MOTION ESTIMATION IN X-RAY IMAGE SEQUENCES WITH BI-DISTRIBUTED TRANSPARENCY

V. Auvray[1,2], P. Bouthemy[1] and J. Liénard[2]

[1]IRISA/INRIA, Campus Beaulieu, 35042 Rennes Cedex, France
[2]General Electric Healthcare, 283 rue de la Miniere, 78530 Buc, France

ABSTRACT

This paper is concerned with motion estimation in transparent X-Ray image sequences. Most of these medical images can be divided into areas containing at most two moving transparent layers. We will call it *bi-distributed transparency*. The first contribution of this paper is a motion estimation framework for the two-layer transparency case, able to handle noisy and low-contrasted X-Ray image sequences. It involves three steps: block-matching, affine fit and gradient-based parametric estimation. This estimation scheme is then extended to the bi-distributed transparency case. The second step is now formulated as a joint motion segmentation-estimation problem solved by the iterative minimization of a MRF-based energy function. This framework has been applied to synthetic and real image sequences with quite satisfactory results.

1. INTRODUCTION

This paper is concerned with motion estimation in X-Ray image sequences. Their motion content is specific since the image formation is ruled by the phenomenon of transparency. We explicitly tackle the transparency issue using the fundamental equation introduced in [1]. It states that, considering two image layers 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:

$$r(x, y, u, v) =$$
$$I(x + u_x + v_x, y + u_y + v_y, t - 1) + I(x, y, t + 1)$$
$$-I(x + v_x, y + v_y, t) - I(x + u_x, y + u_y, t) = 0 \quad (1)$$

We propose in this paper an estimation scheme based on (1) adapted to the noisy and low-contrasted X-Ray image sequences. X-Ray medical image sequences can globally involve several layers. In most cases however, they can be divided in areas including at most two moving transparent layers. We refer to this configuration as the *bi-distributed transparency*.

Eq.1 can be extended to a higher number of layers but at the cost of more constraining assumptions (constant motion over a large time interval in particular). Instead, we believe that it is sufficient to address the bi-distributed transparency to be able to estimate motions present in a large majority of X-Ray images, due to anatomical constraints.

This paper is organized as follows. Section 2 describes an estimation scheme for two-layer transparent images, along with simulations on synthetic images. Section 3 is devoted to the motion estimation framework for images with bi-distributed transparency. Results on real clinical sequences are reported in Section 4. Section 5 contains concluding remarks.

2. TWO-LAYER TRANSPARENCY CASE

2.1. Transparent motion estimation

To estimate motions in a two-layer transparent image sequence, we minimize:

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

where $r(x, y, u^i, v^i)$ is given by (1) and B_i are small blocks dividing the image.

Several methods have been proposed to solve (2), leading to the computation of dense velocity fields ([2]) or block-based displacements ([3]). A compromise has to be met between measurement accuracy on one hand and robustness to noise, computational load and sensitivity to the parameter setting on the other hand.

Since the X-Ray images we are dealing with are noisy and low-contrasted, we focus on a robust estimation scheme.

Block-matching techniques that test every possible displacements in a given range supply interesting results. Applied on 32x32 blocks, they proved to be able to handle large displacements and to be robust to noise. Their main drawback is their inability to process untextured blocks.

Gradient-based parametric estimation techniques, as proposed in [4], are also well adapted to X-Ray clinical image sequences. The considered motions (heart beating, lungs dilation, diaphragm translation) can be modeled with 2D

polynomial models, such as the affine motion model. Parametric estimation methods are known to be fast and efficient, while robust to noise and low contrast. They imply an iterative minimization technique that is quite reliable if properly initialized.

We propose to take advantage of these two approaches by combining them. We perform a block-matching technique first, which is applied on a multiresolution representation of the X-Ray images for low computation time. Then, we fit two affine motion models on the sets of translation vectors estimated on every block, previously sorted into two groups using a sliding scheme (i.e. successive of 3×3 blocks), based on a K-NN technique. This step is carried out using a robust estimation method to discard false estimations on homogeneous blocks. We finally consider the gradient-based parametric estimation method on the whole image, initialized with the two affine motion models computed at the previous step.

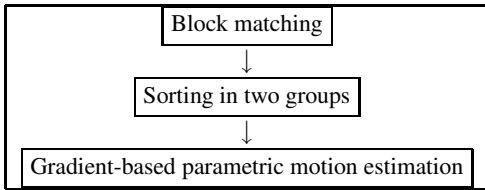The proposed scheme is summarized on Fig.1.

