OPTICAL FLOW CONSTRAINT EQUATION EXTENDED TO TRANSPARENCY

M. Pingault[1], *D. Pellerin*[1,2]
[1]Laboratoire des Images et des Signaux (LIS),
INPG, 46 Av. Félix Vialet 38031 Grenoble Cedex, France
[2]ISTG, Université Joseph Fourier, Grenoble, France
e-mail: pingault,pellerin@lis.inpg.fr

ABSTRACT

Motion transparency phenomena in image sequences are frequent but classical methods of motion estimation are unable to deal with them. This paper describes a method for estimating optical flow by a generalization of the brightness constancy assumption to additive transparencies. This assumption is based on three successive images of a sequence. Thus, by a development to its second order, we obtain an extension of the optical flow constraint equation. The approach assumes that motion is translational on a region large enough in order to regularize the aperture problem. In the way of avoiding outliers, due to a non respect of the brightness constancy assumption, a robust multi-resolution method is used. It is composed of a low-pass pyramid and a M-estimator technique. This method offers some good results on artificial and natural image sequences.

1 INTRODUCTION

In nature, motions in transparencies are very frequent. These phenomena happen when at least two elements are superposed in an image sequence and when these elements are animated by different motions. The most well-known examples of motion transparencies are reflections on transparent surfaces (e.g. windows, water). These phenomena can also be found in partial-transparencies, aka partial occlusions. These kinds of transparencies happen when the object in the foreground is fragmented. The easiest examples are occlusions caused by gates or the branches of a tree.

Several methods exist to estimate the motion of transparent objects with algorithms based on regression techniques [4, 2]. These methods estimate multiple motions in a whole image. However, they use local measures and so can make out only one velocity vector in each pixel of the image sequence. Other methods can be used to deal with transparencies at a local level. It is the case for Shizawa and Mase [7], who set up a model for transparency, starting from the Optical Flow Constraint Equation (OFCE). It turns the motion of objects into uniform translations with the help of spatial and temporal derivatives, which are often used in motion estimation. This method based on the OFCE extend velocity estimations which, until now, were possible only with sequences having a single motion, i.e. in which there was no transparency. Vernon [8] approaches things differently, since he does not model the motions of each object separately as usual. On the contrary, he takes into account the transparency phenomenon inside a system of equations itself.

This article starts with the assumption of motion translation on the condition that there is no illumination change. It enables an equation to be used to solve the problem of additive transparency. This equation is a generalization of OFCE based on three successive images of a sequence when it is made of two transparent objects. In this way the optical flow of the sequence can be estimated. Section 2 explains the modelisation of an image sequence which leads to a generalization of OFCE to additive transparencies. Section 3 shows how M-estimators are used to make a more robust estimation. Section 4 provides some experimental results in natural image sequences.

2 GENERALIZATION OF OFCE TO ADDITIVE TRANSPARENCIES

2.1 The modeling of an image sequence

The basic assumption lies in the necessity to make a model – with a convolution filter – of the motion passing through successive images. It means that we work at the level of a uniform translation motion, which is the common assumption in the field of motion estimation. Before modeling the sequence, we want to put forward another assumption – the absence of intensity variation, which means that the lighting of the setting does not change.

Later on, we will deal with the case of additive transparency, where there are only two objects in motion $u_1(x, y)$ and $u_2(x, y)$ [5]. A velocity filter is allotted to each object: $f(x, y)$ and $g(x, y)$. So, thanks to convolution operator $*^n$, which means that we convolute n times the motion filter with its corresponding object, each image $i_n(x, y)$ of a sequence presenting a phenomenon of additive transparency can be represented by:

$$i_n(x,y) = g(x,y) *^n u_1(x,y) + f(x,y) *^n u_2(x,y) \quad (1)$$

where x and y represent the horizontal and vertical spatial coordinates. Index n represents the sampling of time.

2.2 Brightness constancy assumption

If we apply the spatial Fourier transform to equation (1). If we take, then, three successive images from the sequence and solve the system in order to suppress both transparent objects and keep only the two filters with the successive images, we reach the following equation:

$$\begin{aligned} I_n(f_x,f_y) = & (F^{-1}(f_x,f_y) + G^{-1}(f_x,f_y)) \times I_{n+1}(f_x,f_y) \\ & -F^{-1}(f_x,f_y) \times G^{-1}(f_x,f_y) \times I_{n+2}(f_x,f_y) \quad (2) \end{aligned}$$

This is a non-linear equation related to a second degree equation. It shows that the transparent motion is entirely symmetrical. Indeed, it is possible to invert the filters and the objects without changing the equation.

