluded plane at $\mathbf{k}_m$. This oriented structure varies with the number of Dirac components in $\tilde{I}_2(\mathbf{k})$ and is therefore not stable. More importantly, the amplitude of each distortion line decreases rapidly after leaving the occluded plane due to the hyperbolic behavior of the term $\frac{2\pi}{(\mathbf{k}-\mathbf{k}_m)^T\hat{\boldsymbol{\eta}}}$. As a result, we cannot use the distortion orientation reliably.

In Fig. 2, we display a real occlusion example, in which a right moving box is covering a left moving picture. Both horizontal motions are nearly constant, as shown in the epipolar slice. The section planes of the spectrum in Fig. 3 also indicate only two horizontally constant motions. The structure of the distortion is indistinct.

In summary, the oriented structure of the distortion in the occlusion spectrum is only observable if the occluded signal has very few spectral components. For occluded signals with many spectral components, this structure is blurred after the superposition of differently located distortion lines. In this sense, we can say that the conclusion of Beauchemin and Barron [3] is a special case of our analysis. Moreover, the influence of the oriented distortion with hyperbolic form is in most spectral regions negligible or only comparable to the influence of noise. The main energy of the occlusion spectrum is still contributed by two spectral lines (in 1D) or planes (in 2D). This fact indicates clearly that we are not able to use the distortion to reliably distinguish the occluding velocity from the occluded one.

2.3 Spectral Multiple Motion Model

Though the oriented structure of the distortion is not a suitable feature to identify the occluding velocity in occlusion analysis, we can still detect and analyze the orientation of dominant energy planes to estimate occluding and occluded velocity. The exception is at low frequencies where the determination of the orientation of spectral planes is more susceptible to distortion than at high frequencies. The recipe we learn from our analysis is that we have