Fig. 1. Two-layer transparent motion estimation scheme

2.2. Results on synthetic images

This two-layer transparent motion estimation scheme was applied to realistic synthetic images. Two high-radiation images acquired with a GE Innova system were used as reference layers. They were moved with known affine motions and multiplied together to generate the composite attenuation map. Realistic image sequences were simulated out of the latter using the model developed in [4]. It accounts for the radiation and imaging chain, and integrates quantic and electronic noises, the MTF (Modulation Transfer Function) of the detector and the scattering effect.

The estimation scheme was applied to 250 such image sequences, generated with random affine motions inducing motions of magnitude smaller than 8 pixels. Images similar to record exams (high-dose diagnostic images, $\sigma = 10$ on 12 bits images, 20% scatter) as well as to fluoroscopic images (low-dose interventional images, $\sigma = 20$, 20% scatter) were processed. Tab.1 contains the mean differences between the estimated velocities and the ground truth.

Results without MTF simulation are excellent. They are still very accurate for record images with MTF, and remain good for the difficult images corresponding to fluoroscopic exams. The required accuracy can depend on the targeted application of the motion estimation.

Noise	10	20
NoMTF	0.32	0.59
MTF	0.47	1.65

Table 1. Mean motion estimation errors in pixels for different noise levels, without and with MTF simulation, for two-layer transparent image sequences (see main text).

3. MOTION ESTIMATION WITH BI-DISTRIBUTED TRANSPARENCY

3.1. Motion estimation framework

We consider here images that can be segmented into areas which contain at most two moving transparent layers (Fig.2). Most of the clinical X-Ray sequences belong to such configuration, since three anatomies undergoing coherent motions rarely superimpose.

We will follow the same scheme as in Section 2 (Fig.1). The block-matching step still works correctly on bi-distributed transparent images, except for the (rare) blocks located over borders between two regions.

However, we will modify the second step of the algorithm: the affine motion fit. Instead, the translational vectors estimated by the block-matching step will be exploited to segment the image into two-layer regions and improve the motion estimations as explained in subsection 3-2. Finally, the gradient-based parametric estimation scheme will be applied on the segmented image.
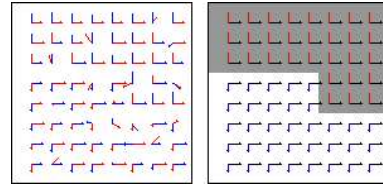

Fig. 2. Example of a bi-distributed transparency situation. On the left, the translation vectors estimated on each block. On the right, the true velocities and different regions that we aim at extracting. (Here, two two-layer regions for a total of three moving layers in the image)

3.2. Segmentation framework

We propose a joint segmentation and estimation scheme based on a Markov Random Field (MRF) modeling. In [5], a relatively similar problem is addressed and a mechanism is proposed to compute multiple transparent motions and their corresponding spatial supports. A dense velocity field is estimated based on (1), and then a EM-algorithm is used to fit parametric models to the velocity estimates and to segment the image. This method proceeds in a sequential way.

In contrast, we propose a *joint* segmentation and motion estimation framework. Such an approach allows us to cope with noisy low-contrasted medical image sequences. Indeed, a regularization framework is required to estimate trans-

parent motions in that context. This implies a segmentation stage associated with an alternate estimation scheme.

We aim at labeling the set of S blocks according to the pair of layers they are belonging to. Let $e = \{e_s\}$ denote the label field with $e_s = \{e_s(1), e_s(2)\}$. Let us assume that the image comprises a total of K transparent layers. To each layer is attached a motion model of parameters θ_k (six parameters). Let $\Theta = \{\theta_k, k = 1, K\}$.

The global energy function is defined by:

$$F(e, \Theta) = \sum_{s \in S} \rho\Big(r_s(\theta_{e_s(1)}, \theta_{e_s(2)})\Big)$$
$$+ \mu \sum_{s,t \in C} \Big(\big(1 - \delta(e_s(1), e_t(1))\big)\big(1 - \delta(e_s(1), e_t(2))\big)$$
$$+ \big(1 - \delta(e_s(2), e_t(1))\big)\big(1 - \delta(e_s(2), e_t(2))\big)\Big) \qquad (3)$$

The first term of Eq.3 makes Eq.1 be verified on each block s with two affine motion fields of parameters $\theta_{e_s(1)}$ and $\theta_{e_s(2)}$ respectively. We use the robust Tukey function $\rho(.)$ to discard outliers. The second term enforces the segmentation to be reasonably smooth, $\delta(.,.)$ being equal to 1 if the two labels are the same and 0 otherwise. The μ parameter weights the relative influence of the two terms.