The last step consists in calculating the inverse Fourier transform while keeping in mind that the motion filters are Dirac delta distribution (for example: $f(x,y) = \delta(x + v_{x_1}, y + v_{y_1})$). By this method we come to the extension of the brightness constancy assumption in the field of additive transparency:

$$\begin{aligned} i_n(x,y) = & i_{n+1}(x + v_{x_1}, y + v_{y_1}) + i_{n+1}(x + v_{x_2}, y + v_{y_2}) \\ & - i_{n+2}(x + (v_{x_1} + v_{x_2}), y + (v_{y_1} + v_{y_2})) \quad (3) \end{aligned}$$

where (v_{x_1}, v_{y_1}) and (v_{x_2}, v_{y_2}) represent the velocity vectors of objects in transparency $u_1(x,y)$ and $u_2(x,y)$.

In the field of usual motion estimation, where there is only one motion in each part of the sequence, the brightness constancy equation – obtained through the translation assumption of moving objects – is the following:

$$i_n(x,y) = i_{n+1}(x + v_x, y + v_y) \quad (4)$$

Thus, the similarity between the two equations enables us to apply classical techniques such as robust estimation and multi-resolution that have been already used in order to solve our new equation.

2.3 Motion constraint equation

We develop the equation (3) with the Taylor expansion on each velocity to its second order:

$$\begin{aligned} i_t(x,y) = & (v_{x_1} + v_{x_2})\ i^x_{n+1}(x,y) + (v_{y_1} + v_{y_2})\ i^y_{n+1}(x,y) \\ & -(v_{x_1} + v_{x_2})\ i^x_{n+2}(x,y) - (v_{y_1} + v_{y_2})\ i^y_{n+2}(x,y) \\ & +\frac{(v^2_{x_1} + v^2_{x_2})}{2}\ i^{x^2}_{n+1}(x,y) + \frac{(v^2_{y_1} + v^2_{y_2})}{2}\ i^{y^2}_{n+1}(x,y) \\ & -\frac{(v_{x_1} + v_{x_2})^2}{2}\ i^{x^2}_{n+2}(x,y) - \frac{(v_{y_1} + v_{y_2})^2}{2}\ i^{y^2}_{n+2}(x,y) \\ & +(v_{x_1}v_{y_1} + v_{x_2}v_{y_2})\ i^{xy}_{n+1}(x,y) \\ & -(v_{x_1} + v_{x_2})(v_{y_1} + v_{y_2})\ i^{xy}_{n+2}(x,y) \quad (5) \end{aligned}$$

with,

$$i_t(x,y) = i_n(x,y) - 2 * i_{n+1}(x,y) + i_{n+2}(x,y),$$
$$i^*_n(x,y) = \frac{\partial i_n(x,y)}{\partial *},\ i^{**}_n(x,y) = \frac{\partial i_n(x,y)^2}{\partial * \partial *}.$$

Equation (5) represents the extension of the Optical Flow Constraint Equation (OFCE) to additive transparency. This equation, relevant to each pixel of the image sequence, has eight unknowns quantities. This is the aperture problem [10]. To resolve this, we need to state that the motion is locally constant inside a window which is large enough, at least eight pixels, around the pixel being studied. We choose a window of 21×21 pixels. Thus, equation (5) can be written in matrix notation in the form of a least-squares problem.

3 ROBUST MULTI-RESOLUTION ESTIMATION

The method we use to reduce noise in image sequences is based on a multi-resolution pyramid with two levels which allows large displacements to be estimated. Between levels, a motion compensation technique is applied. This technique enables us to make a more accurate estimation and also reduce the number of unknown quantities when solving the problem with robust least-squares for reducing the influence of outliers.

3.1 First step

First of all, we estimate velocities with the least-squares method applied to a series of three sub-sampled images [3] (level 1 of the pyramid) taken from the sequence being analyzed. To obtain better results, we use the M-estimator technique. It differs from the least-squares technique by minimizing a function of the estimation error rather than the square number of the error. We chose as function the "Tukey's biweight". Minimization by M-estimators can be shown as a robust least-squares (see [6] for details).

At the end of this step, the velocities ($(\hat{v}_{x_1}, \hat{v}_{y_1})$ and $(\hat{v}_{x_2}, \hat{v}_{y_2})$) of the two objects in additive transparency are estimated. This is obtained by finding the solutions to two second degree equations which are established from the components of vector resulting from robust least-squares. We can bring this estimation down to level 0 of the pyramid (the original images) by oversampling the velocities (bilinear interpolation).

