

# Multiresolution Parametric Estimation of Transparent Motions and Denoising of Fluoroscopic Images

Vincent Auvray<sup>1,2</sup>, Jean Liénard<sup>2</sup>, and Patrick Bouthemy<sup>1</sup>

<sup>1</sup> IRISA/INRIA, Campus de Beaulieu, 35042 Rennes Cedex, France

<sup>2</sup> General Electric Healthcare, 283 rue de la Minière, 78530 Buc, France

**Abstract.** We describe a novel multiresolution parametric framework to estimate transparent motions typically present in X-Ray exams. Assuming the presence of two transparent layers, it computes two affine velocity fields by minimizing an appropriate objective function with an incremental Gauss-Newton technique. We have designed a realistic simulation scheme of fluoroscopic image sequences to validate our method on data with ground truth and different levels of noise. An experiment on real clinical images is also reported. We then exploit this transparent-motion estimation method to denoise two layers image sequences using a motion-compensated estimation method. In accordance with theory, we show that we reach a denoising factor of 2/3 in a few iterations without bringing any local artifacts in the image sequence.

## 1 Introduction

X-Ray fluoroscopic image sequences are widely used by cardiologists during interventional exams. Since radiation is kept as low as possible to protect the patient's health, the images are corrupted with a large amount of noise and must be processed. We are concerned with motion-compensated temporal filtering, which requires reliable motion estimation [1]. It is however difficult to directly apply motion estimation methods defined for video sequences to X-Ray images, since the image formation process is ruled by the phenomenon of transparency. Unlike in usual video images, there is no occlusion in the image when an organ covers another but a grayvalue addition. The principle of brightness consistency of points along their trajectories is in particular no longer valid. Motion estimation methods for X-Ray images have to explicitly tackle the transparency issue. If some work have tried to directly extend usual motion estimation strategies to the transparency case [2], most of them model transparency in the spatial domain using the fundamental equation introduced by Shizawa and Mase [3], or its discrete version developed in [4]. The latter states that, if one considers the image sequence  $I$  as the superposition of two layers  $I_1$  and  $I_2$  ( $I = I_1 + I_2$ ), respectively moving with velocities  $v_1 = (v_{1x}, v_{1y})$  and  $v_2 = (v_{2x}, v_{2y})$ , we have:

$$\begin{aligned} r(x, y, v_1, v_2) &= I(x + v_{1x} + v_{2x}, y + v_{1y} + v_{2y}, t - 1) + I(x, y, t + 1) \\ &\quad - I(x + v_{1x}, y + v_{1y}, t) - I(x + v_{2x}, y + v_{2y}, t) = 0 \end{aligned} \quad (1)$$

It implicitly assumes that  $v_1$  and  $v_2$  are constant over time interval  $[t - 1, t + 1]$ . Even if the hypothesis of constant velocity remains problematic at a few specific instants of the heart cycle, Eq.(1) allows reliable estimations most of the time since the temporal velocity variations are usually reasonably smooth. This approach can be extended to  $n$  layers by considering  $n + 1$  images while extending the motion invariance assumption. We will focus on the two-layer case, but it is straightforward to extend our work to  $n$ -transparent layers.

This paper is organized as follows. In Section 2, we describe a parametric estimation method based on an efficient multiresolution minimization. Experimental results are given in Section 3 on realistic synthetic and real X-Ray images. We develop in Section 4 a motion-compensated denoising method for image sequences involving transparency. Finally, Section 5 contains concluding remarks.

## 2 Global Parametric Estimation

### 2.1 Transparent Motion Constraint with Parametric Models

A first category of methods estimate motions in transparency in the frequency domain [5], but these techniques have then to assume that the motion is constant over a large time interval (dozen of frames) and are therefore unapplicable to clinical image sequences involving time-variant movements.

To compute the velocity fields by solving equation (1) in the space domain, we have to minimize

$$J(v_1(\cdot), v_2(\cdot)) = \sum_{(x,y) \in \mathfrak{S}} r(x, y, v_1(x, y), v_2(x, y))^2 \quad (2)$$

where  $r(x, y, v_1(x, y), v_2(x, y))$  is given by Eq.(1) and  $\mathfrak{S}$  denotes the image grid. Several methods have been proposed to solve this problem, making different assumptions on the velocity fields. The more flexible the hypothesis on the motions, the more accurate the estimations, but also the more complex the algorithm. Thus, a compromise has to be reached between measurement accuracy on the one hand and robustness to noise, computational load and sensitivity to the program tuning on the other hand.

