Oriented Structure of the Occlusion Distortion: Is It Reliable?

Corresponding Author: Weichuan Yu
Department of Diagnostic Radiology, Yale University
BML 332, P.O.Box 208042, New Haven, CT 06520-8042, USA
Tel/Fax: +1 203 7857294/7374273, Email: weichuan@noodle.med.yale.edu

Gerald Sommer
Institute of Computer Science, University Kiel
Preusserstrasse 1-9, D-24105 Kiel, GERMANY
Tel/Fax: +49 431 5604-70/-81, Email: gs@ks.informatik.uni-kiel.de

Steven Beauchemin
346 Herzberg Laboratories, Carleton University
1125 Colonel By Drive, Ottawa, ON K1S 5B6, CANADA
Tel/Fax: +1 613 520-2600, ext. 2282/520-4334, Email: beau@csd.uwo.ca

Kostas Daniilidis
GRASP Laboratory, University of Pennsylvania
3401 Walnut Street, Philadelphia, PA 19104-6228, USA
Tel/Fax: +1 215 8988549/5732048, Email: kostas@grasp.cis.upenn.edu
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 rapid decreasing energy contribution and possible superposition of different distortion components, which further blurs the weak orientation structure. We also indicate that the superposition principle of Shizawa and Mase for multiple motion estimation needs to be adjusted.

Index terms: Occlusion, Spectral Analysis, Oriented Structure, Optical Flow

1 Introduction

The motivation of studying multiple motions (including occlusion and additive transparency) in the spectral domain is mainly due to the inability of spatial motion models [3]. The spectrum of multiple motions was first analyzed by Fleet [8]. Beauchemin and Barron [2, 3, 4] formulated in the frequency domain an explicit model of occlusion. 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. Here we point out that the orientation of the distortion term cannot be used reliably to identify an arbitrary occluding velocity. Next we start with the spectral analysis of a 1D occlusion sequence. The conclusion of Beauchemin and Barron [3] is proved 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 come to 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 [14] needs to be adjusted. This manuscript finally discusses some other related works 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 \( \chi(x) \) and modeled the occlusion in the spatial domain as

\[
I(x, t) = \chi(x - v_1 t) I_1(x - v_1 t) + [1 - \chi(x - v_1 t)] I_2(x - v_2 t),
\]

where \( I_1(x) \) is a 2D occluding signal moving with a velocity \( v_1 = (v_{1x}, v_{1y})^T \) and \( I_2(x) \) is a 2D occluded signal moving with a velocity \( v_2 = (v_{2x}, v_{2y})^T \). This equation is the seed of all other occlusion models which come in succession. For simplicity we focus on the constant velocity models in this paper.

2.1 1D Occlusion Spectrum

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

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

with

\[
u(x) = \begin{cases} 1 & x \geq 0 \\ 0 & \text{otherwise} \end{cases}.
\]

The Fourier transform of equation (2) reads

\[
\hat{I}(\omega_x, \omega_t) = \hat{u}(\omega_x) \delta(\omega_x v_1 + \omega_t) * \hat{I}_1(\omega_x) \delta(\omega_x v_1 + \omega_t) + \hat{I}_2(\omega_x) \delta(\omega_x v_2 + \omega_t)
\]

\[
-\hat{u}(\omega_x) \delta(\omega_x v_1 + \omega_t) * \hat{I}_2(\omega_x) \delta(\omega_x v_2 + \omega_t),
\]

where

\[
\hat{u}(\omega_x) = \pi \delta(\omega_x) + \frac{1}{i\omega_x}.
\]

Here \(*\) means convolution and \( \hat{\cdot} \) denotes the Fourier transform of the corresponding signal.