3.2 Second step

We use the first estimation of velocities to solve the OFCE in a robust way through motion compensation (6) of the original images (level 0 of the pyramid).

$$i_n(x,y) = i^1_{n+1}(x+\widetilde{v}_{x_1}, y+\widetilde{v}_{y_1}) + i^2_{n+1}(x+\widetilde{v}_{x_2}, y+\widetilde{v}_{y_2}) - i_{n+2}(x+(\widetilde{v}_{x_1}+\widetilde{v}_{x_2}), y+(\widetilde{v}_{y_1}+\widetilde{v}_{y_2})) \quad (6)$$

with,

$$\widetilde{v}_{k_i} = v_{k_i} - \hat{v}_{k_i}$$

We develop equation (6) with the Taylor expansion, but this time we do it to the first order. We do not need to develop the equation in its second order since the motion compensation suppresses the symmetry existing between the two objects in transparency when the second image splits into two images – i^1_{n+1} and i^2_{n+1} (6).

$$\begin{aligned} i_t(x,y) = & \quad \widetilde{v}_{x_1}(i^{1x}_{n+1}(x,y) - i^x_{n+2}(x,y)) \\ & + \widetilde{v}_{x_2}(i^{2x}_{n+1}(x,y) - i^x_{n+2}(x,y)) \\ & + \widetilde{v}_{y_1}(i^{1y}_{n+1}(x,y) - i^y_{n+2}(x,y)) \\ & + \widetilde{v}_{y_2}(i^{2y}_{n+1}(x,y) - i^y_{n+2}(x,y)) \end{aligned} \quad (7)$$

with,

$$i_t(x,y) = i_n(x,y) - i^1_{n+1}(x,y) - i^2_{n+1}(x,y) + i_{n+2}(x,y)$$

We have a new linear equation with four unknown quantities for each pixel of the image in question. We can solve this problem by applying the spatial constancy of motion to quite a large window (more than four pixels). We have chosen the same size – 21×21 pixels – to reduce noise in the sequence. We also use the technique of M-estimators to solve equation (7).

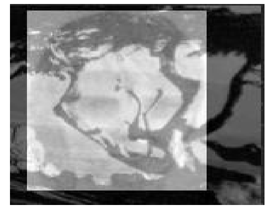
Once we have carried out the second estimation, we obtain the new velocity estimations of the original image $v_{k_i} = \hat{v}_{k_i} + \widetilde{v}_{k_i}$.

4 APPLICATION OF THE METHOD

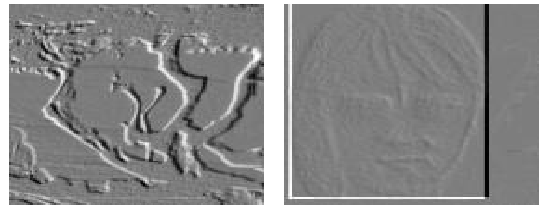
In order to assess the results of the estimation, we have chosen to show the separation of two objects in transparency through the motion compensation method. The final result corresponds to two images representing the difference between the objects, between moment n and moment $n+1$.

This method has been tested on a sequence made of real images whose motion was designed artificially in order to be able to monitor the velocities and, thus, to quantify them. The "Titou" sequence is composed of the "Translation Tree" [1] on which we superposed an image of a face (Fig. 1) in additive transparency. The motion of the "Translation Tree" sequence is a translation motion, collinear to horizontal axis, with an amplitude of 1.73 to 2.26 pixels per image. The motion of the face image was created in order to obtain a velocity equal to one pixel per image collinear to vertical axis.

To define the estimation error, we use the Fleet angular error [1]. Table 1 shows the angular error for all the optical flow.



(a)



(b) (c)

Figure 1: "Titou" sequence : superposition of the image of a face on the "Translation Tree" sequence at the speed of $(1, 0)$ pixels per image. (a) Image taken from the sequence. (b) Separation of "Translation Tree". (c) Separation of the face image.

	Angular error	
Velocities	Mean	Standard deviation
"Translation Tree"	1.67°	1.99°
Face image	7.26°	12.92°

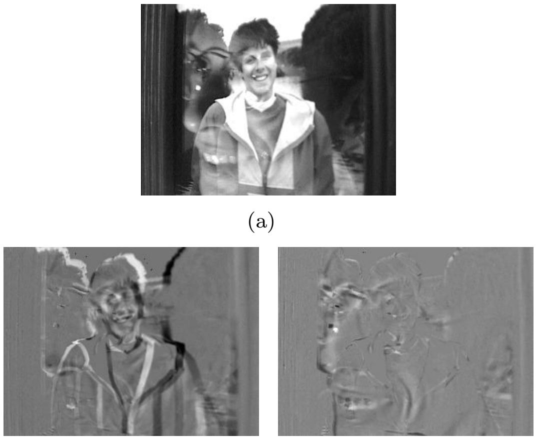
Table 1: Angular error for "Titou" sequence.

