Solving Motion Planes by Projection and Ring Integration

Gaspard Petit Sébastien Roy
Département d'informatique et de recherche opérationelle
Université de Montréal
{petitgas, roys}@iro.umontreal.ca

Abstract

We present a new method to find motion planes in energy based and spatio-temporal derivative optical flow. Because our method makes few assumptions about the motion model and the number of motions present in the sampling window, we are able to recover simple single motion as well as complex distributions involving transparency and occlusions. We also discuss the effects of spectral overlapping in the case of energy-based methods and present some results on synthetic and natural sequences.

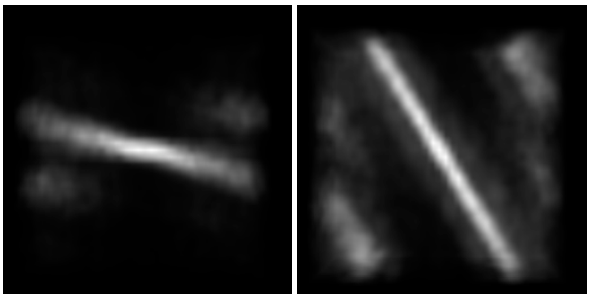

Figure 1: **Warping artifacts in the frequency domain:** when motion is not exactly one pixel per frame (**left:** less than one pixel, **right:** more than one pixel) warping artifacts begin to appear.

1. Introduction

When computing optical flow involving occlusions or transparency, local motion is often ambiguous and needs to be resolved in a larger window or globally as an energy minimization problem. The method presented in this paper was developed to preserve information in the form of distributions of local motions and provides a way to estimate ambiguity. This information can then be used to resolve the system globally. The method resolves planes that pass through the origin in a 3D sampled space. Such planes occur naturally in energy and spatio-temporal derivative methods.

1.1. Energy motion planes

Energy based motion estimation approaches rely on the principle that a linear motion of textures will draw oriented lines in time. These lines, in turn, form planes that intersect at the origin in the sequence's spectrum. Energy based motion estimation is effective for egomotion [5] but can also be used for optical flow where Gabor-like filters are used to locally estimate frequencies [4].

Parametrizing energy motion planes is not trivial. In the frequency domain, low frequencies are close to the origin, giving little information about the orientation of the plane while high frequencies give accurate information but are sensitive to noise. In addition, when motion is high, spectral overlapping occurs and the signal appear to "wrap around" (fig.1).

The quality if the motion distribution recovered depends mainly on the support of the filters used (usually determined by the window size in the spatial domain) and the response to the spatial textures, the resolution in time and wether the assumption that motion is constant over time is true or not. Taking more samples in time provides higher accuracy (especially for high motion) but may break the assumption of constant motion.

1.2. Spatio-temporal derivative planes

Spatio-temporal derivatives plotted in 3D will also lie on the same plane if they represent the same motion. This can easily be seen from the constant brightness constraint which describes a plane where the normal is the motion in the spatial domain:

$$\langle v_x, v_y, v_t \rangle \cdot \nabla I = 0$$

Each sample draws a line in the derivative space. Two samples or more are necessary to solve the plane. Assuming that motion is constant in a small neighborhood, samples may be taken inside a window. They could also be taken at different scales, thus, the spatio-temporal derivative filters would be "tuned" to different frequencies in the image texture.

1.3. Existing work

Traditional implementations make assumptions about the number or type of motions present in the sampling window. Heeger [4] assumes a single motion plane that can be solved analytically using gabor filters. Chen *et al.* [2] makes no assumption as far as the number of motions is concerned, but the method is sensitive to noise and under-sampling artifacts. Mann and Langer [5] support various motion speeds but the orientation has to be the same for all. Pingault [6] makes an *a priori* estimate for the number of motions and uses 3D gaussians and expectation maximization (EM) to model the motion planes. Extra planes are then discarded *a posteriori* using thresholding. Our method is most similar to Yu *et al.* [12] which takes responses of conic filters to map precisely the planes in a spherical (θ, ϕ) space. The main advantage

of our method is that we do not need to find the number of motions by clustering the non-zero values near the θ axis, and counting the clusters and model the spherical signal using EM to recover the orientation of the plane corresponding to each cluster. Instead, we propose integrate energy along rings and generate a motion distribution where a simple voting scheme can then be used to identify the dominant motions. This removes the need for the iterative EM and, because we do not make the assumptions that the motion distribution is gaussian, we obtain a distribution that can be more complex and allows a motion that is not purely translational.

