

MULTIRESOLUTION PARAMETRIC ESTIMATION OF TRANSPARENT MOTIONS

V. Auvray^{1,2}, P. Bouthemy¹ and J. Lienard²

¹IRISA/INRIA, Campus Beaulieu, 35042 Rennes Cedex, France

²General Electric Healthcare, 283 rue de la Miniere, 78530 Buc, France

ABSTRACT

A new framework dealing with motion estimation in transparent images is presented. It relies on a block-oriented estimation with efficient multiresolution function minimization. A downhill simplex method provides an appropriate initialization to this scheme. The estimated velocity vectors are greatly improved by an original postprocessing stage which performs a single motion estimation on differences of warped images. Finally, a regularization step is carried out. It is demonstrated on a large set of simulations that a quarter pixel precision can be attained on noise-free images. The case of noisy images is also addressed and provides satisfactory results, even in the case of low-contrasted medical images. An example on real clinical images is also reported with promising results.

1. INTRODUCTION

Medical images sequences acquired by a X-ray device involve specific motion contents since the image formation process is ruled by the phenomenon of transparency. Unlike in usual video images, there is no apparent occlusion when an object covers another but a grayvalue addition. Any attempt to apply classical motion assumptions and models to such images is useless. We have to explicitly tackle the transparency issue. This paper is concerned with the estimation of transparent motions. We will focus on the two layers case and we will discuss a novel approach involving parametric motion models and multiresolution motion estimation.

The issue of computing transparent motions has already been investigated, but mostly for video sequences and particular transparent effects: images in a mirror, water reflection, moving object observed through a curtain or a fence...

A first class of methods adapts usual motion estimation techniques to the transparent case [1]. A second category of solutions explicitly model transparency using the fundamental equation introduced by Shizawa and Mase [2]. It states that, considering two image layers I_1 and I_2 moving respectively with velocities $u = (u_x, u_y)$ and $v = (v_x, v_y)$ assumed to be constant from time instant $t - 1$ to $t + 1$, we have:

$$\begin{aligned} r(x, y, u, v) = & \\ & I(x + u_x + v_x, y + u_y + v_y, t - 1) + I(x, y, t + 1) \\ & - I(x + v_x, y + v_y, t) - I(x + u_x, y + u_y, t) = 0 \quad (1) \end{aligned}$$

Different methods have been proposed to estimate the transparent motions from this equation using three successive images: minimizing a global energy function with regularizing a global energy function with regularizing [3], adapting wavelets or B-spline decomposition [4], using Markov random fields or block-matching [5]. Another way is to formulate the problem in the frequency domain [6], but the time interval over which motion has to be assumed constant is then much larger.

The paper is organized as follows. Section 2 describes the main aspects of the designed multiresolution estimation stage. Section 3 presents an important post-processing stage and experimental results are reported in Section 4.

2. TRANSPARENT MOTION ESTIMATION

2.1. Transparent motion constraint

We start from equation (1). Actually, it is correct only if it is assumed that I_1 is moving with the velocity u at $p = (x, y)$, and also at location $p + v$ at time $t - 1$. The same holds for I_2 at p and $p + u$. This constraint on the velocities of the two layers practically implies that (1) is rigorously verified only when translational motions are concerned (at least in a given region).

Therefore, to be able to handle more complex motions, we divide the image into small blocks where the motion is assumed to be translational. The expression to be minimized over each block B_i is given by:

$$J(u^i, v^i) = \sum_{(x,y) \in B_i} r(x, y, u^i, v^i)^2 \quad (2)$$

where $r(x, y, u^i, v^i)$ is given by equation (1)

Two plots of the function $J(u, v)$ are displayed in Fig.1 to illustrate its properties. Three images were constructed by considering two layers moving with known velocities in the

x direction. Then, the plot can be restricted to the (u_x, v_x) space. Let us point out that these plots exhibit two symmetrical minima since u and v are playing reverse roles as it can be observed in equation (1). This ambiguity is not a problem yet, since we do not need to associate the velocities u and v with a specific layer.

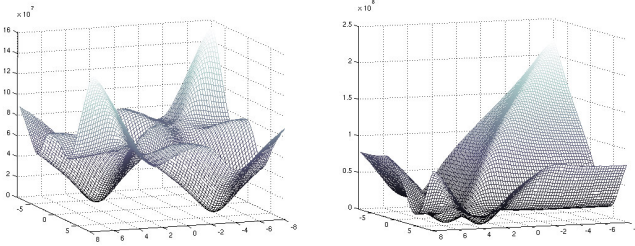


Fig. 1. Two plots of the function $J(u, v)$ restricted to the (u_x, v_x) subspace. True velocities: left: $u = (3, 0)$, $v = (-4, 0)$, right: $u = (3, 0)$ and $v = (5, 0)$

2.2. Multiresolution transparent motion estimation

Given three consecutive images involving two transparent motions, we estimate the velocity of the two layers by minimizing $J(u, v)$. If the velocity magnitudes were small, we could consider a linearized version of J , and then minimize it using an efficient iterative minimization scheme.

