Motion estimation of opaque or transparent objects using triads of Gabor filters

A. Spinéi\textsuperscript{a}, D. Pellerin\textsuperscript{a, b, *}

\textsuperscript{a}Laboratoire des Images et des Signaux, Institut National Polytechnique de Grenoble, 46 Av. Félix Viallet, 38031 Grenoble Cedex, France
\textsuperscript{b}Institut des Sciences et Techniques, Université Joseph Fourier, Grenoble, France

Received 19 July 1999; received in revised form 27 July 2000

Abstract

A new energy-based method for motion estimation of opaque or transparent objects is described, which uses a reduced number of spatio-temporal Gabor filters. This method is based on the simplifying hypothesis of additive transparency and it copes very well with two specific cases: transparent objects in motion over a fixed background, or over a background moving at the same speed but going in opposite directions. © 2001 Elsevier Science B.V. All rights reserved.

Zusammenfassung

Wir beschreiben eine neue, auf Energie-Messung beruhende Methode zur Bestimmung der Bewegung opaker oder transparenter Objekte. Diese Methode beruht auf der Benutzung einer reduzierten Anzahl von raum-zeitlichen Gabor-Filtern. Sie stützt sich auf die vereinfachende Hypothese additiver Transparenz und kann auf zwei Spezialfälle angewendet werden: Transparente, bewegte Objekte vor festem Hintergrund oder vor einem Hintergrund, der sich mit fester Geschwindigkeit in entgegengesetzter Richtung bewegt. © 2001 Elsevier Science B.V. All rights reserved.

Résumé

Nous décrivons une nouvelle méthode énergétique d’estimation du mouvement d’objets opaques ou transparents, qui utilise un nombre réduit de filtres spatio-temporels de Gabor. Cette méthode repose sur l’hypothèse simplificatrice d’une transparence additive et s’applique dans deux cas particuliers: objets transparents en mouvement sur un fond fixe, ou sur un fond se déplaçant à la même vitesse mais avec une direction opposée. © 2001 Elsevier Science B.V. All rights reserved.

Keywords: Transparent motion estimation; Energy-based method; Gabor filters; Psychophysical results

*Correspondence address. Laboratoire des Images et des Signaux, Institut National Polytechnique de Grenoble, 46 Av. Félix Viallet, 38031 Grenoble Cedex, France. Tel.: + 33-4-76-57-43-69; fax: 33-4-76-57-47-90.
E-mail address: pellerin@lis-viallet.inpg.fr (D. Pellerin).
1. Introduction

Motion transparency consists in the superposition of at least two “images”, moving at independent speeds in the same region of the visual field, which leads to the perception of more than one velocity field. This phenomenon appears in a scene which contains transparent objects such as a window, opaque objects which create diffuse reflections, or fragmented occlusions by objects such as trees and fences. Shadows are also a widespread example of transparency. The analysis and understanding of motion transparency is rather complex and traditional motion estimation techniques are not able to cope with them.

Several methods were perfected in order to estimate transparent motion or more generally multiple motion. Some algorithms [1,3] based on regression techniques, are able to compute a global multiple motion for the entire image, but use local measurements that are only capable of determining a single-velocity estimate at each point of the image. Energy-based models [6] inspired by psycho-physical experimentations are able to detect motion transparency and to estimate only certain parameters such as the direction of two motions. Finally, a generalization of the constraint equation of optical flow has been proposed [7]. This technique presents two major drawbacks: it has only been tested with random textures and it requires high-order derivatives and a large number of calculations (convolution by 376 spatio-temporal Gabor filters).

In this paper, we shall present an energy-based model for transparent motion estimation, using a small number of spatio-temporal Gabor filters. Our model copes very well with two specific cases of motion transparency, by allowing the detection of transparent regions and the estimation of the speed in each region of the image. We shall begin by presenting the frequency patterns induced by opaque and transparent motions (Section 2). Later on, we shall briefly describe a model for the estimation of opaque motion based on triads of Gabor filters (Section 3), then we shall extend this model to motion transparency (Section 4). We shall present the general ideas behind our algorithm in Section 5 and we shall analyse two cases of motion transparency (Sections 6 and 7).