2. Pre-Filtering

In energy based methods, if the filter used does not naturally have a limited support, the image should be filtered to prevent discontinuities on the edges. For example, in a windowed Fourier transform we multiply the signal in our window by a sine: $I'(x) = I(x)\sin(\frac{2\pi x}{size_x}) + 1$. While several authors use a gaussian filter, we find that it tends to blur the spectrum (see fig.2). If this step is neglected, the shap borders will interfere with the motion analysis: they will be considered as non-moving discontinuities with full range spectrum and will add an artificial plane at $t = 0$.

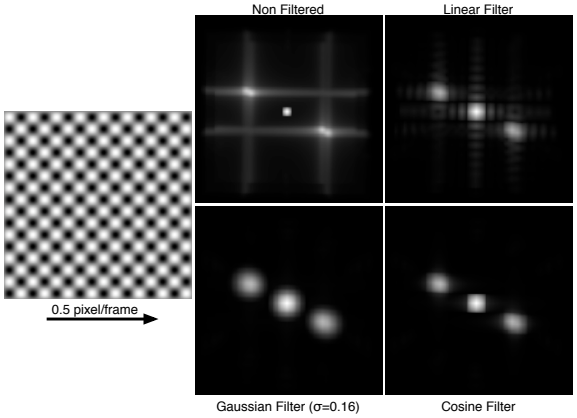

Figure 2: **Effect of various filters on energy:** neglecting spatio-temporal filtering or using the wrong image filter generates artifacts in frequency space.

The next step consists in attenuating the low frequencies using a high-pass filter. This filter will get rid of the DC component as well as frequencies that provide little information about the motion and tend to interfere with the rest of the process. A typical standard deviation for this filter is 20% of the nyquist frequency.

3. Projection

The projection remaps the energy of the sequence onto the surface of a sphere. This step is similar to [12] except that instead of using conic filters, it simply projects by casting rays along the normals of the sphere thus integrating the spectrum F onto the surface $\mathcal{S}$:

$$N_{\theta,\phi} = \langle \sin\theta\cos\phi, \sin\phi, \cos\theta\cos\phi \rangle$$

$$\mathcal{S}(N) = \int F(rN)dr$$

Where F is the 3D Fourier or derivative samples. Only half the sphere needs to be processed, and rays may be cast halfway inside since the energy is even-symmetric. All motion planes pass through the origin, therefore we expect each plane to appear a line on the surface of the sphere (fig.3).

The projection step is not necessary for the derivative approach. The spatio-temporal derivatives already describe a line in 3D that goes through the origin. Therefore, instead of casting rays, we can project the derivative lines on the surface of the sphere directly.

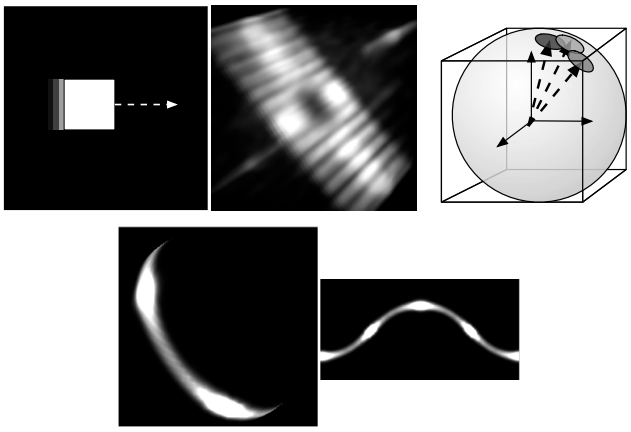

Figure 3: **Projection of the energy on the surface of a sphere:** the **top row** shows the sequence of a square that moves to the right at a speed of $\langle 1, 0 \rangle$ along with the *log*-energy of its 3D Fourier transform. The black hole in the center is a result of the high pass filter. Rays are then cast from the center of the cube and projected onto the surface of a sphere. **Bottom left:** projection on an actual sphere. **Bottom right:** unwrapped (θ, ϕ) texture map.