In other words, a penalty μ is added when introducing a region border involving a change in one layer only, and a penalty 2μ when both layers are different. According to the targeted application, μ can be set either to favor data-driven velocity estimations (small μ), or to favour smooth segmentation (high μ). We have determined μ in a content-adaptive way: $\mu = med_{s \in S}\big(r_s(\theta_{e_s(1)}, \theta_{e_s(2)})\big)$.

Eq.3 is minimized iteratively, using the IRLS method to estimate the parameters of the motion models when the labels are fixed, and the ICM technique to estimate the labels once the motion parameters are fixed. The former stage stops either when the update on the velocities is below a threshold, the latter stops when no labels have been changed. There is also a limit on the iterations. Such an alternate iterative minimization method converges if properly initialized.

To this end, we apply the Hough transform on the displacements computed in the block-matching step. We choose here a three-dimensional parameter space (i.e., a simplified affine motion model, with two translational and one scale components) that roughly matches the anatomic motions. We resort to a continuous increment of the accumulation matrix based on a confidence value depending on the computed 2-layer displaced frame difference. The Hough transform allows us to determine the number of moving layers present in the image, and it provides us with a rough initialization of the affine motion models of the selected layers. This process is fast since it is carried out at the level of blocks, and since we resort to pixel discretization. Then, the label field is initialized by minimizing the first term of Eq.3 only (i.e., we consider a maximum likelihood criterion).

Noise	10	20
NoMTF	0.48	0.96
MTF	0.59	1.85

Table 2. Mean estimation errors in pixels for different noise levels, without or with MTF simulation for bi-distributed transparent configurations (see main text).

3.3. Results on synthetic images

This joint motion estimation-segmentation framework for bi-distributed transparency was applied to the same kind of synthetic images as in Section 2. The first layer covers the whole image, the other one is split in two parts undergoing two different motions. Therefore, we have three moving layers ($K = 3$) and two different two-layer regions. Tab.2 presents the mean estimation errors obtained on 150 generated examples. Results in terms of motion estimation accuracy are quite satisfactory in every configuration. Let us note that we get 33% of oversegmentation ($K_{est} > 3$) and 6% undersegmentation ($K_{est} < 3$).

4. RESULTS ON CLINICAL SEQUENCES

We report the results obtained on two different clinical sequences. The two first columns of Fig.3 show two images of the sequence and the third one contains the motion estimation results. In the fourth column, we plot the extracted layers, and in the fifth one the segmentation in two-layer regions. The framework runs in about 15 seconds with a PC 2.5MHz, 1Go of memory.

The first clinical sequence is a fluoroscopic exam with bi-transparency, corresponding to a $5cm$x$5cm$ area within lungs (bright area on the left of the image) and heart (dark region on the right) between ribs. The observed parts of the lungs and heart are moving in a similar way in this area of the anatomy, so that they should be considered as forming a single layer. Their motion, as well as the static background corresponding to the ribs, is correctly estimated by the algorithm. As a result, the two-layer segmentation map only contains one region over the whole image.

