COMBINED 3D OBJECT MOTION ESTIMATION IN MEDICAL SEQUENCES

Bertrand Delhay[1]*, Patrick Clarysse*[1]*, Stephane Bonnet*[2]*, Pierre Grangeat*[2]*, Isabelle E. Magnin*[1]

[1]CREATIS, UMR CNRS 5515, 69621 Villeurbanne, France
E-Mail: {bertrand.delhay,patrick.clarysse, isabelle.magnin}@creatis.insa-lyon.fr
[2]Laboratoire d'Electronique et de Technologie de l'Information (LETI),
CEA,17 rue des Martyrs, 38054 Grenoble cedex 9, France
E-mail: {Pierre.Grangeat, Stephane.Bonnet@cea.fr

ABSTRACT

A method for the combined estimation of object's motion and boundaries in 3D image sequences is proposed. Based on the object concept of MPEG-4 standard and on a preliminary segmentation, it partitions the images in 3 types of 3D blocks. Motion is first accurately determined on boundary blocks and initialize the estimation in the objects' interior. This initialization is refined with a BBGDS algorithm to detect some disparities in the motion field. Synthetic experiments show that the accuracy is equivalent to classical block matching algorithms with exhaustive search but in less than half of the time. Moreover, the description of objects in terms of motion and shape is readily available.

1. INTRODUCTION

Computed Tomography (CT) is an imaging modality characterized by a high spatial resolution. Following technical evolutions, increased scanner rotation speeds (0.5 s per turn) with a higher number of detectors have been achieved. Thus, fast imaging of moving organs is now possible and will be very helpful in surgeries and radiotherapies, for example. However some artifacts may still appear in images due to organs' motion during acquisition and unappropriate image reconstruction techniques. A method for compensating motion during the reconstruction stage have been proposed, which is based on an accurate description of object's locations and displacements [1-2]. Displacements fields are computed from low spatial resolution images and used to reduce artifacts when reconstructing high resolution image sequences. The quality of the image reconstruction directly depends on the quality of the estimates displacement fields.

In this paper, we propose an efficient and fast 3D segmentation and motion estimation algorithm. The classical exhaustive block matching algorithm like Full Search (FS) can provide an accurate motion estimation but is time-consuming. Several refinements have been proposed to speed up the process ([3],[4]or [5]). But they are mostly based on the assumption that the error surface, which represents the correlation values at each positions in a search window (SW), decreases monotonously, which is sometimes false, especially with blocks located on objects boundaries. Other studies intend to reduce the size of the search window adding a priori knowledge on objects shape ([6], [7]) but such motion estimation techniques show a decreasing efficiency with arbitrarily shaped objects. The MPEG-4 standard [8], provides a coding scheme for arbitrarily shaped video objects (VOs). Each VO is described by all temporal instances (VOPs) of the object in terms of shape, texture and motion. The encoder divides an image into specific blocks. We use this principle to develop our 3D motion estimation technique.

2. METHOD

The main steps of the algorithm which is inspired from a motion estimation method in 2D image sequences based on the MPEG-4 standard [9] is described hereafter. The proposed method, named Combined Object Motion Estimation (COME), processes 3D volumes sequences. We choose a block matching approach because large displacements are to be estimated and differential methods are not well-adapted in such a situation. Moreover, within the context of the targeted application, the motion estimation algorithm must be accurate as well as fast. We will then use different optimization strategies to find block displacements.

The COME algorithm can be divided into three steps: partitioning of the volume into 3D blocks, regularized boundary block motion estimation, motion propagation from outside to the inside of the objects.

2.1. 3D VOPs and Block labelling

A segmentation algorithm is supposed to provide a binary 3D VOP for each object of interest (Figure 1(a)) in the current volume. For each VOP, the volume is divided into regular blocks (Figure 1(b)). There are three types of block labels: Boundary Block (BB), Empty Block (EB), Full Block (FB) defined as

$$\text{BB:} \quad \sum_{\boldsymbol{x}\in B} S(\boldsymbol{x}) < \Delta x \cdot \Delta y \cdot \Delta z \tag{1a}$$