Since large motions can be encountered, we introduce a multiresolution framework exploiting Gaussian pyramids of the three consecutive images. At its coarsest level L , motions are small enough to allow a minimisation using the conjugate gradient algorithm of the function $J(u, v)$ linearized around $(0, 0)$, which supplies first estimates of the two motions, denoted (u^L, v^L) . Then, we proceed the same way on the next resolution level $L - 1$ but by linearizing J around $(2.u^L, 2.v^L)$. The method is then iterated through the successive resolution levels until the finest one.

Actually, we have stated that the estimates are more accurate if this scheme (linearization and minimization) was iterated a few times at a given resolution level.

2.3. Initialisation with a simplex algorithm

Such a minimization scheme is efficient and fast, however it is also sensitive to the initialization, especially since we are dealing with medical X-ray images involving low contrast and high noise.

To properly initialize the multiresolution iterative estimation method, we resort to the downhill simplex method. This minimization technique can be applied to nonlinear functions. For a function defined on a space of dimension n ,

it selects $n + 1$ points in that space. In our case, to minimize J in the four-dimensional space (u, v) , it will consider five samples of velocities (u, v) . At each iteration, it substitutes for the point corresponding to the highest value of J a new point sought on a line perpendicular to the hyperplane containing the other n test points [7]. This provides an appropriate initialization for the considered minimization problem. Computational load is limited since we use the simplex algorithm at the coarsest image resolution only.

3. POSTPROCESSING STAGE

3.1. Warping and single motion estimation

It can be stressed from the plots of the function J in Fig.1, that the global minimum might be difficult to attain when the two layers velocities are close (example on the right of Fig.1). (By the way, if the velocities are very similar, then the transparent motion issue vanishes and comes to a single motion situation!) To overcome this problem, we have designed a postprocessing stage as explained below.

The multiresolution estimation stage delivers the estimated velocities $(\hat{u}_{me}, \hat{v}_{me})$. If for instance \hat{u}_{me} is a correct estimation, we can make the corresponding layer (let us say I_1) implicitly disappear without identifying it. To do so, we compensate one layer motion using \hat{u}_{me} so that it appears static, and we compute the differences of the successive warped images:

$$\begin{aligned} \Delta I_{\hat{u}_{me}, t+1}(x, y, t) &= \tilde{I}^{\hat{u}_{me}}(x, y, t) - \tilde{I}^{2 \cdot \hat{u}_{me}}(x, y, t - 1) \\ &= \tilde{I}_2^{\hat{u}_{me}}(x, y, t) - \tilde{I}_2^{2 \cdot \hat{u}_{me}}(x, y, t - 1) \end{aligned} \quad (3)$$

$$\begin{aligned} \Delta I_{\hat{u}_{me}, t}(x, y, t + 1) &= I(x, y, t + 1) - \tilde{I}^{\hat{u}_{me}}(x, y, t) \\ &= I_2(x, y, t + 1) - \tilde{I}_2^{\hat{u}_{me}}(x, y, t) \end{aligned} \quad (4)$$

with $\tilde{I}^w(x, y, t) = I(x + w_x, y + w_y, t)$ the image warped using w .

$\Delta I_{\hat{u}_{me}, t+1}(x, y, t)$ stands for a one-layer sequence moving with $s = \hat{u}_{me} + v$. Its motion can be estimated using a single motion estimation method, and we can reestimate v as $\hat{v}_{pp} = \hat{s} - \hat{u}_{me}$. A similar development can be inferred to get a new estimate of u , denoted \hat{u}_{pp} from $t = u + \hat{v}_{me}$.

Thus, if one of the two velocities is correctly estimated, this scheme offers a way to compute the second one.

To summarize, we build the images \tilde{I} warped with \hat{u}_{me} and \hat{v}_{me} respectively, we compute the successive images ΔI from relations (3) and (4), we estimate the velocities s and t using the robust multiresolution motion estimation method described in [8] (with a constant motion model), and we recover \hat{u}_{pp} and \hat{v}_{pp} .

We have now to select the best estimates out of the four possible pairs $(\hat{u}_{me}, \hat{v}_{me})$, $(\hat{u}_{me}, \hat{v}_{pp})$, $(\hat{u}_{pp}, \hat{v}_{me})$ and $(\hat{u}_{pp}, \hat{v}_{pp})$.