4. Integration

The motion distribution is found by integrating rings around the sphere. We define an axis $\langle u, v, 1 - \sqrt{u^2 + v^2} \rangle$ on the surface of the sphere and find a point p_0 perpendicular to this axis:

$$p_0 = \langle \begin{array}{l} (\sqrt{u^2+v^2} - 1)\cos(\arctan\frac{v}{u}), \\ (\sqrt{u^2+v^2} - 1)\sin(\arctan\frac{v}{u}), \\ \sqrt{u^2+v^2} \rangle \end{array}$$

We use quaternions to rotate p_0 around the axis integrate $\mathcal{S}$ along the ring:

$$\mathcal{P}(u,v) = \int_0^{2\pi} \mathcal{S}(Rot_\theta \cdot p_0)d\theta$$

Again, in practice, because the signal is even-symmetric, only one hemisphere needs to be computed and rings can be limited to 180^o instead of 360^o.

The result is a gauss map $\mathcal{P}(u, v)$ that gives the response of each motion plane oriented with a normal $\langle u, v, \sqrt{1 - u^2 - v^2} \rangle$. To enhance this map, its minimum value is substracted from all other responses. This minimum value corresponds to white noise in the original signal and low frequencies that contributed to several or all orientations (thus, giving no relevant information about motion).

The gauss map can be represented as a planar map, as shown in fig.4.

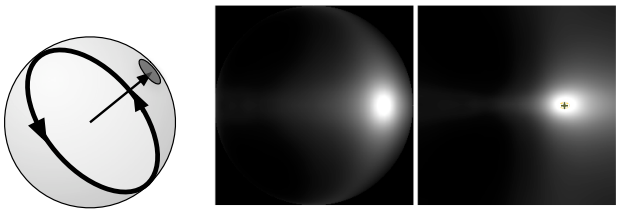

Figure 4: **Integration of rings around the sphere:** we define a normal $\langle x, y, z \rangle$ with coordinates $\langle x, y, \sqrt{1-x^2-y^2} \rangle$ and integrate the energy on the ring that is perpendicular. **Middle:** the gauss map where each point is the responses of a motion plane. **Right:** the equivalent planar map where we find the actual coordinates of the motion: the marked maximum corresponds to the vector $\langle 1.017, 0 \rangle$

After normalization, the motion distribution can be used as a probability distribution. The range of values ($\max(\mathcal{P}) - \min(\mathcal{P})$) also provides an indication of the ambiguity of the motion. Further analysis can be performed from the motion distribution. The window size could be readjusted from the motion ambiguity (unless the spatial signal is a superposition of different signals, a more localized support should result in less ambiguous motion). Energy minimization could be used to resolve motion globally from local patches, or additional parametrization could be applied for specific motion models. For instance, one can analyze the translation stretch caused by parallax of a sequence (see fig. 5 c).

5. Spectral Aliasing in Fast and Slow Motion

With energy based motion analysis, spectral aliasing usually occurs when the motion speed is higher than one pixel per frame. Since the filter is periodic, the motion that is seen at a given frequency of the filter is actually a modulus of that frequency. In general, we assume that the motion is smaller than half the wavelength, but this is not always true. Aliasing can be temporal or spatial, but as pointed out by Mann and Langer [5], in natural sequences, spatial aliasing is less important because of optical blur.

Mann and Langer describes a way to rectify the plane when there is single motion (or a bow-tie). This rectification cannot be used in our method since rectifying for one motion plane might interfere with another valid plane.

Instead, we choose to wrap the rays that we use when projecting the energy on the surface of our sphere. This way, the rays will follow the planes as they wrap and should hold a greater amount of energy when they reach the surface of the sphere. In fig.5, we show how this method allows us to detect motion of 8 pixels using a sampling of $31 \times 31 \times 9$ even if the temporal aliasing is quite important at that speed.