2. Frequency patterns induced by motion

Opaque motion: Energy-based methods for velocity estimation in image sequences are based on the fact that motion induces specific energy patterns in the frequential domain. The power spectrum, associated with an image \( i(x, y) \) in translation with a uniform velocity \( v = (v_x, v_y) \) occupies a plane in the spatio-temporal frequency domain whose equation is \( f_t + v_xf_x + v_yf_y = 0 \), where \( f_x, f_y \) represent spatial frequencies and \( f_t \) the temporal frequency. This plane passes through the origin and has a slope proportional to the speed. The principle of energy-based methods consists in localizing the non-zero energy plane or velocity plane using oriented spatio-temporal filters, in order to directly obtain the components of the velocity vector \((v_x, v_y)\).

Transparent motion: Two fundamental transparency types were distinguished: multiplicative and additive transparencies [4]. In this article, we are interested in additive transparency. Thus, for the sum of two images \( i_1(x, y) \) and \( i_2(x, y) \) in translation with uniform velocities \( v_1 = (v_{1x}, v_{1y}) \) and \( v_2 = (v_{2x}, v_{2y}) \), the energy is concentrated in two planes whose equations are \( f_t + v_{1x}f_x + v_{1y}f_y = 0 \) and \( f_t + v_{2x}f_x + v_{2y}f_y = 0 \).

3. Opaque motion estimation using triads of Gabor filters

Spatio-temporal pass-band filters such as Gabor energy filters are not selective to velocities, but rather are tuned to particular spatio-temporal frequencies: therefore, on their own, they do not allow velocity estimation. We demonstrated an effective way of combining their responses in order to estimate the motion. A cosine (or even) phase 3D Gabor filter is represented as a spatio-temporal Gaussian window multiplied by a cosine wave

\[
G_c(x, y, t) = \cos[2\pi(f_{sx}x + f_{sy}y + f_{st}t)] \frac{1}{\sqrt{2\pi \sigma_x \sigma_y \sigma_t}} \times \exp \left[-\left(\frac{x^2}{2\sigma_x^2} + \frac{y^2}{2\sigma_y^2} + \frac{t^2}{2\sigma_t^2}\right)\right],
\]

...
where \((f_{x_0}, f_{y_0}, f_{z_0})\) is the central frequency of the filter and \((\sigma_x, \sigma_y, \sigma_z)\) is the standard deviation of the spatio-temporal Gaussian window. The sum of the squared output of the cosine-phase filter and the squared output of the sine-phase filter gives a measure of energy that is independent of the phase of the signal. The energy response to a moving random texture conforming to the white noise hypothesis (the texture has an uniform energy spectrum) of such a 3D Gabor filter [2] is

\[
R_{f_{z_0}} = \frac{k^2}{8\pi \sqrt{\delta}} \exp\left[ -\frac{4\pi^2 \sigma_x^2 \sigma_y^2 \sigma_z^2}{\delta} (v_x f_{x_0} + v_y f_{y_0} + f_{z_0})^2 \right]
\]

with \(\delta = (v_x \sigma_x \sigma_z)^2 + (v_y \sigma_y \sigma_z)^2 + (\sigma_x \sigma_y)^2\) and \(k\) a normalization constant (the energy due to the local image contrast).

The speed component \(v_x\) can be computed directly using the energy responses of three Gabor filters (a triad) with identical bandwidths (the standard deviations are \(\sigma_t\) in temporal and \(\sigma_s = \sigma_x = \sigma_y\) in spatial), centered on the same spatial frequency \(f_{x_0}\) = 0.25 and \(f_{y_0}\) = 0 (noted \(\phi = 0\) in polar coordinates) but placed on three different temporal frequencies \(f_{z_0} = -0.25; 0\) and \(0.25\) (Fig. 1(a)). Actually, the energy responses of the three Gabor filters could be detailed as

\[
\begin{align*}
R_{-0.25} &= C \alpha^{(1-v_x)^2}, \\
R_0 &= C \alpha^{v_x^2}, \\
R_{0.25} &= C \alpha^{(1+v_x)^2}
\end{align*}
\]

with the notations

\[
C = \frac{k^2}{8\pi \sqrt{\delta}}, \quad D = \frac{\pi^2 \sigma_z^2}{4}.
\]

\[
\alpha = \exp\left[ -\frac{D}{v_x^2 + v_y^2 + (\sigma_x / \sigma_z)^2}\right].
\]

We call “energy coefficient” the variable \(C\) which depends on the local image contrast. We can deduce, by combining the preceding equations, the expression of the speed component

\[
v_x = \frac{1L_{-0.25} - L_{0.25}}{6L_0}
\]

with the notation \(L_{f_{z_0}} = \ln R_{f_{z_0}} - (\ln R_{-0.25} + \ln R_0 + \ln R_{0.25})/3\).