3.2. Choice strategy

An intuitive strategy to select the final estimates of u and v would be to compute the value of $J(u, v)$ for the four pairs and to keep the pair providing the smallest value. This could be relevant for noise-free images, but does not work as soon as noisy sequences are involved.

The expression of $J(u, v)$ is formed by the algebraic sum of four images. If the original images were corrupted by a noise of variance σ^2 , $J(u, v)$ would be affected by a noise of variance $4.\sigma^2$, and make the selection unreliably.

We propose instead to apply a regularization scheme. We want to keep at the end of this stage the two velocities in each block which minimize (2) correctly, while being coherent with the velocities of the neighboring blocks.

To achieve that, we have defined an alternate estimation process. First, we compute the two affine motions for the two moving layers of the image, and then we select in each block the pair of velocities which is the most coherent with the estimated motion models.

The two affine motions models are computed from the velocity vectors selected at the previous iteration, using a robust estimation technique in order to discard the outliers if any. In each block, the two mean displacements corresponding to the two estimated motion models are evaluated and compared (using the L_2 norm) to the four candidate pairs, and the closest pair is selected.

Once again, this iterative algorithm works correctly if it is properly initialized. The experiments we have carried out have led us to initialize the velocity field of the two transparent layers with $(\hat{u}_{me}, \hat{v}_{me})$ in each block. The velocities are first organised in two groups $\{\hat{u}_{me}\}$ and $\{\hat{v}_{me}\}$ supposed to correspond to the two involved layers using a clustering technique. This operation is necessary to make sure that u and v are referring to the same layers from block to block.

The overall framework is summarized in Fig.2.

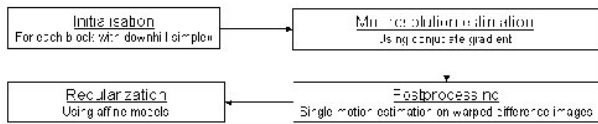


Fig. 2. Overall transparent motion estimation framework

4. EXPERIMENTAL RESULTS

We have generated realistic X-Ray image sequences to assess the performance of our method with an available ground truth. To do so, we have chosen two clinical images (standing for the two layers) as measured on X-Ray imaging systems, and moved them by known translational and affine motions. The displacement values were randomly chosen while ensuring an effective displacement at each pixel between -8 and 8 pixels per frame between two images. The images

Noise	0	10	20
ME	0.33	0.72	1.59
PP	0.3	0.77	1.69
RE	0.25	0.62	1.43
PR	0.20	0.34	0.87

Table 1. Mean estimation errors in pixels for different noise levels and different versions of the estimator (see main text).

intensities were multiplied by a factor accounting for the exposition mode simulated: the higher the X-Ray dose to simulate, the higher the simulated images intensities and the better the image contrasts.

Since X-Ray imaging physically involves multiplicative transparencies, we have multiplied the two layers and passed the composite image through a log operator to be able to work with additive transparencies.

The final 12 bits images were corrupted by an independent gaussian noise whose standard deviation depends on the simulated dose. Our benchmark includes noise-free images as a reference, images representative for high dose exams ($\sigma = 10$, typical diagnostic exams) and for low dose exams ($\sigma = 20$, typical interventional exams). The mean level of the simulated images is about 500.

The experiments reported hereafter involve a first layer undergoing affine motion, the other layer undergoing a translation. The whole estimation framework runs in about 10 seconds for 288×288 images on a Pentium IV 2.4 GHz and 1 Go of memory. Tab.1 contains the mean estimation errors on the velocity vectors over the blocks for 250 random image generations.

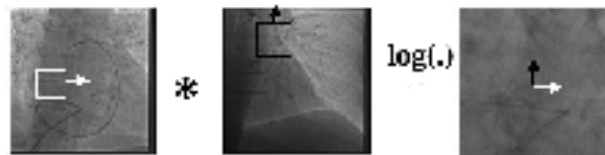


Fig. 3. Generation of transparent images

Four versions of the estimation method are compared to evaluate the contributions of the different modules. *ME* just accomplishes the first two steps: initialisation and multiresolution estimation. *PP* goes also through the postprocessing step but selects the best velocity pair candidate based on the function $J(u, v)$ only, without performing the regularisation step. *RE* adopts the candidate selected at the end of the regularisation step (that is the complete framework proposed in this paper), whereas *PR* just keeps the mean velocities in each block corresponding to the two estimated affine motion models.