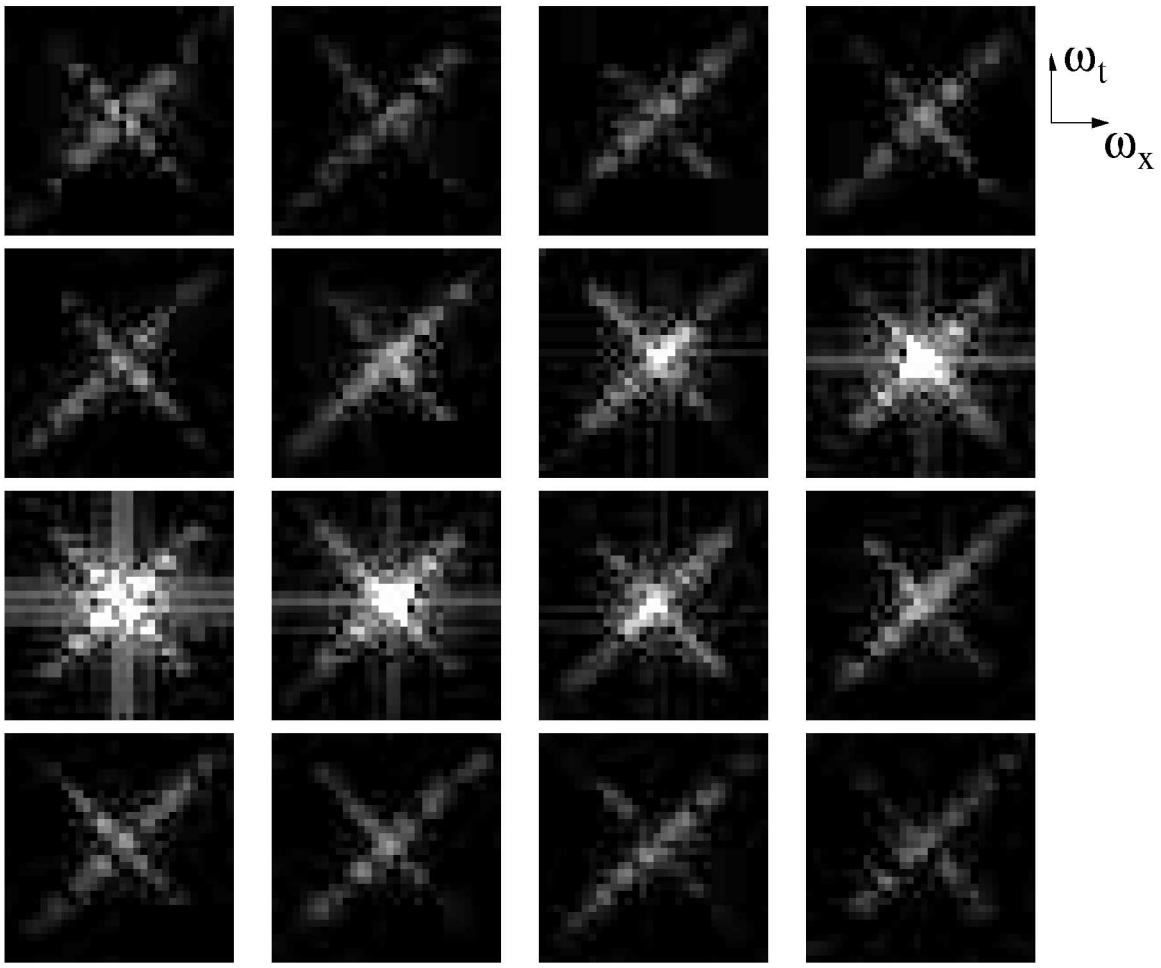

Fig. 3. The (ω_x, ω_t) sections of the $32 \times 32 \times 32$ spectral cube. We only display one of every two section planes. The origin of each section lies in the middle of the image. In row 1 are the 1st, 3rd, 5th, and 7th section (from left to right) and the 9th, 11th, $\cdots$, 31st (ω_x, ω_t) section are arranged similarly in row 2, row 3, and row 4. Two dominant spectral planes indicate two horizontal constant motions vividly. The oriented structure of distortion is indistinct. However, the disturbance of distortion in low frequency regions (see Sections 15, 17, and 19) is apparent.

to consider the spectrum only above a lower bound of the frequency to improve the robustness of motion estimation.

Another advantage of the above knowledge is that we may treat a different kind of multiple motion, namely, additive transparency, in the same manner. We may construct an additive transparency sequence by simply substituting $\chi(\mathbf{x} - \mathbf{v}_1 t)$ in (1) with a real constant $a(a \in (0,\ 1))$ [3]. The corresponding spectrum is then characterized by two oriented planes without distortion

$$\tilde{I}(\mathbf{k}, \omega_t) = a\tilde{I}_1(\mathbf{k})\delta(\mathbf{k}^T\mathbf{v}_1 + \omega_t) + (1-a)\tilde{I}_2(\mathbf{k})\delta(\mathbf{k}^T\mathbf{v}_2 + \omega_t). \quad (18)$$

The additive transparency is therefore very related to the occlusion in the sense that both occlusion and additive transparency are characterized by two dominant spectral planes. Taking into account that the distortion in the occlusion spectrum is negligible or only comparable to noise, we may model both occlusion and additive transparency in the spectral domain as multiple planes passing through the origin. It is known that the corresponding flow vectors are obtained from the normal vectors of these planes. This model can be viewed as a generalization of the spatiotemporal energy model of single motion [1], [10]. This spectral model seems similar to the superposition principle proposed by Shizawa and Mase [15]. However, we would like to mention two noticeable differences:

- Shizawa and Mase proposed that multiple motions are characterized as multiple planes both in the (I_x, I_y, I_t) derivative space and in the frequency domain. We argue that, even though the model is correct in the (I_x, I_y, I_t)-space, it is not realistically applicable in the case of transparency. According to the additive transparency model, the derivative of the intensity profile in a transparency scene is a superposition of two component derivatives derived from functions I_1 and I_2 (cf. (1)). As we cannot decompose the sum of derivatives into two components properly, the estimation of transparency speeds is infeasible in the (I_x, I_y, I_t) derivative space.
- In the case of occlusion, the spectral planes of the occluding and occluded signals are disturbed by the distortion at low frequencies. We have to truncate low-frequency components in order to fit multiple planes robustly.

3 Discussion

Related Multiple Motion Models. Chen et al. attacked the problem of multiple motion estimation using the *Harmonic retrieval framework* [6]. In this framework, occlusion is also treated as noise in the recovering of multicomponent frequencies corresponding to multiple motions. The equivalence between multiple motions and multiple spectral planes was pointed out as well. The authors did not, however, study the explicit structure of the occlusion distortion.

Recently, Langer and Mann [13] categorized image motions according to the dimensionality of both image points and

velocities. They specifically studied one category termed *optical snow*, in which the speeds at each image point form a 1D curve. Occlusion and additive transparency are regarded as a special case of optical snow because each image point at the occlusion boundary or in the transparency has two velocities. If all speeds in the optical snow category happen to have the same direction, the spectral planes corresponding to these motions form a *bow tie* structure in the spectral space

$$\alpha v_x \omega_x + \alpha v_y \omega_y + \omega_t = 0, \tag{19}$$

where $\alpha \in I\!R$ and (v_x, v_y) denote horizontal and vertical motion parameters.

This categorized analysis provides an interesting point of view of image motions. But, as the authors correctly pointed out, it does not give a detailed description of occlusion and additive transparency. Besides, though the spectral bow tie structure in the optical snow category is similar to the multiple plane structure of occlusion or transparency, the motions in the optical snow are constrained to have the same direction. For occlusion analysis, this constraint is too strong.

Numerical Limitations. The good performance of a motion estimation algorithm relies on correct motion modeling. Nevertheless, an ideal model does not guarantee feasibility of the model. In practice, there exists a severe problem in obtaining the energy spectrum of an image sequence due to the block effect of the discrete Fourier transform (DFT). This is an important problem in frequency-based techniques. In order to avoid the block effect of DFT, we take a local Fourier transform (LFT), i.e., a DFT windowed by a Gaussian. According to the convolution theorem in Fourier analysis, a DFT of the image sequence windowed by a spatio-temporal Gaussian is equivalent to a convolution between the spectrum of the image sequence and the spectrum of the Gaussian function, which is also a Gaussian. Consequently, the spectrum is blurred and frequency resolution decreases. For compensation, we have to enlarge the window size so that we can improve the frequency resolution. In the enlarged window, however, we may not be able to approximate the actual motion robustly with a simple motion model, as we can in a smaller window. Moreover, using a larger window means including more time frames in the estimation. On one side, using multiple frames improves the robustness of the optical flow estimation algorithm (e.g., [12]). On the other side, if we include a very long sequence in the estimation, the motion is less likely to be constant over so large time interval.