In [6], dense velocity fields are computed by adding a regularization term to (2), allowing local motion variations to be correctly estimated at the price of a high sensitivity to noise and complex computations. On the contrary, strong assumptions on the velocities are made in [7] by considering  $v_1$  and  $v_2$  constant on blocks of the image (and therefore accounting for a limited range of motions), to allow fast and robust motion estimation. In [4], the velocities are decomposed on a B-spline basis, so that this method can account for complex motion, while staying relatively tractable. However, the structure of the basis has to be carefully adapted to every particular situation and the computational load becomes high if fine measurement accuracy is needed.

We propose instead to represent the velocity fields with 2D polynomial models, which can account for the considered motions (heart beating, lungs dilation,

diaphragm translation), while requiring a few parameters for each layer. We believe that affine models describe the anatomic movements accurately enough, at least within a given area, while keeping the model simple to handle both the transparency issue and the high level of noise in the images. Our framework would also work with higher-order polynomial models, such as quadratic ones, if needed.

Hence, the velocity vector at point  $(x, y)$  for the layer  $i$  is given by:

$$v_{ix}(x, y) = c_{i,1} + a_{i,1} \cdot x + a_{i,2} \cdot y \quad \text{and} \quad v_{iy}(x, y) = c_{i,2} + a_{i,3} \cdot x + a_{i,4} \cdot y \quad (3)$$

Criterion (2) becomes a function of 12 parameters for the whole image:

$$J(\theta_1, \theta_2) = \sum_{(x,y) \in \mathfrak{S}} r(x, y; \theta_1, \theta_2) \quad \text{with} \quad \theta_i = (c_{i,1}, c_{i,2}, a_{i,1}, a_{i,2}, a_{i,3}, a_{i,4}) \quad (4)$$

## 2.2 Multiresolution Transparent Motion Estimation

If the velocity magnitudes were small, we could consider a linearized version of (4), and then minimize it using an efficient iterative minimization scheme.

Since large motions can occur in practice, we introduce a multiresolution incremental framework exploiting Gaussian pyramids of the three consecutive images. At its coarsest level  $L$ , motions are small enough to allow for a minimisation using the conjugate gradient algorithm applied to the objective function (4) once linearized, which supplies first estimates of the two motions, denoted  $(\hat{\theta}_1^L, \hat{\theta}_2^L)$ . At the level  $L-1$ , we initialize  $(\theta_1^{L-1}, \theta_2^{L-1})$  with  $(\tilde{\theta}_1^{L-1}, \tilde{\theta}_2^{L-1})$  where  $\tilde{c}_{i,k}^{L-1} = 2\hat{c}_{i,k}^L$  ( $k = 1, 2$ ) and  $\tilde{a}_{i,l}^{L-1} = \hat{a}_{i,l}^L$  ( $l = 1, 4$ ). We then write  $\theta_i^{L-1} = \tilde{\theta}_i^{L-1} + \Delta\theta_i^{L-1}$ , and we minimize the objective function  $J(\theta_1, \theta_2)$  with respect to  $\Delta\theta_i^{L-1}$ , once  $J$  is linearized around  $(\tilde{\theta}_1^{L-1}, \tilde{\theta}_2^{L-1})$ . This incremental Gauss-Newton method is then iterated through the successive resolution levels until the finest one.

## 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.

We resort to the downhill simplex method to provide an appropriate initialization at the level  $L$ . This minimization technique can be applied to nonlinear functions. For a function defined on a space of dimension  $n$ , it selects  $n + 1$  samples. In our case, to minimize  $J$  in the 12-dimensional space of the affine motion parameters, it will involve 13 samples of parameters. At each iteration, it substitutes for the sample corresponding to the highest value of  $J$  a new sample found on a line perpendicular to the hyperplane containing the other  $n$  test points [8].

Computational load is limited since we use the simplex algorithm at the coarsest image resolution only.

### 3 Results of Transparent Motion Estimation

#### 3.1 Image Simulation Process

**Image Modeling.** To synthesize images representative for X-Ray exams, we have modeled an imaging system as shown on Fig.1. The anatomy  $A$  attenuates the input dose  $R_{in}$  to give  $R_{out} = R_{in} \cdot A$ . A part of the input radiation is also scattered, which is modeled by adding to  $R_{out}$  a part of a large low-pass version of  $R_{out}$ :  $S = s \cdot R_{out}^{LP}$  ( $s$  ranging from 10 to 50%). A Poisson quantic noise typical for radiation is added. Finally, the X-photons are converted into grayscale levels on the receiver screen, which has a given Modulation Transfer Function  $MTF$ .