In the case of a 2D movement, two-speed components are theoretically sufficient (white noise hypothesis). The speed component \(v_y\) can be computed in the same manner by using a second triad of Gabor filters centered on the spatial frequency \(f_{x_0} = 0\) and \(f_{y_0} = 0.25\) (noted \(\phi = \pi/2\) in polar coordinates) (Fig. 1(b)). This method is described in detail in [8,9].

4. Motion transparency estimation

The response \(R_{f_{z_0}}\) of a Gabor filter is computed using a specific frequential pattern: an energy plane for 2D motion (or an energy line for 1D motion)
under a white noise hypothesis. In the case of additive transparent motions, the response of a Gabor filter is the sum of the contributions of each motion plane (Fig. 1(a)). If the speeds of the two transparent surfaces are $v_1 = (v_{1x}, v_{1y})$ and $v_2 = (v_{2x}, v_{2y})$, the responses of the two triads of filters oriented at $\phi = 0$ and $\pi/2$ (with respectively $a = x$ and $y$) can be written as

\[
\begin{align*}
R_{0, 0.25a} &= C_{1a}x_1^{(1-v_1a)} + C_{2a}x_2^{(1-v_2a)}, \\
R_{0a} &= C_{1a}x_1^a + C_{2a}x_2^a, \\
R_{0, 25a} &= C_{1a}x_1^{(1+v_1a)} + C_{2a}x_2^{(1+v_2a)}
\end{align*}
\]

with the notations

\[
C_{1a} = \frac{k_{1a}^2}{8\pi\sqrt{\delta}}, \quad C_{2a} = \frac{k_{2a}^2}{8\pi\sqrt{\delta}}, \quad D = \frac{\pi^2 \sigma^2}{4},
\]

\[
x_1 = \exp\left[-\frac{D}{v_{1x}^2 + v_{1y}^2 + (\sigma_v/\sigma_i)^2}\right],
\]

\[
x_2 = \exp\left[-\frac{D}{v_{2x}^2 + v_{2y}^2 + (\sigma_v/\sigma_i)^2}\right].
\]

We introduced energy coefficients $C_{1a}, C_{2a}$ and $C_{1a}, C_{2a}$, proportional to the local contrast in the image, in the $Ox$ or $Oy$ direction, in order to increase the robustness of our approach when the white noise hypothesis is not satisfied.

Therefore, in the $Ox$ direction (and with the same reasoning for the perpendicular direction), we have four variables: the two speeds $v_{1x}, v_{2x},$ and the two energy coefficients $C_{1x}$ and $C_{2y}$. A unique solution cannot be obtained in our case, because we dispose of only 3 equations (a triad of filters) for 4 variables. In order to reach a solution, we shall restrict our study to two cases of transparent motion:

- **Only one moving “image”:** This is a very common situation in natural scenes, with transparent objects moving over a fixed opaque background, or inversely opaque moving objects seen through a fixed transparent textured surface or its equivalent (a window with a reflected image, a fence, etc.).

- **Two “images” with opposite speeds:** In this case, we have two moving objects at the same speed but going in opposite directions. This particular case will allow the comparison of our model’s performance with psychophysical results obtained by human subjects.

It is obvious that, even if the filters used in the case of transparent and opaque motion are identical, the interpretation of their energy responses is radically different. For a better understanding, we shall introduce some specific terminology. We will call an “opaque triad” a triad for which responses were processed using the estimation algorithm described in [9], in order to compute the speed for an opaque motion. A “transparent triad” is a triad identical to the previous one, but the responses are processed in a different way in order to cope with transparency.