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

$$
\tilde{I}(\omega_x, \omega_l) = [\pi \tilde{I}_1(\omega_x) + \frac{1}{i \omega_x} \tilde{I}_1(\omega_x)] \delta(\omega_x v_1 + \omega_l) + (1 - \pi) \tilde{I}_2(\omega_x) \delta(\omega_x v_2 + \omega_l)
$$

$$
+ \frac{i}{\omega_x} \delta(\omega_x v_1 + \omega_l) * \tilde{I}_2(\omega_x) \delta(\omega_x v_2 + \omega_l).
$$

(6)

The first two terms in equation (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 equation (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

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

(7)

where $m$ indexes the components and $c_m$ denotes the coefficients. This reformulation makes no difference in representing signals (cf equation (13) in [3]). After this reformulation the distortion term then reads

$$
\frac{i}{\omega_x} \delta(\omega_x v_1 + \omega_l) * \tilde{I}_2(\omega_x) \delta(\omega_x v_2 + \omega_l) = \sum_m \frac{i c_m}{\omega_x - \omega_m} \delta(\omega_x v_1 + \omega_l - \omega_m(v_1 - v_2)).
$$

(8)

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

This orientation structure formed by the distortion lines characterizes the occlusion spectrum and many approaches (e.g. [9, 3]) 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, which is common for usual real signals, this kind of orientation may disappear due to the superposition (see figure 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.
The Theorem 1 in [3] (cf equation (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. The Theorem 2 in [3] left this superposition effect out of account and turned out to be a specific conclusion only. We may explain this point by demonstrating that equation (11) in [3] is only a special case of equation (6). Using the same cosine functions in Theorem 1 as occluding and occluded signal
\[
\begin{align*}
I_1(x - v_1 t) &= c_1 \cos(\omega_1 (x - v_1 t)) \\
I_2(x - v_2 t) &= c_2 \cos(\omega_2 (x - v_2 t)),
\end{align*}
\]
we have the corresponding spectra
\[
\begin{align*}
I_1(\omega_x) \delta(\omega_x v_1 + \omega_t) &= \frac{c_1}{2} \left[ \delta(\omega_x - \omega_1, \omega_t + \omega_1 v_1) + \delta(\omega_x + \omega_1, \omega_t - \omega_1 v_1) \right] \\
&= \frac{c_1}{2} \delta(\omega_x - \omega_1, \omega_t + \omega_1 v_1) + \frac{c_1}{2} \delta(\omega_x + \omega_1, \omega_t - \omega_1 v_1) \\
I_2(\omega_x) \delta(\omega_x v_2 + \omega_t) &= \frac{c_2}{2} \left[ \delta(\omega_x - \omega_2, \omega_t + \omega_2 v_2) + \delta(\omega_x + \omega_2, \omega_t - \omega_2 v_2) \right] \\
&= \frac{c_2}{2} \delta(\omega_x - \omega_2, \omega_t + \omega_2 v_2) + \frac{c_2}{2} \delta(\omega_x + \omega_2, \omega_t - \omega_2 v_2) .
\end{align*}
\]
Substituting (9) into (6) yields
\[
\tilde{I}(\omega_x, \omega_t) = \frac{\pi}{2} c_1 \left[ \delta(\omega_x - \omega_1, \omega_t + \omega_1 v_1) + \delta(\omega_x + \omega_1, \omega_t - \omega_1 v_1) \right] \\
- \frac{\pi}{2} c_1 \left[ \frac{1}{\omega_x - \omega_1} \delta(\omega_x v_1 + \omega_t) + \frac{1}{\omega_x + \omega_1} \delta(\omega_x v_1 + \omega_t) \right] \\
+ \frac{\pi}{2} c_2 \left[ \delta(\omega_x - \omega_2, \omega_t + \omega_2 v_2) + \delta(\omega_x + \omega_2, \omega_t - \omega_2 v_2) \right] \\
+ \frac{\pi}{2} c_2 \left[ \frac{1}{\omega_x - \omega_2} \delta(\omega_x v_1 + \omega_t - \omega_2 (v_1 - v_2)) + \frac{1}{\omega_x + \omega_2} \delta(\omega_x v_1 + \omega_t + \omega_2 (v_1 - v_2)) \right].
\]
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 figure 1 we display a 1D random dot occlusion sequence smoothed by a Gaussian window and the corresponding energy spectrum. The spectrum is characterized as 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
Figure 1: **Left:** Gaussian windowed 1D random dot occlusion sequence. The occluding velocity is 1 [pixel/frame] and the occluded velocity is −1 [pixel/frame]. **Middle:** The energy spectrum of the occlusion sequence. For displaying we delete the DC component. The occluding and the occluded signal are clearly characterized as 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. **Right:** In order to display the superposition effect of distortion lines, we set a threshold equal to 1% 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 more indistinct the distortion structure is.