**Transparent Image Sequence Generation.** To reproduce configurations typical of cardiac exams, we have formed one layer by taking a real image of an abdominal exam where the spine and ribs are visible, and the second layer with an X-Ray image of the heart (displayed on Fig.2).

To build the attenuation maps of these two images, we have inverted the chain described on Fig.1. We have thus transformed the grayscale image into the X-photonic image, compensated for the scatter, and inferred the attenuation maps from the dose used during the exams. Since we cannot invert the noise operator, we have chosen high radiation exams guaranteeing nearly noise-free images.

The two attenuation images were moved with the simulated velocity field and multiplied to generate the anatomy corresponding to the superposition of the two layers. A fluoroscopic acquisition was simulated following the image formation process given in Fig.1 to get realistic X-Ray images with two transparent layers undergoing known motions. Finally, to work with additive transparency instead of multiplicative one, we applied a log operator to the composite image.

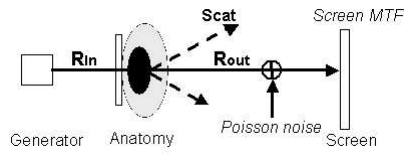


Fig. 1. Image formation model

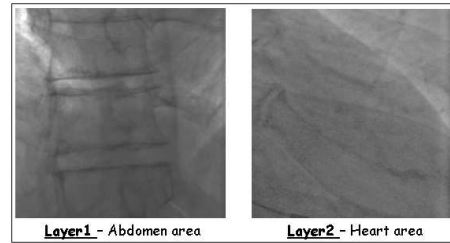


Fig. 2. Real X-Ray images used as layers for the image sequence generation

#### 3.2 Results on Generated Examples

We have generated 250 image sequences of three frames as described above, the abdominal layer undergoing a translation and the heart layer an affine motion. The motion parameter values are randomly chosen while ensuring a displacement at each pixel in the range of  $-8$  to  $8$  pixels. The images are coded on 12 bits, and their mean value is typically 500.

The estimation framework runs in 5 seconds for  $288 * 288$  images on a Pentium

IV (2.4 GHz and 1 Go). Tab.1 contains the mean estimation errors on the velocity vectors for different noise levels (standard deviation of  $\sigma = 10$  or  $\sigma = 20$  grayscale values), scatter rate and with or without MTF simulation.

The estimations are quite satisfactory if we do not take the MTF into account. Even with a large scatter rate, the mean velocities errors are below 1 pixel. However, the motion estimation method seems sensitive to the correlated noise introduced by the MTF. For images corresponding to a moderate radiation level (noise with  $\sigma = 10$ ), the estimations remain reliable. Nevertheless, for a higher noise level which may be encountered in real images ( $\sigma = 20$  and 20% scatter, 50% scatter being a quite extreme situation in practice), the velocities are affected by a mean error of about 2.5 pixels. This would imply some extension of the method to handle MTF effects in highly noisy situations.

**Table 1.** Mean estimation errors in pixels for the proposed framework, with different noise levels and scatter rate, with or without MTF addition (see main text).

	NoMTF		MTF	
Noise	10	20	10	20
No scatter	0.30	0.55	0.49	2.09
20% scatter	0.36	0.76	0.71	2.84
50% scatter	0.38	1.27	1.04	4.54

### 3.3 Results on Clinical Image Sequences

We have also carried out experiments on real medical image sequences. X-Ray medical image sequences usually involves several layers as a whole, but only one or two layers at a time within delimited areas. The segmentation of the image into areas corresponding to two-layers-configurations is beyond the scope of this paper.

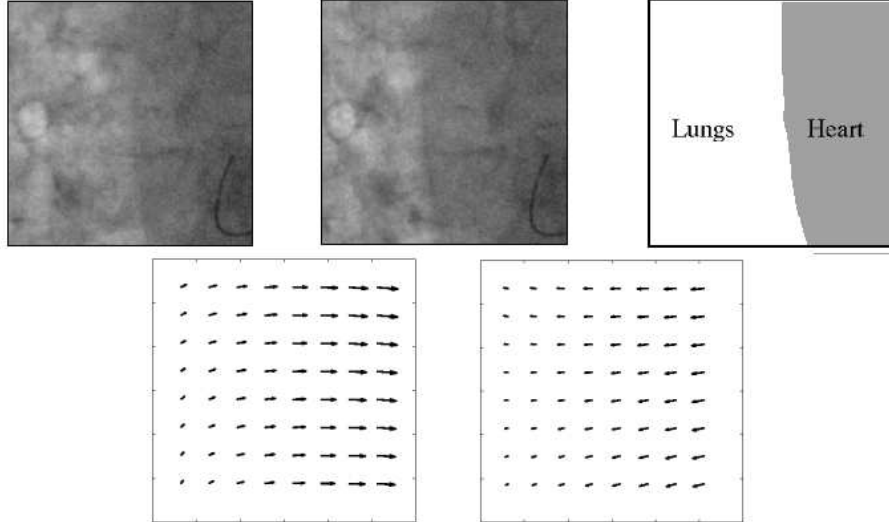
Therefore, we have selected a part of a cardiac fluoroscopic sequence corresponding to a two-layer transparency case. It represents an area of about 5cm x 5cm on the right of the heart, and was acquired at 30 Hz. Let us point out that these fluoroscopic exams are usually carried out without contrast media injection.