Even without the warping of the rays, experimental results indicate that spectral overlap has little effect on our method. Since the wrapped planes are not aligned with the origin, they become diffused on the surface of the sphere during the projection step.

6. Conclusions

We presented a simple yet robust method to find multiple motion planes for energy-based motion estimation. The method makes no assumption about the number or the type of motion in the sampled window and is robust to the spatial and temporal aliasing. We showed that the maximum response in the motion distribution map corresponds to the plane of the dominant motion, thereby making the method suitable for simple motion estimation. In addition, it is possible to perform further analysis on the local maxima to learn and take into account other motions present in the sampling window. Depending on the resolution of the motion density map, an implementation of this method takes a fraction of a second to compute and one could easily imagine further optimizations where the rings are evaluated from coarse to fine over a region of interest.

References

[1] E. Bruno and D. Pellerin "Robust motion estimation using spatial Gabor-like filters," Signal Processing, vol. 82, n. 2, pp. 297-309, 2002.

[2] W.-G. Chen, G.B. Giannakis and N. Nandhakumar, "A harmonic retrival framework for discontinous motion estimation" IEEE Transactions on Image Processing, vol. 7, n. 9, pp. 1242-1257, 1998.

[3] D.J. Fleet and A.D. Jepson, "Computation of component image velocity from local phase information" International Journal of Computer Vision, vol. 5, n. 1, pp. 77-104, 1990.

[4] D.J. Heeger, "Optical flow using spatiotemporal filters," *International Journal of Computer Vision*, vol. 1, n. 4, pp. 297-302, 1988.

[5] R. Mann and M. Langer, "Estimating camera motion through a 3D cluttered scene," First Canadian Conference on Computer and Robot Vision (CVR 2004). London, Ontario, Canada, 2004

[6] M. Pingault "Estimations fréquentielle et temporelle du mouvement en transparence additive dans les squences d'images," PhD thesis, Laboratoire des Images et des Signaux, Grenoble, France, 2003

[7] M. Shizawa and K. Mase, "Principle of superposition: A common computational framework for analysis of multiple motion" In : IEEE Workshop on Visual Motion, October 1991.

[8] M. Shizawa and K. Mase, "Simultaneous multiple optical flow estimation," In : IEEE Conference on Computer Vision and Pattern Recognition CVPR90, Atlantic City, USA, 1990.

[9] M. Shizawa and K. Mase, "A unified computational theory for motion transparency and motion boundaries based on eigenenergy analysis," In : IEEE Conference on Computer Vision and Pattern Recognition CVPR91, pp. 289-295, Lahaina, Maui, HI, USA, 1991.

[10] I. Stuke, T. Aach, C. Mota and E. Barth "Estimation of multiple motions: regularization and performance evaluation" In : Proceedings of the IS&T SPIE Annual Symposium on Electronic Imaging, Santa Clara, 2003.