the superposition effect of the distortion, we set a threshold equal to 1% 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 more dominant orientation structure due to superposition of many distortion lines. Note that the gray-value difference among different distortion lines is below 1% of the maximal value of the spectrum. If we raise the threshold to 10%, 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 reliably to 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 [16] and assumed that
the occlusion window moves independently of both occluding and occluded signal (cf equation (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 equation (2)).
This difference is the reason why the orientation of the distortion in our model does not
depend on the static spectrum of the occluding signal (i.e. \( \vec{I}_1(\omega_z) \)). Besides, the shape of the
characteristic function \( \chi(x) \) is not explicitly described in [16, 9] (cf. equation (12) in [9]),
while Beauchemin and Barron modified this function into a step function (cf. equation (4)
in [3]). We believe that this specification describes the occlusion boundary more explicitly
and makes the geometric imagination more easily.

2.2 2D Occlusion Spectrum

In this subsection we extend the above analysis to a 2D occlusion sequence [15]. We only
need to replace the \( \chi(x) \) in (1) with a 2D Heaviside unit step function \( U(x) \)
\[
U(x) = \begin{cases} 
1 & x^T \hat{n} \geq 0 \\
0 & \text{otherwise}
\end{cases},
\]
where \( x \) denotes 2D spatial Cartesian coordinates and \( \hat{n} \) is a unit vector normal to the
occluding boundary.

We denote the spatial frequency vector as \( k = (\omega_x, \omega_y)^T \) and the temporal frequency as
\( \omega_t \). Then, the Fourier transform of the image sequence reads
\[
\vec{I}(k, \omega_t) = \vec{U}(k)\delta(k^T v_1 + \omega_t) * \vec{I}_1(k)\delta(k^T v_1 + \omega_t) + \vec{I}_2(k)\delta(k^T v_2 + \omega_t) \\
- \vec{U}(k)\delta(k^T v_1 + \omega_t) * \vec{I}_2(k)\delta(k^T v_2 + \omega_t)
\]
with
\[
\vec{U}(k) = 2\pi [\pi\delta(|k|) + \frac{\delta(k^T \hat{n}_\perp)}{ik^T \hat{n}}].
\]
Here \( \hat{n}_\perp \) denotes a unit vector perpendicular to \( \hat{n} \). Note that this equation is different to
(3.17) in [8], where the coefficient \( 2\pi \) and the Dirac term \( \delta(k^T \hat{n}_\perp) \) are missing due to a
derivation error.
Substituting (13) into (12) yields
\[
\tilde{I}(k, \omega_t) = [2\pi^2 \tilde{I}_1(k) + A(k)]\delta(k^T v_1 + \omega_t) \\
+ (1 - 2\pi^2) \tilde{I}_2(k)\delta(k^T v_2 + \omega_t) + B(k, \omega_t)
\]  
(14)

with
\[
\begin{cases}
A(k) &= \frac{2\pi}{ikT\eta} \delta(k^T \hat{\eta}_\perp) * \tilde{I}_1(k) \\
B(k, \omega_t) &= \frac{i2\pi}{kT\eta} \delta(k^T \hat{\eta}_\perp) \delta(k^T v_1 + \omega_t) * \tilde{I}_2(k)\delta(k^T v_2 + \omega_t).
\end{cases}
\]  
(15)

The first two terms in expression (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 \(A(k)\). The distortion term \(B(k, \omega_t)\) 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 equation (7) to 2D space
\[
\tilde{I}_2(k) = \sum_m c_m \delta(k - k_m).
\]  
(16)

The distortion term is then reformulated as
\[
B(k, \omega_t) = \frac{i2\pi}{kT\eta} \delta(k^T \hat{\eta}_\perp) \delta(k^T v_1 + \omega_t) * \sum_m c_m \delta(k - k_m, k_m^T v_2 + \omega_t)
\]
\[
= i2\pi \sum_m \frac{c_m}{(k - k_m)^T \eta} \delta((k - k_m)^T \hat{\eta}_\perp, k_m^T v_1 + \omega_t - k_m^T (v_1 - v_2)).
\]  
(17)

Now it is clear to see that \(B(k, \omega_t)\) consists of a set of 3D distortion lines with the same orientation formed by \(\delta(k^T \hat{\eta}_\perp) \delta(k^T v_1 + \omega_t)\). Each distortion line intersects the occluded plane at \(k_m\) (we can prove it by setting \(k = k_m\) in equation (17)). This oriented structure varies with the number of Dirac components in \(\tilde{I}_2(k)\) and is therefore not stable. More importantly, the amplitude of each distortion line decreases rapidly after leaving the occluded plane due to the hyperbolic property of the term \(\frac{2\pi}{(k - k_m)^T \eta}\). As a result, we cannot use the distortion orientation reliably.