Two frames of this sequence are displayed on Fig.3 along with two computed transparent motion fields. 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, so that an observer perceives only two “motions”, one corresponding to the static background and the other to the group “heart+lungs”. The magnitude of the motion over a cycle is 25 pixels. The images have a low contrast and are corrupted by an important noise ( $\sigma \simeq 20$ ).

The estimated affine motion models are coherent with the movements observed on the sequence: the background is static (and therefore the corresponding velocity field is not plotted in Fig.3) and the (affine) motion of the heart matches the anatomic truth. Its magnitude is correctly decreasing in the lungs area with

the distance to the heart.

The images are noisy, low contrasted and contain complex movements. Moreover, the motion is not perfectly constant over three consecutive images, which does not impair the estimation here. Even in this complicated situation, the proposed transparent motion estimation framework supplies convincing results.



**Fig. 3.** Top: From left to right: Images at time instants 1 and 8 of the real fluoroscopic image sequence, anatomical moving regions. Bottom: computed velocities of the mobile layer with our parametric transparent motion estimation method during the diastole (at time 6), and during the systole (at time 13).

## 4 Motion-Compensated Image Sequence Denoising with Transparency

### 4.1 Filtering Approach

Fluoroscopic images are acquired at a very low radiation level so that they need to be denoised to be tractable for the clinician. Their high acquisition rate (typically 30 Hz) is favourable to temporal filtering, whereas spatial filtering is difficult since noise correlation is very perceptible at such a high frame rate.

A direct application of motion-compensated temporal filter for video images (such as those reviewed in [9]) is impossible since it would require to have first separated the layers, which is a complex process. Instead, we propose a motion-compensated temporal filter adapted to transparent images. Rewriting (1) as:

$$\begin{aligned}
 I(x, y, t + 1) = & I(x + v_{1x}, y + v_{1y}, t) + I(x + v_{2x}, y + v_{2y}, t) \\
 & - I(x + v_{1x} + v_{2x}, y + v_{1y} + v_{2y}, t - 1)
 \end{aligned}
 \tag{5}$$

we can note that, if the velocities  $v_1$  and  $v_2$  are available, the image at time  $t + 1$  can be predicted from the two previous images. If the estimated velocities

are accurate enough, the predicted image  $\hat{I}(x, y, t + 1)$  is supposed to match the observed image  $I(x, y, t + 1)$  except for the noise realizations, thus providing the means to smooth out the noise.

## 4.2 Recursive Filtering Method

We adopt a temporally recursive linear filter technique specified as follows:

$$\tilde{I}_{t+1}(x, y) = \alpha(t) \cdot \hat{I}_{t+1}(x, y) + (1 - \alpha(t)) \cdot I_{t+1}(x, y) \quad (6)$$

Let us note  $(\sigma_{\tilde{I},t}^2)_{t \in \mathbb{N}}$  the noise variance of the successive denoised images. From Eq.(5), we can infer that the noise variance of the predicted image equals  $\sigma_{\tilde{I},t-1}^2 + 2\sigma_{\tilde{I},t}^2$ , so that (6) implies:

$$\sigma_{\tilde{I},t+1}^2 = \alpha(t)^2 \sigma_{\tilde{I},t-1}^2 + 2 \cdot \alpha(t)^2 \sigma_{\tilde{I},t}^2 + (1 - \alpha(t))^2 \sigma_I^2 \quad (7)$$

with  $\sigma_I^2$  the variance of the original noise.  $\hat{\alpha}(t) = \sigma_I^2 / (\sigma_{\tilde{I},t-1}^2 + 2 \cdot \sigma_{\tilde{I},t}^2 + \sigma_I^2)$  guarantees the smallest  $\sigma_{\tilde{I},t+1}^2$ . With this setting of  $\alpha(t)$ , the sequence  $(\hat{\alpha}(t))_{t \in \mathbb{N}}$  has two positive fixed points: a repulsive one 0 and an attractive one  $2/3$ .