5. Opaque and transparent motion estimation

We shall use a specific mechanism to combine the information obtained from an opaque triad and from a transparent triad (as a matter of fact, the filtering will be done only once). The two estimations are independent. The estimation with the transparent triad allows the identification (detection) of the “transparent regions” (Fig. 2) which
satisfy the energy model of two different motion planes. In order to obtain the final optical flow, we combine the opaque triad results for the opaque regions with the transparent triad results for the transparent regions. We will use a “transparent/opaque” criterion, which will be described in the next section.

6. Case of only one moving “image”

6.1. Computing the speed

In the particular case \( \mathbf{v}_2 = (0,0) \), which can be assimilated with the very common case of fixed background, the expression of \( \alpha_2 \) (Eq. (1)) becomes

\[
\alpha_2 = \exp \left[ -\frac{D}{(\sigma_x/\sigma_t)^2} \right]
\]

Knowing that \( \mathbf{v}_1 = (v_{1x}, v_{1y}) = (v_x, v_y) \) and \( \mathbf{v}_2 = (v_{2x}, v_{2y}) = (0,0) \), we may express the responses of the two triads of filters oriented at \( \phi = 0 \) and \( \pi/2 \) (with resp. \( a = x \) and \( y \)):

\[
\begin{align*}
R_{-0.25a} &= C_{1a}\alpha_1^{(1-v_x)^2} + C_{2a}\alpha_2, \\
R_{0a} &= C_{1a}\alpha_1^{v_x^2} + C_{2a}, \\
R_{0.25a} &= C_{1a}\alpha_1^{(1+v_y)^2} + C_{2a}\alpha_2.
\end{align*}
\]

In order to eliminate the energy coefficients \( C_{1x}, C_{2x}, C_{1y}, \) and \( C_{2y} \), we define \( T \) as

\[
T = \frac{\alpha_2}{\alpha_1} = \exp \left[ -\frac{D}{(\sigma_x/\sigma_t)^2} \right] \left( \frac{D}{v_x^2 + v_y^2 + (\sigma_x/\sigma_t)^2} \right)
\]

\[
\approx 1.
\]

We shall demonstrate that this rather rough approximation leads to satisfactory results, which will be a starting point for an iterative process, guiding the computation toward precise estimations.

We may define for the two triads of filters oriented at \( \phi = 0 \) and \( \pi/2 \) (with resp. \( a = x \) and \( y \)):

\[
E_a = \frac{R_{-0.25a} - \alpha_2 R_{0a}}{R_{0.25a} - \alpha_2 R_{0a}} = \frac{\alpha_1^{2v_x - T}}{\alpha_1^{2v_x} - T}.
\]

In this way, we find two quadratic equations with the variable \( \alpha_1^{2v_x} \). One more shorthand notation will be used: \( \beta_a = \alpha_1^{2v_x} \). The following solution will be obtained:

\[
\beta_a = \begin{cases} 
\frac{T(E_a - 1) + \sqrt{T^2(1 - E_a)^2 + 4E_a}}{2E_a} & \text{if } v_a \geq 0, \\
\frac{T(E_a - 1) - \sqrt{T^2(1 - E_a)^2 + 4E_a}}{2E_a} & \text{if } v_a < 0.
\end{cases}
\]

Moreover, we may write

\[
\ln(\beta_x) = \frac{D2v_x}{v_x^2 + v_y^2 + (\sigma_x/\sigma_t)^2}
\]

and

\[
\ln(\beta_y) = \frac{D2v_y}{v_x^2 + v_y^2 + (\sigma_x/\sigma_t)^2}
\]

and then

\[
v_x^2 \left[ \ln(\beta_x) + \left( \frac{\ln(\beta_y)}{\ln(\beta_x)} \right)^2 \right] - 2Dv_x + \left( \frac{\sigma_x}{\sigma_t} \right)^2 2\ln(\beta_x) = 0
\]

which is a quadratic equation in \( v_x \). The discriminant is noted \( \Delta \). So the solution could be written as

\[
v_x = \frac{2D - \sqrt{\Delta}}{2[\ln(\beta_x) + (\ln(\beta_y)/\ln(\beta_x))]}.
\]

The solution obtained in Eq. (3) is not an exact solution because we used an approximation (Eq. (2)) in our calculation. It is necessary to reintegrate \( v_x \) and \( v_y \) in the expression of \( \alpha_1 \) and to compute the new values of \( \beta_x \) and \( \beta_y \), leading to more precise estimations for \( v_x \) and \( v_y \). We demonstrated via numerical computation that only one iteration is enough for a good estimation (error \( \ll 1\% \)).