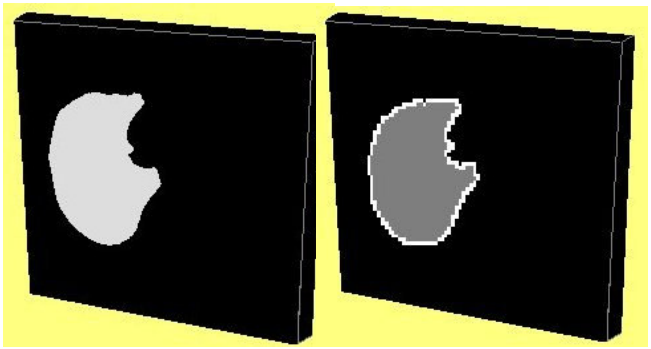
$$\text{EB:} \quad \sum_{\boldsymbol{x}\in B} S(\boldsymbol{x}) = 0 \tag{1b}$$

$$\text{FB:} \quad \sum_{\boldsymbol{x}\in B} S(\boldsymbol{x}) = \Delta x \cdot \Delta y \cdot \Delta z \tag{1c}$$

where $S(\boldsymbol{x})$ stands for the binary value of voxel $\boldsymbol{x}$ within the block B, issued from the segmentation. BB and FB contain voxels belonging to the object of interest; Δx, Δy and Δz are the

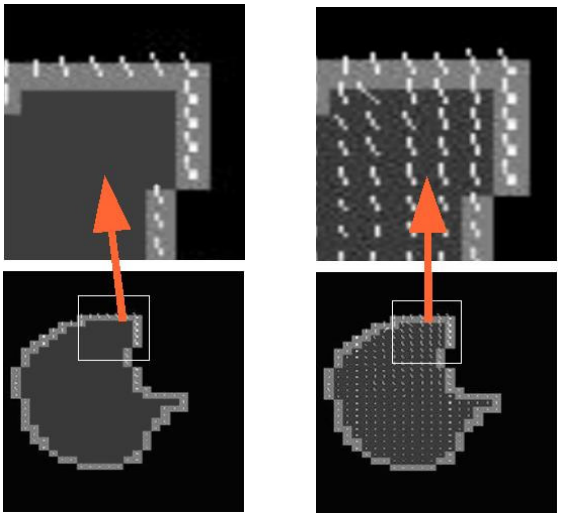
0-7803-8554-3/04/$20.00 ©2004 IEEE.

blocks size. By definition, a BB includes voxels which belong to the considered VOP and its neighboring VOPs. We add two constraints: (i) an EB and a FB cannot be juxtaposed and (ii) a FB lying on the volume boundary is necessarily a BB. Conventional Block Matching algorithms use to scan the blocks in the top-down and left-right order. We propose to first process the BBs, tacking into account that they partially include objects voxels. Displacements vectors for FBs, highly correlated with BBs, are computed in a second step. No motion search is performed for EBs.



(a) (b)

Fig. 1. *(a) A 3D Binary Mask issued from the segmentation. (b) block partitioning and labeling in boundary blocks (white), full blocks (gray) and empty blocks (black)*



(a) (b)

Fig. 2. *Different steps of the 3D motion estimation (2D views) (a) regularized motion is computed on BBs. (b) vector displacements are computed over the whole object with propagation through FBs*

2.2. Motion estimation on Boundary Blocks

2.2.1. Initial estimation

A displacement vector is computed for each BB and only for BB, to estimate the motion of the object's contours between the current volume (time t) and the following volume (time $t+1$). In order to ensure the accuracy of the motion vectors of the BBs, we use a speed-up Block Matching exhaustive search which computes a Weighted Mean Absolute Difference measure (WMAD) (2) to find the best location in the SW:

$$WMAD(\boldsymbol{d}) = \frac{1}{\sum_{\boldsymbol{x}\in B_i} S(\boldsymbol{x})} \cdot \sum_{\boldsymbol{x}\in B_i} \left\{ \left| I^{t+1}(\boldsymbol{x}) - I^{t}(\boldsymbol{x}+\boldsymbol{d}) \right| \cdot S(\boldsymbol{x}) \right\} \tag{2}$$

where $\boldsymbol{d}$ is the expected displacement and belongs to SW. $I^t(x)$ is the intensity at time t of voxel $\boldsymbol{x}$ within the block B_i. The main interest of this step is that the number of blocks for which exhaustive search is performed is largely reduced.

For all the BBs, only the voxels which belong to the object are taken into account for the correspondance search using $WMAD$. Thus, we exclude neighboring object's motions. To accelerate the process, the accumulative property of the $WMAD$ measure is used to stop the search when a certain threshold set by the initial value is reached.