## 4.3 Denoising Results

We have applied the proposed denoising scheme to 100 image sequences generated as explained in subsection 3.1, with MTF simulation and a scatter rate of 20%. Since we can in practice estimate the noise in the original images from acquisition parameters, we can recursively compute  $\hat{\alpha}(t)$  for an optimal denoising. Our temporal filter used these settings of  $\hat{\alpha}(t)$  (reported in Tab.2 with the theoretical denoising factor  $\hat{\sigma}_{\tilde{I},t}^2 / \sigma_I^2$ ) and the velocities estimated by the transparent motion estimation method of Section 2.

The difference between denoised and original image sequences is quite noticeable if visualized in a live manner at 30 Hz, but cannot be efficiently highlighted from two printed frames. Therefore, we rather present our simulation results in Tab.2. No local artifacts are observed on the denoised images: contrasts are preserved because of the motion compensation and no noise coloration nor structural artifacts have appeared.

## 5 Conclusion

We have designed a novel and efficient multiresolution parametric framework to estimate transparent motions for two layers by using affine motion models. We have generated realistic X-Ray image sequences to assess the performance of our method with an available ground truth. The results on medium-level radiation images are quite satisfactory, and the performance smoothly decreases with higher noise and scatter unless the MTF effect is incorporated. Experiments on

**Table 2.** Noise reduction for a transparent image sequence with two layers. Theoretical values compared with experiments for different radiation configurations (MTF simulation and 20% scatter).

Frame $t$	n	0	1	2	3	4	5	6	7	8
Theoretical	$\hat{\alpha}(t)$	0.250	0.286	0.315	0.324	0.329	0.331	0.332	0.333	0.333
values	$\hat{\sigma}_{I,t}^2/\sigma_I^2$	0.750	0.714	0.685	0.676	0.670	0.669	0.668	0.667	0.667
Simulations ( $\sigma_I = 10$ )	$\sigma_{I,t}^2/\sigma_I^2$	0.653	0.648	0.659	0.661	0.655	0.655	0.657	0.655	0.657
Simulations ( $\sigma_I = 20$ )	$\sigma_{I,t}^2/\sigma_I^2$	0.670	0.667	0.670	0.668	0.670	0.668	0.669	0.672	0.671

real clinical image sequences were also reported with convincing results.

An application of this method for temporal denoising transparent image sequences with two moving layers was also presented. Preliminary experiments have shown that we can reach in a few iterations a denoising factor of 2/3 without adding any local artifacts in the denoised image sequences.

We now plan to extend our method to handle the MTF effect and to carry out more experiments on real X-ray image sequences. We also plan to tackle the "distributed two-layer configuration". The proposed method will then be applicable if we can segment the image in its different regions of two layer configuration.

## References

1. Dekeyser, F., Bouthemy, P., Pérez, P.: Spatio-temporal Wiener filtering of image sequences using a parametric motion model. In: IEEE Int. Conf. on Image Processing, Vancouver (2000)
2. Irani, M., Rousso, B., Peleg, S.: Computing occluding and transparent motions. *International Journal of Computer Vision* **12** (1994) 5–16
3. Shizawa, M., Mase, K.: Principle of superposition: A common computational framework for analysis of multiple motions. In: IEEE Workshop on Visual Motion, Princeton, New-Jersey (1991) 164–172
4. Pingault, M., Bruno, E., Pellerin, D.: A robust multiscale B-spline function decomposition for estimating motion transparency. *IEEE Trans. on Image Processing* **12** (2003) 1416–1426
5. Pingault, M., Pellerin, D.: Motion estimation of transparent objects in the frequency domain. *Signal Processing* **84** (2004) 709–719
6. Stuke, I., Aach, T., Mota, C., Barth, E.: Estimation of multiple motions: regularization and performance evaluation. *Image and Video Communications and Processing 2003*, SPIE **5022** (2003) 75–86
7. Stuke, I., Aach, T., Mota, C., Barth, E.: Estimation of multiple motions by block matching. In: 4th ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD 2003), Luebeck (2003) 358–362
8. Press, W., al: Downhill Simplex Method in Multidimensions. In: Numerical recipes in C : The Art of Scientific Computing. Cambridge university press (1988-1992) 408–412
9. Brailean, J., Kleihorst, R., Efstratiadis, S., Katsaggelos, K., Lagendijk, R.: Noise reduction filters for dynamic image sequences: A review. *Proceedings of the IEEE* **83** (1995) 1272–1292