We use the knowledge of the image sequences to quantify the separation of objects in transparency through the motion compensation method. For that, we show the difference between the images of the sequences without transparencies, and we compare them to the images obtained with the compensation method. We have shown the results in Table 2. This time, the results are given according to the amplitude, knowing that the images in this sequence are encoded on eight bits, which means on 256 grey levels.

	Compensation error	
Velocities	Mean	Standard deviation
"Translation Tree"	−0.07°	3.61°
Face image	1.60°	4.10°

Table 2: Compensation error for "Titou" sequence.

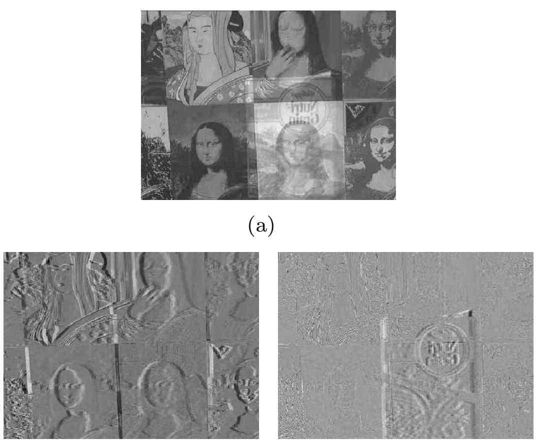
This algorithm has also been tested on two natural image sequences. These sequences are obtained by a reflection on a transparency surface. The first one is named "Transparency sequence" [2] (Fig. 2). This sequence is composed of the reflection of someone's face on the piece of glass covering a photograph. The difficulty of this sequence is that moving objects do not have a real uniform translational motion but are uni-



(a) (b) (c)

Figure 2: "Transparency sequence". (a) Image taken from the sequence. (b) Separation of the photograph. (c) Separation of the reflection.

formly accelerated. The second one is the "Monalisa" sequence [9] (Fig. 3). It is created by moving a movie camera to the right. It shoots a portrait, and on the piece of glass covering this portrait is reflected a pack of muesli moving to the left. The important fact of this sequence is the size of the pack of muesli which does not cover all the portrait. In these two sequences, the results are a good standpoint of separation by the compensation technique of the estimated motion. Only a few artifacts can be seen when the gradient in images is strong, as when there are frontiers.



(a) (b) (c)

Figure 3: "Monalisa" sequence. (a) Image taken from the sequence. (b) Separation of the portrait. (c) Separation of the pack of muesli.

5 CONCLUSION

We have described a velocity estimation method when faced to sequences of images with transparency phenomena. This technique is a new way of approaching the issue. Indeed, its method of modeling includes transparency, not in superposing several models without transparency, but in considering it as a real entire model. We can calculate the velocities by solving a system of equations based on three successive images of a sequence, with the help of derivatives and a least square only. In order to make a more robust estimation and to avoid wrong data, we introduced the use of M-estimators. The results we obtained on artificial and natural sequences prove the efficiency of our new approach.

References

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

[2] M. Black and P. Anandan. The robust estimation of multiple motions : Parametric and piecewise-smooth flow field. *Computer Vision and Image Understanding*, 63(1):75–104, 1996.

[3] P. Burt. The pyramid as a structure for efficient computation. *Mutiresolution Image Processing and Analysis*, Springer-Verlag:6–35, 1984.

[4] M. Irani, B. Rousso, and S. Peleg. Computing occluding and transparent motions. *Internationnal Journal of Computer Vision*, 12(1):5–16, 1994.

[5] J. Mulligan. Motion transparency is restricted to two planes. *Investigation Ophthalmology and Visual Science (supp.)*, 33:1049, 1992.

[6] J. Odobez and P. Bouthemy. Robust multiresolution estimation of parametric motion models. *Journal of Visual Communication and Image Representation*, 6(14):348–365, December 1995.

[7] M. Shizawa and K. Mase. Multiple optical flow-fundamental constraint equations and a unified computational theory for detecting motion transparency and motion boundaries. *Systems and Computers in Japan*, 25(5):52–74, 1994.

[8] D. Vernon. Decoupling fourier components of dynamic image sequences: a theory of signal separation, image segmentation and optical flow estimation. In S. Verlag, editor, *European Conference on Computer Vision ECCV98*, pages 68–85, June 1998.

[9] Y. Weichuan. *Local orientation analysis in images and image sequences using steerable filters.* PhD thesis, University of Kiel, Germany, 2000.

[10] Y. Weiss and E. Adelson. A unified mixture framework for motion segmentation: incorporating spatial coherence and estimating the number of models. In *IEEE Conf. on Computer Vision and Pattern Recognition CVPR-96*, pages 321–326, June 1996.