2.2.2. Regularization

As the initial estimation is local, a smoothing constraint is added to ensure the region coherency of the Boundary Motion. The motion vectors are regularized by minimizing the following energy term:

$$E_{tot}(\boldsymbol{d}) = E_e(\boldsymbol{d}) + \alpha E_i(\boldsymbol{d}) \tag{3}$$

$$E_e(\boldsymbol{d}) = WMAD(\boldsymbol{d}) \tag{4}$$

$$E_i(\boldsymbol{d}) = \sum_{B_n \in N_k} \left| \overline{\boldsymbol{d}} - \overline{\boldsymbol{d}}_{\boldsymbol{n}} \right|^2 \tag{5}$$

The energy component E_e (4) represents the dissimilarity between the current block and the displaced reference block while the component E_i (5) penalizes the disparity between neighboring vectors. Parameter α balances the contribution of the two energy components. We use the euclidean norm to evaluate the distance between the estimated displacement $\boldsymbol{d}$ of the block B_k and the estimated displacement $\boldsymbol{d_n}$ of neighboring blocks B_n which belongs to N_k, the 3D neighborhood of B_k. The minimization of (3) is performed through the Iterated Conditional Mode (ICM) algorithm [10]. The motion field on the objects' boundaries produced by the regularization process is thus smoothed. Figure(2(a)) shows the motion vectors obtained at this step.

2.3. Motion estimation within the object (FBs)

2.3.1. Initial estimation of FBs

The previous step provides a smooth and accurate motion estimation for all the BBs. The motion for FBs is provided by the BBs, at the periphery, and spread out to FBs inside the object.

2.3.2. Refinement of FBs

It is obvious that the deformation of human organs is not uniform, especially within the lungs. The refinement step progressively

modifies those initial values using a block-based gradient descent search (BBGDS) algorithm [5]. The dissimilarity values that result from checking positions in the SW describe an error surface. With the initialized displacement, we might be closer to the true motion vector and the BBGDS has a higher chance to find an optimum. In this way, we are able to detect some disparity in the displacement field. Thus, an efficient motion estimation for the FBs is achieved avoiding an exhaustive search. At this step, a motion vector for all blocks of the objects of interest is available. From that, it is straightforward to compute a dense motion field (at each voxel of the volume) using 3D linear interpolation.

3. RESULTS

3.1. Data and evaluation criterion

Our method is evaluated on two synthetic volume pairs from a X-Ray 3D chest acquisition provided by CEA-LETI in Grenoble , France. Volumes dimensions are 512x512x12 and each voxel is coded in Hounsfield units. In the first pair, a 3D global uniform FFD deformation [11] has simulated the respiratory motion starting from a real acquisition. In the second set, discontinuities between the objects have been introduced by simulating diaphragm and lung interactions. As the respiratory motion is simulated, the theoretic displacement field is completely known. Error measures evaluate the shift between theoretical and estimated displacement vectors in magnitude (6) and angle (7). The DFD (8) and computing time are also calculated to assess the results.

$$QME(\boldsymbol{x}) = \left|\boldsymbol{v} - \boldsymbol{v}'\right| \tag{6}$$

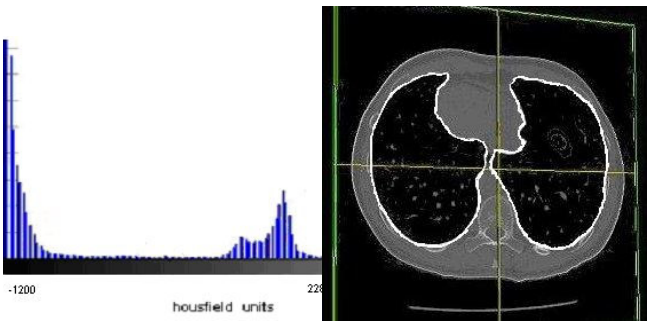
$$AME(\boldsymbol{x}) = arccos\left(\left\langle \frac{\boldsymbol{v}}{|\boldsymbol{v}|} \cdot \frac{\boldsymbol{v}'}{|\boldsymbol{v}'|} \right\rangle\right) \tag{7}$$

$$DFD(\boldsymbol{x}+t) = I(\mathbf{x}, t+1) - I(\mathbf{x}+\mathbf{v'}, t) \tag{8}$$

where $\boldsymbol{v}$ represents the theoretical displacement, $\boldsymbol{v'}$ the estimated displacement and $\langle\cdot\rangle$, the scalar product.