6.2. Principle of region discrimination

In order to separate transparent regions from opaque regions, we shall use the estimated speed to compute the energy coefficients \( C_{1x} \) and \( C_{2x} \) in direction \( Ox \) for each of the transparent moving surfaces (we shall take a similar approach for the energy coefficients \( C_{1y} \) and \( C_{2y} \) in the \( Oy \) direction):

\[
C_{1x} = \frac{R_{-0.25x} - R_{0.25y}}{\alpha_{1-\epsilon_x} + \alpha_{1+\epsilon_x}}.
\]
and

\[ C_{2x} = R_{0x} - C_{1x}^2. \]

If the two energies have close values, we may safely assume that we are in the presence of transparency. On the contrary, it is well known that two superposed images, with very different contrasts, could not induce the transparency perception to an observer. Moreover, if one of the two energies is very small, it is probably due to noise. In this case, we are in the presence of opaque motion. In this way, we are able to separate the transparent and the opaque motion. We defined an energy similarity measure \( S_x \) in direction \( O_x \) (\( S_y \) in direction \( O_y \) is computed in the same manner) which expresses the imbalance between the two contrasts:

\[ S_x = \frac{C_{1x}}{C_{2x}} + \frac{C_{2x}}{C_{1x}}. \]  

(4)

For an “ideal” transparency between two images with the same contrast, \( S_x = 2 \). However, it is obvious that the ratio between \( C_{1x} \) and \( C_{2x} \) could be rather high, so a fairly loose threshold must be established. Experimentally, we found that a threshold of \( S_x = 20 \) is very robust in the case of artificial image sequences containing white spectra transparent objects and background with similar contrasts. The thresholding is carried out on each of the two orthogonal triads used for the 2D motion estimation and the transparency region is the gathering of points detected in transparent motion for the two triads.

6.3. Results

We chose for the spatio-temporal Gabor filters the following standard deviations: \( \sigma_x = \sigma_y = 4 \) in spatial and \( \sigma_t = 1 \) in temporal. We used an efficient recursive implementation of the energy Gabor filters [9] which allows a reduction (1:4) of the amount of operations needed for the filtering. In this way, the motion estimation and the region discrimination require only 132 elementary operations/pixel (additions, multiplications).

Random texture sequence: The fixed background is a matrix of 128 × 128 pixels covered with a white noise texture. Two squares of size 20 × 20 pixels are moving on the background: one of them is transparent and it has a translation speed of \( v_{\text{obj1}} = (1, -1) \) pixels/image, the other one is opaque and it has a different speed, for example \( v_{\text{obj2}} = (-1, 1) \) pixels/image (Fig. 3(a)). With the method of opaque motion estimation (only one speed for each pixel of the image) (Fig. 3(b)), the averages and the standard deviations of the estimated optical flow were, respectively (−0.97,0.98) and (0.08,0.08) on the opaque square, (0.46,−0.48) and (0.11,0.09) on the transparent square. The speed is correctly estimated for the opaque object, but it is largely underestimated in the case of the transparent object because of the influence of the fixed background. With our method of opaque and transparent motion estimation (Fig. 3(c)), the average and the standard deviation of the estimated optical flow were, respectively (0.96,−0.97) and (0.08,0.08) on the transparent square. We notice...
a quality discrimination and a good estimation. We observe also that on the border of the opaque object, the occlusion region is assimilated to an additive transparency.