It can be observed that the postprocessing step improves the quality of the estimates on noisy images, as soon as the final

estimate selection is made as explained in 3.2. The proposed framework *RE* supplies very good estimations, reaching a subpixelic precision on noisy images typical for fluoroscopic exams.

But the very best strategy here adopts the velocity vectors corresponding to the estimated affine motion models. It should nevertheless be pointed out that these motions models are here exact for the simulated images, which will not be the case for real image sequences with transparent motions. We have also applied our estimation framework to real medical images. We present our results on a cardiac fluoroscopic sequence acquired at 30 Hz. It represents an area about 5cm*5cm on the right of the heart. Two frames of this sequence are presented on Fig.4 along with computed transparent motions.

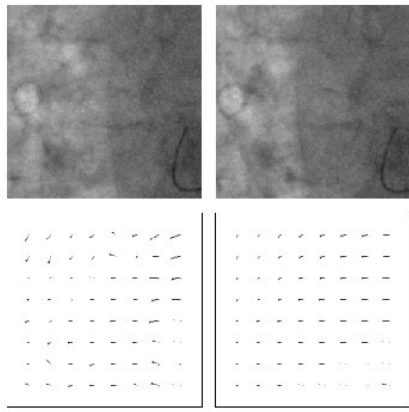


Fig. 4. Top: Images number 1 and 8 of the sequence. Bottom: Regularized estimated velocities using the first three frames. Left: Final estimation in each block using the *RE* algorithm (two vectors are displayed per block), right: the two estimated affine motion models (*PR* algorithm).

The heart (appearing dark here) is beating on the right of the images over a static background corresponding to the spine and ribs. The bright tissues of the lungs are following the heart motion. The magnitude of the motion over a cycle is 25 pixels. The images have a low contrast and are corrupted by an important noise ($\sigma \simeq 20$).

Our estimations are coherent with the motions observed on the sequence. Two motions have been found, which correspond to the anatomic truth: the background, which is static (and therefore does not appear on the arrow fields in Fig.4) and the (affine) motion of the heart. The magnitude of the estimated affine motion is correctly decreasing in the lungs area with the distance to the heart. The images are noisy, low contrasted and contain complex motions. Moreover, the motions are not perfectly constant over the three consecutive images. Even in this complicated situation, the proposed estimation framework gives promising results.

5. CONCLUSION

We have designed a complete and novel multiresolution framework to estimate transparent motions for two layers. In particular, it involves an efficient initialisation and an original postprocessing stage to greatly improve the velocity vectors estimated by the ME step. It was demonstrated on a large set of simulations that a quarter pixel precision can be attained on noise-free images. The processing of noisy images was also addressed, and even in the case of low contrasted medical images, satisfactory results are obtained for a reasonable amount of temporal noise. An example on real clinical images was also reported with encouraging results. More experiments on real X-ray image sequences will be carried out, and we will also tackle situations involving more than two transparent layers.

6. REFERENCES

- [1] M. Irani, B. Rousso, and S. Peleg, "Computing occluding and transparent motions," *International Journal of Computer Vision*, vol. 12, no. 1, pp. 5–16, 1994.
- [2] M. Shizawa and K. Mase, "Simultaneous multiple optical flow," in *Int. Conf. on Pattern Recognition*, Atlantic City, New Jersey, 1990, vol. 1, pp. 274–278.
- [3] J. Toro, F. Owens, and R. Medina, "Multiple motion estimation and segmentation in transparency," in *Proc. of the IEEE Int. Conference on Acoustics, Speech and Signal Processing*, Istanbul, June 2000, pp. 2087–2090.
- [4] M. Pingault, E. Bruno, and D. Pellerin, "A robust multiscale B-spline function decomposition for estimating motion transparency," *IEEE Trans. on Image Processing*, vol. 12, no. 11, pp. 1416–1426, Nov. 2003.
- [5] I. Stuke, T. Aach, C. Mota, and E. Barth, "Estimation of multiple motions using block-matching and Markov random fields," in *SPIE Conf. Visual Communications and Image Processing VCIP'2004*, San Jose, California, Jan 2004, pp. 274–278.
- [6] M. Pingault and D. Pellerin, "Motion estimation of transparent objects in the frequency domain," *Signal Processing*, vol. 84, no. 4, pp. 709–719, April 2004.
- [7] W.H.Press et al, *Numerical recipes in C : The Art of Scientific Computing*, chapter Downhill Simplex Method in Multidimensions, pp. 408–412, Cambridge university press, 1988-1992.
- [8] J.-M. Odobez and P. Bouthemy, "Robust multiresolution estimation of parametric motion models," *Journal of Visual Communication and Image Representation*, vol. 6, no. 4, pp. 348–365, Dec. 1995.