3.2. Tests

The first step concerns the extraction of anatomic structures and the VOPs construction. In thorax images, the intensity distribution exhibits two well distinct peaks as shown in Figure (3(a)). The segmentation of the lungs from thoses images relies on the optimal thresholding method presented in [12], where a threshold is automatically determined. Morphological erosions and dilations are applied to obtain the final voxel classification (Figure 3b). This step provides a good description of the VOPs used as input to our algorithm. SW dimensions are set to 15x15x5 according to respiration magnitude. Block dimensions are 7x7x3. The number of iterations for the regularization process is set to 5 and the coefficient α is set to 1. Computing time and accuracy of COME are compared with the 3D Block Matching with exhaustive search (FS). Results are presented in Table 1. COME algorithm achieves thes same accuracy as the reference FS method with a reduced processing time (1/2). Figure (4) displays different VOP motion fields. In the second sequence, a more realistic respiratory motion has been simulated. In particular, because of the pleural liquid, organs move differently at their interface (shear motion) and the diaphragm pushes the lung in the vertical direction. The benefit of the local motion refinement is clearly seen in figures (5). It shows the dense motion field on the first slice. Vertical displacement is detected whereas in the interior of the left lung no vertical motion has been imposed for BBs of the object. Figure (6) emphasis the ability of the algorithm to estimate different motion on opposite sides of object's interface. Moreover, with the VOP concept, it provides the objects together with their motion. With a 3 GHz Pentium, processing time is almost decreased by three including segmentation and interpolation time. The performance will be even better with a bigger volume, because the FBs/BBs ratio will increase.



(a) (b)

Fig. 3. *a) The two characteristic peaks on a chest tomographic histogram b) Resulting segmentation for the lung*

4. CONCLUSION

We presented a 3D fast combined motion estimation and object extraction method. Promising preliminary results were obtained on synthetic volumes. It might be very helpful and faster to compute a motion field only for a region of interest (VOP) with applications such as radiotherapy or biopsy where the whole motion field of the thorax is not needed. The test on realistic respiration simulations shows that motion discontinuities on objects boundaries are better handled with our method than with classical algorithms.

In the short term, the next step of our study is to evaluate COME on in vivo acquisitions. We plan to use multi-volume frames to investigate temporal prediction. By tacking into account previous estimations and redundancies present in real situation we envisage to track motion.

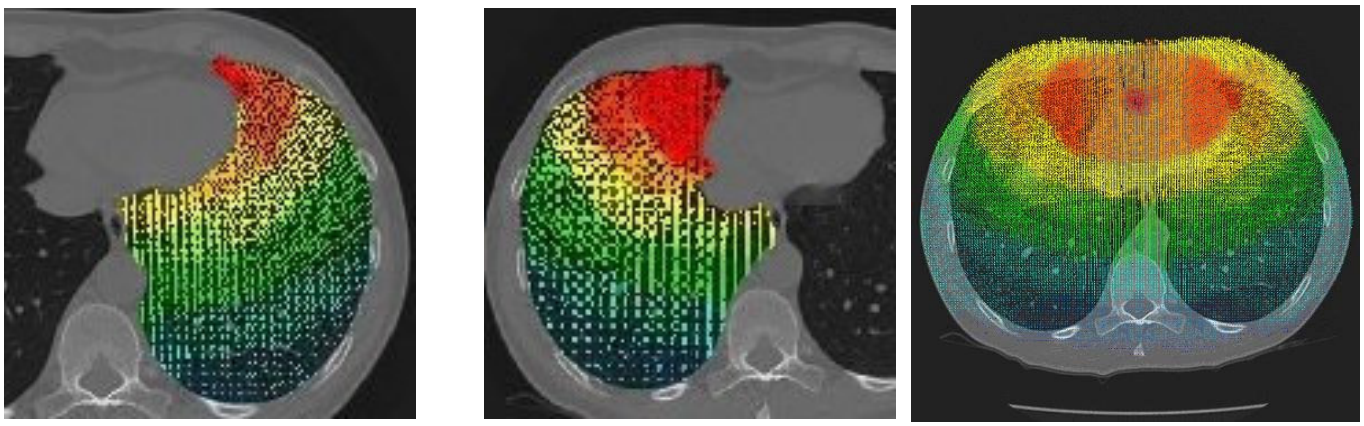
Finally, the time processing is still too long for imposed temporal resolution and we plan to investigate temporal and spatial multiresolution processes.