In figure 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 figure 3 also indicate only two horizontally constant motions. The structure of the distortion is indistinct.
Figure 2: **Left and Middle:** The first and 16-th frame of the occlusion sequence. The white window in the 16-th 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. **Right:** 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).

In summary, the oriented structure of the distortion in 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 meaning we 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 nature is in most spectral regions negligible or only comparable to the influence of noise. The main energy proportion of occlusion spectrum is still on the two spectral lines/planes. This fact indicates clearly that we are not able to use the distortion reliably to 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 still can 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. As a recipe, we may 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
Figure 3: The \((\omega_x, \omega_t)\) 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, 5-th, and 7-th section (from left to right) and the 9-th, 11-th, \cdots, 31-th \((\omega_x, \omega_t)\) sections 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. But the disturbance of distortion in low frequency regions (see section 15, 17, and 19) is apparent.

multiple motions namely additive transparency in the same manner. We may construct an additive transparency sequence by simply substituting \(\chi(x - v_1t)\) in (1) with a real constant \(a (a \in (0, 1))\) \[3\]. The corresponding spectrum is then characterized by two oriented planes
without distortion
\[
\hat{I}(k, \omega_l) = a\hat{I}_1(k)\delta(k^T v_1 + \omega_l) + (1 - a)\hat{I}_2(k)\delta(k^T v_2 + \omega_l).
\] (18)

The additive transparency is therefore very related to the occlusion in the sense that both
occlusion and additive transparency are characterized with two dominant spectral planes.
Taking into account that the distortion in the occlusion spectrum is negligible or only com-
parable to noise, we may model both occlusion and additive transparency in the spectral
domain as multiple planes passing through the origin. The corresponding motion speeds are
described by the normal vectors of these planes. This model can be viewed as a generaliza-
tion of the spatiotemporal energy model of single motion [1, 10]. This spectral model seems
similar to the superposition principle proposed by Shizawa and Mase [14]. But we would
like to mention two different points:

- 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
  the model is correct in the \((I_x, I_y, I_t)\)-space, it is not feasible in case of transparency
  since we can hardly assume that the intensity profile is differentiable.

- In case of occlusion, the spectral planes of occluding and occluded signal are disturbed
  by the distortion at low frequencies. We have to truncate low frequency components
  in order to fit multiple planes robustly.

3 Discussions

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

Recently, Langer and Mann categorized image motions according to the dimensionality
of both image points and velocities. They specifically studied one category named optical
snow, in which the speeds at each image point form a 1-D curve. The occlusions and additive transparency are regarded as a special case of the optical snow category because each image point at occlusion boundary or in 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 then form a bow tie structure in the spectral space

\[ \alpha v_x \omega_x + \alpha v_y \omega_y + \omega_z = 0, \]

(19)

where \( \alpha \in IR \) 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 the correct model of motions. Nevertheless, an ideal model does not guarantee that this model is feasible. 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 actually one main problem in the frequency-based techniques. In order to avoid the block effect of DFT, we take a local Fourier transform (LFT), i.e. Gaussian windowed DFT. According to the convolution theorem in the Fourier analysis, the Gaussian windowed DFT of the image sequence is equivalent to the 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 we suffer under the limited resolution. For compensation, we have to enlarge the window size so that we can improve the frequency resolution with finer interval. But in the enlarged window, we may no more be able to approximate the actual motion robustly with a simple motion model, as we used to do in a smaller window. Moreover, using a larger window means including more 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 into estimation, it is also fragile
to assume that the motion is constant over so large time interval.

One possible solution to this problem may be using motion models with higher order, which still remains a challenging topic [5]. Another promising alternative is using Gabor based or wavelets based approaches (e.g. [7, 11, 13]). But we still have to find out an elegant solution to reduce the prohibitive computation cost of these approaches.

References