**Natural appearance sequence**: We chose to create an artificial sequence of images with natural appearance (or spectrum). We used two images ("clown" and "mandrill" which are included in the standard package of Matlab computing software. One of the two images is the background and the other one is the transparent moving object. We re-sampled these images: 128×128 pixels for the background and 40×25 for the transparent object. The transparent rectangle has a translation speed of \( v = \frac{1}{2} \log \left( \frac{\gamma_x + \sqrt{\gamma_x^2 - 4}}{2} \right) \) for \( v > 0 \),

\[
  v_x = \frac{1}{2} \log \left( \frac{\gamma_x - \sqrt{\gamma_x^2 - 4}}{2} \right) \quad \text{for} \quad v_x < 0
\]

with the notation \( \gamma_x = (R_{0.25x} + R_{-0.25x})/R_{0x} \),

and \( \alpha \) computed with the approximation \( \alpha = \exp[-D(\sigma_e/\sigma_x)^2] \).

We compute easily the energy coefficients

\[
  C_{2x} = \frac{R_{-0.25x} - R_{0.25x}}{x(1 + v_x)(x^{v_x} - x^{-v_x})}
\]

and

\[
  C_{1x} = \frac{R_{0x}}{x^{v_x}} - C_{2x}.
\]

At last, we used the same energy similarity measure \( S_x \) (Eq. (4)) and the same transparency thresholding.

**7. Case of two "images" with opposite speeds**

**7.1. Speed expression**

The case of two “images” with opposite speeds translates by the equality \( v_1 = -v_2 \). As we already mentioned it, this particular case allows the comparison of our model's performance with psychophysical results obtained by human subjects. By a similar process to the case of “one moving image”, we demonstrated [8] that the speed for each triad could be written as

\[
  v_x = \frac{1}{2} \log \left( \frac{\gamma_x + \sqrt{\gamma_x^2 - 4}}{2} \right) \quad \text{for} \quad v > 0,
\]

\[
  v_x = \frac{1}{2} \log \left( \frac{\gamma_x - \sqrt{\gamma_x^2 - 4}}{2} \right) \quad \text{for} \quad v < 0
\]

We choose to compare the results of our method with the psychophysical results recently obtained by Murakami [5]. As he suggested in his paper.
we used a random-dots stimulus developed by Mulligan [4] which is highly interesting for the study of the mechanisms used by the visual system for motion transparency analysis.

7.2.1. The stimulus

In the first stage, two random-dot patterns will be generated on two virtual screens (Fig. 5). Each point has the value 0 ("black") or 1 ("white"), the chance of appearance of the two values being identical. The two virtual screens are in motion in opposite directions with a speed of 1 pixel/image. The subject will observe the superposition of the two patterns through a "fixed frame. There is a simple rule to compute the value of the pixels: when a \( a \) dot ("0" or "1") is superposed with a \( j \) dot ("0" or "1"), the luminance of the resulting pixel is noted \( L_{ij} \).

7.2.2. Results

Our method is a low-level method, which furnishes only results at pixel level, whereas transparency perception is a global measure (the subject decides if the image is transparent or not). We decided to use the same energy similarity measure \( S_x \) (Eq. (4)). The mean value \( S_x \) computed for all the pixels of the image will allow us to have a global value which characterizes the global image transparency. A direct comparison of \( S_x \) with psycho-physical results is not possible, since we have two radically different measures. Still, we can define a global value \( T_x \) which is proportional to motion transparency: \( T_x = 5 - S_x \). The value 5 represents only an empirical offset introduced in order to have \( T_x \) positive values, the appearance of the graphic being exactly the same.

In Fig. 5(d) the surface described by \( T_x \) for different values of \( L_{00} \) and \( L_{01} \) is represented. This result presents a strong similarity to those obtained by Murakami for human subjects. We observed, in the same manner as him, that the "transparency detection" points are essentially located under the main diagonal of the experimentation domain. This supports the idea that the mechanism underlying motion transparency low-level detection and estimation is probably based on the energy properties of motion [6].

8. Conclusions

We described a new energy-based method for estimation of opaque and transparent motions. This technique uses a simplifying hypothesis of additive transparency in order to use a reduced number of Gabor filters organized in triads and it is capable of detecting motion transparency regions and estimating the two motions on these regions.
for two specific cases: the case of transparent moving objects over a fixed background and the case of transparent motion with identical speeds oriented in opposite directions. We explained how this method leads to results which are very similar to those obtained by psychophysical experimentation with human subjects. In order to extend our approach to the general case of motion transparency with two completely independent speeds, it will be interesting, for future research, to use not only groups of three Gabor filters (triads) but also four or five such filters.

References