5. ACKNOWLEDGEMENT

We thank E. Revounkova and O. Pantea for their participation during this work.

6. REFERENCES

[1] S.Bonnet, A.Koenig, S.Roux, P.Hugonnard,R.Guillemaud, P.Grangeat "Dynamic X-Ray Computed Tomography", *Proceedings of the IEEE*, vol. 91, No 10, pp. 1574-1587, October 2003

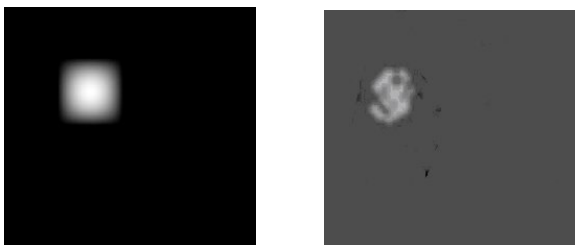


(a) (b) (c)

Fig. 4. *a,b) dense motion field for different VOPs c) the whole dense field*

Table 1. *Error measurements for the two compared methods*

Accuracy		COME	FS
DFD (HU)	Mean	85.2	83.9
	Std. Dev.	72.2	72.1
AME (Deg)	Mean	12	12
	Std. Dev.	11.5	10.8
QME (Voxel)	Mean	0.52	0.51
	Std. Dev.	0.45	0.35
Time		1mn 07s	2mn 35s



(a) (b)

Fig. 5. *Comparison between theoretical dense motion field on the first slice in the vertical direction (a) and estimated motion field (b)*

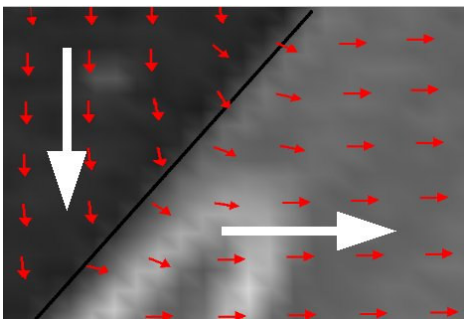


Fig. 6. *Estimated discontinuities on object's boundary*

[2] S.Roux, L.Desbat, A.Koenig and P.Grangeat, "Efficient acquisition for periodic dynamic CT", *Nuclear Science Symposium conference records*, vol. 3, pp.1612-1616, 2002.

[3] T. Koga, K. Iinuma, A. Hirano, Y. Iijima, and T. Ishiguro, "Motion compensated interframe coding for video conferencing", *Proc. Nat.Telecommunications Conf*, pp. 531-535, 1981.

[4] A.Beghdadi, M.Mesbah, J.MonteilA, "A fast incremental approach for accurate measurement of the displacement field", *Image and Vision Computing*, vol. 21, pp. 383-399, 2003

[5] L. K. Liu and E. Feig, "A block-based gradient descent search algorithm for block motion estimation in video coding", *IEEE Trans. Circuits Syst.Video Technol*, vol. 6, pp. 419-422, 1996.

[6] Y-S. Chen, Y-P. Hung, C-S. Fuh, "A fast Block Matching Algorithm Based on the Winner Update Energy", *Proceedings of the fourth Asian Conference on Computer Vision*, Taiwan, vol. 2, pp. 977-982, 2000.

[7] K-L. Chung, L-C. Chang , "A New Predictive Search Area Approach for Fast Block Motion Estimation", *IEEE Transactions On Image Processing*, vol. 12, issue 6, pp. 648-652, 2003

[8] T.Ebrahimi, C.Horne, "MPEG-4 natural video coding- An overview", *Signal processing: Image communication*, vol. 15, pp. 365-385, 2000.

[9] K.C. Hui, W-C Siu, Y-L Chan, "Fast Motion Estimation of arbitrarily shaped video objects in MPEG-4", *Signal processing: Image communication*, vol. 18, pp. 33-50, 2003.

[10] J. Besag, "On the statistical analysis of dirty pictures ", *J.R.Statistical society*, vol. 48, pp. 259-302, 1986.

[11] T.W. Sedeberg, S.R. Parry, "Free Form Deformation of solid geometric models", *SIGGRAPH*, vol. 30, pp. 151-160, 1986;

[12] S.Hu, E.hoffman, J.M.Reinhardt, "Automatic Lung Segmentation for accurate quantitation of Volumetric X-RAY CT images ", *IEEE transactions on medicals imaging*, vol. 20, issue 6, pp 490-498, 2001.