One possible solution to this problem may be to use motion models with higher order, which still remains a challenging topic [5]. Another promising alternative is to use Gabor-based or wavelets-based approaches (e.g., [7], [11], [14]). An elegant solution to the prohibitive computation cost of these approaches remains to be found.

ACKNOWLEDGMENTS

The authors would like to thank L. Staib for his kind help in correcting the manuscript. They also highly appreciate the reviewers' valuable comments which helped them in clarifying and improving the paper. The financial support provided to W. Yu by DAAD (German Academic Exchange Service) and the US National Institute of Health grant R01-HL-44803-06 (principal investigator: J.S. Duncan) is gratefully acknowledged. G. Sommer was supported by DFG grant 320/1-3. S. Beauchemin was funded by NSERC-R3142A01. K. Daniilidis is grateful for the following support from the US National Science Foundation (NSF): NSF-IIS-0083209, NSF-EIA-0120565, NSF-IIS-0121293, NSF-EIA-9703220, a DARPA/ITO/NGI subcontract to UNC, and a Penn Research Foundation grant.

REFERENCES

[1] E.H. Adelson and J.R. Bergen, "Spatiotemporal Energy Models for the Perception of Motion," *J. Optical Soc. Am.*, vol. 1, no. 2, pp. 284-299, 1985.
[2] S.S. Beauchemin and J.L. Barron, "A Theory of Occlusion in the Context of Optical Flow," *Advances in Computer Vision*, F. Solina, W. Kropatsch, R. Klette, and R. Bajcsy, eds., pp. 191-200, New York: Springer Wien, Nov. 1997.
[3] S.S. Beauchemin and J.L. Barron, "The Frequency Structure of 1D Occluding Image Signals," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 22, pp. 200-206, 2000.
[4] S.S. Beauchemin and J.L. Barron, "On the Fourier Properties of Discontinuous Visual Motion," *J. Math. Imaging and Vision*, vol. 13, no. 3, pp. 155-172, 2000.
[5] W. Chen, G.B. Giannakis, and N. Nandhakumar, "Spatiotemporal Approach for Time-Varying Global Image Motion Estimation," *IEEE Trans. Image Processing*, vol. 5, no. 10, pp. 1448-1461, 1996.
[6] W. Chen, G.B. Giannakis, and N. Nandhakumar, "A Harmonic Retrieval Framework for Discontinuous Motion Estimation," *IEEE Trans. Image Processing*, vol. 7, no. 9, pp. 1242-1257, 1998.
[7] J.G. Daugman, "Complete Discrete 2D Gabor Transforms by Neural Networks for Image Analysis and Compression," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 36, no. 7, 1988.
[8] D.J. Fleet, *Measurement of Image Velocity*. Kluwer Academic Publishers, 1992.
[9] D.J. Fleet and K. Langley, "Computational Analysis of Non-Fourier Motion," *Vision Research*, vol. 34, pp. 3057-3079, 1994.
[10] D.J. Heeger, "Optical Flow Using Spatiotemporal Filters. *Int'l J. Computer Vision*, vol. 1, no. 4, pp. 279-302, 1987.
[11] F. Heitger, L. Rosenthaler, E. Peterhans, and O. Kuebler, "Simulation of Neural Contour Mechanisms: From Simple to End-Stopped Cells," *Vision Research*, vol. 32, no. 5, pp. 963-981, 1992.
[12] B. Jähne, *Spatio-Temporal Image Processing*. Springer-Verlag, 1993.
[13] M.S. Langer and R. Mann, "Dimensional Analysis of Image Motion," *Proc. Int'l Conf. Computer Vision*, vol. I, pp. 155-162, July 2001.
[14] T.S. Lee, "Image Representation Using 2D Gabor Wavelets," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 18, no. 10, pp. 959-971, Oct. 1996.
[15] M. Shizawa and K. Mase, "A Unified Computational Theory for Motion Transparency and Motion Boundaries Based on Eigenenergy Analysis," *IEEE Conf. Computer Vision and Pattern Recognition*, pp. 289-295, June 1991.
[16] W. Yu, K. Daniilidis, S. Beauchemin, and G. Sommer, "Detection and Characterization of Multiple Motion Points," *IEEE Conf. Computer Vision and Pattern Recognition*, vol. I, pp. 171-177, June 1999.
[17] J.M. Zanker, "Theta Motion: A Paradoxical Stimulus to Explore Higher Order Motion Extraction," *Vision Research*, vol. 33, no. 4, pp. 553-569, 1993.