The second clinical sequence is a record sequence with bi-distributed transparency. The diaphragm in the bottom part translates up and down, the bright tissues of the lungs expand and contract, and the gray mass of the heart is beating in the background. A guidewire drives a stent in a coronary. Here again, the motion estimates match the visual observation. The dilation of the heart in the background corresponds to the green layer. The motion of the diaphragm is correctly estimated and corresponds to the red layer. The lungs are static at this time of the cycle and thus are merged with the static background. The obtained segmentation is not perfectly coherent with the anatomy (since a part of the heart is assigned to the red layer of the diaphragm) but it
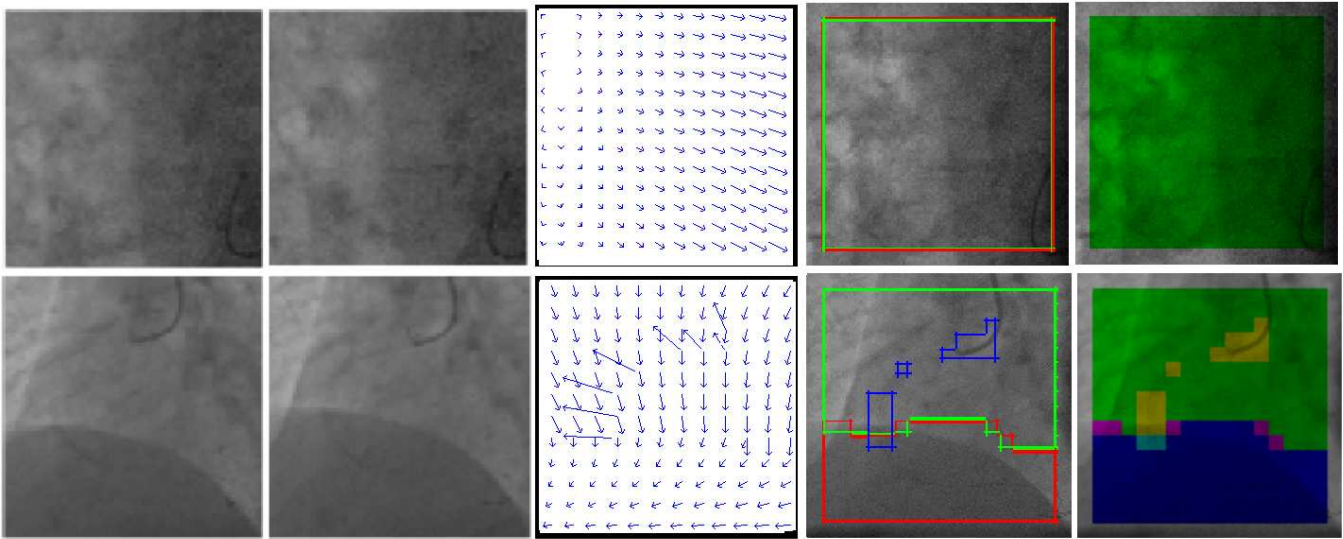
Fig. 3. Motion estimation and segmentation for clinical sequences. From left to right: Two images from the sequence, estimated motions, layer boundaries and region segmentation (see main text). Top row (fluoroscopic sequence of 288x288 images): Frames 10 and 30 of the sequence, estimated motions at time 14 (the arrows length has been multiplied for visualisation). Bottom row (record sequence 576x576): Frames 1 and 15 of the sequence, estimated motions at time 5 (arrows length multiplied by 10).

reflects the observed motions.

Let us point out that the movement of the stent is correctly handled though our modeling is not that adequate for thin regions, and its motion has a high amplitude.

These results are promising in particular regarding motion estimation. The segmentation is less accurate, but for applications such as motion-compensated image denoising this information is not that important *per se*. Motion estimates are wrong only when the assumption of motion constancy between time instant $t-1$ and $t+1$ does not hold (which is the case in 10% to 20% of the sequence length).

5. CONCLUSION

This paper has presented an original motion estimation framework for image sequences with bi-distributed transparency, designed for the noisy and low-contrasted X-Ray image sequences. It involves three steps: block-matching, MRF-based segmentation and gradient-based parametric estimation. The segmentation step results in iteratively minimizing a MRF-based global energy function. It is properly initialized with a Hough transform. This framework has been successfully applied to synthetic and real image sequences.

We plan to extend our framework to identify regions formed by only one layer and those including three or more layers. We also aim at better understanding the impact of the noise correlation due to the MTF on the method performance. Moreover, we want to develop motion-compensated

temporal denoising algorithms for X-Ray image sequences using these motion estimation-segmentation framework.

6. REFERENCES

[1] 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.

[2] I. Stuke, T. Aach, C. Mota, and E. Barth, "Estimation of multiple motions: regularization and performance evaluation," *Image and Video Communications and Processing 2003, SPIE*, vol. 5022, pp. 75–86, May 2003.

[3] I. Stuke, T. Aach, C. Mota, and E. Barth, "Estimation of multiple motions by block matching," in *4th ACIS Int. Conf. on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD 2003)*, Luebeck, October 2003, pp. 358–362.

[4] V. Auvray, J. Liénard, and P. Bouthemy, "Multiresolution parametric estimation of transparent motions and denoising of fluoroscopic images," in *MICCAI'05*, Palm Springs, USA, October 2005, vol. 2, pp. 352–359.

[5] 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.