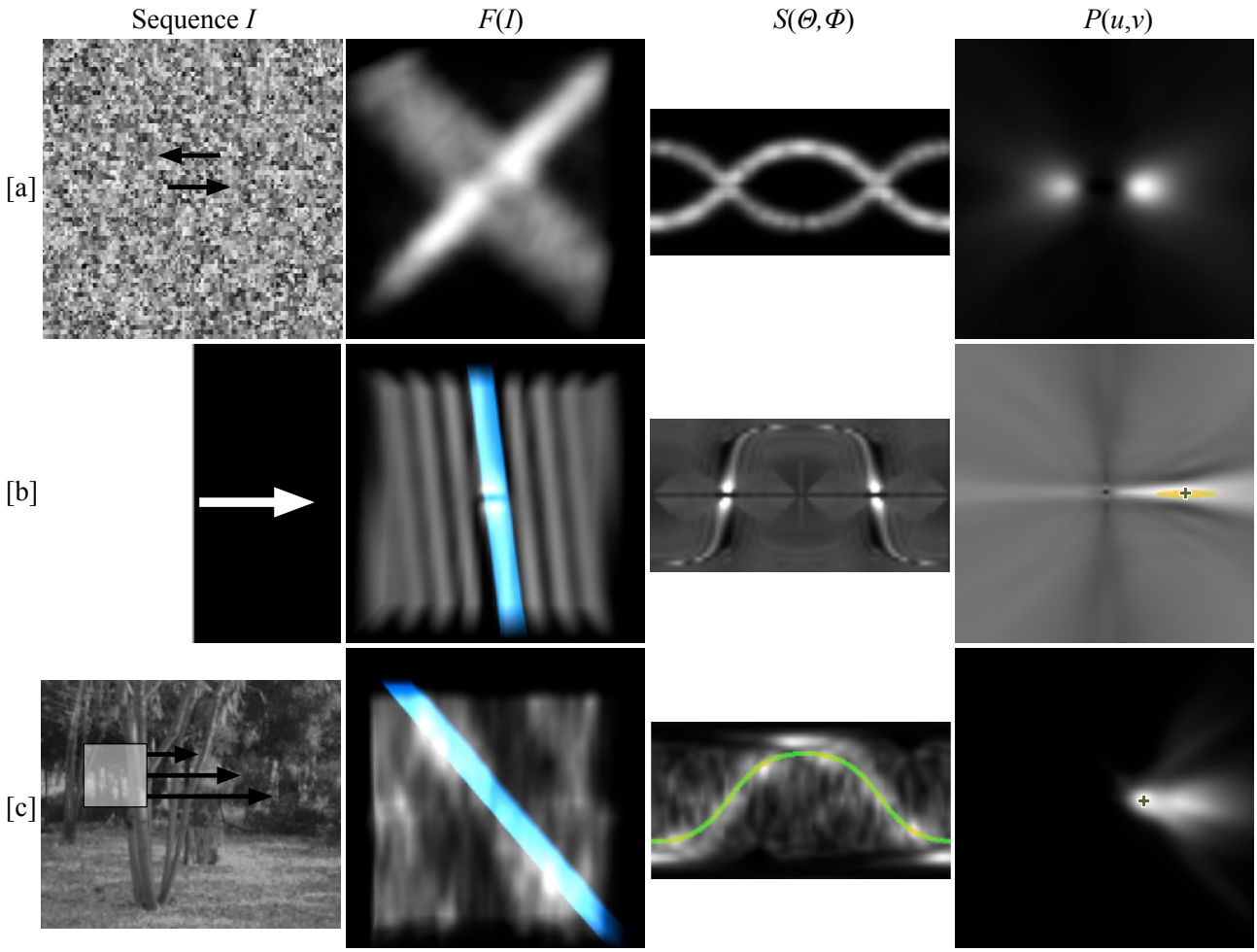

Figure 5: All results used a $31 \times 31 \times 9$ window and an energy approach. The resolution for $\mathcal{S}(\theta, \phi)$ and of the density map $\mathcal{P}(u, v)$ were 64×32 and 64×64 respectively. **a:** Two images of noise with motion $\langle 1, 0\rangle$ and $\langle -1, 0\rangle$ added together . The two motions are recovered (one maximum at $\langle 1.004, 0\rangle$ and another one at $\langle -0.992, 0\rangle$). **b:** A black rectangle on a white background moving at $\langle 8, 0\rangle$ pixels per frame. The high motion creates severe aliasing along the temporal axis. Yet, motion is found at $\langle 7.46, 0\rangle$. For such a sequence, a better approach would be to first subsample the image - reducing the speed and thus the aliasing, but we wanted to show how aliasing affects our method. **c:** Motion with parallax. The window chosen contains multiple motions going to the right (the camera translates to the left and trees of different depth move at different speeds). In the rightmost image, the parallax appears as stretch along the x axis.

[11] W. Yu, G. Sommer and K. Daniilidis, "3D-orientation signatures with conic kernel filtering for multiple motion analysis," In : IEEE Conference on Computer Vision and Pattern Recognition CVPR01, pp. 299-306, Kauai, HI, USA, 2001.

[12] W. Yu, G. Sommer and K. Daniilidis, "Three dimensional orientation signatures with conic kernel filtering for multiple motion analysis," Image and Vision Computing, vol. 21, n. 5, pp. 447-458, 2003.