Region-Based Segmentation of Generic Video Scenes Indexing

Aree A. Mohammed

College of Science, University of Sulaimani, Iraq (e-mail: areeali@hotmail.com)

Abstract— In this work we develop an object extraction method and propose efficient algorithms for object motion characterization. The set of proposed tools serves as a basis for development of object-based functionalities for manipulation of video content. The estimators by different algorithms are compared in terms of quality and performance and tested on real video sequences. The proposed method will be useful for the latest standards of encoding and description of multimedia content – MPEG4 and MPEG7.

Keywords— *Object extraction, Video indexing, Segmentation, Optical flow, Motion estimators.*

I. INTRODUCTION

Extraction of objects of interest from generic video scenes remains a topic of high priority in the framework of object-based coding and indexing of multimedia content. Such tools are mandatory for the latest standards of encoding and description of multimedia content – MPEG4 and MPEG7. Several approaches for object-based segmentation of video have been proposed in literature [1,2,3], and can be qualified as "local" and "global" ones.

The local approaches [3], seek for determining the occluding border between the object of interest and background and are based on differential techniques. The global ones [1,2], proceed by region-based segmentation in the area of interest and allow an object to be extracted by labelling regions belonging to it. It should be stated that in case of generic video scenes when the object and the background are animated by very similar motions, the extraction methods with user interaction are practically the only possible way to reach a viable solution [2].

In this paper we are interested in a fine object-based segmentation with a high quality of object boundary and fine local motion representation by object-based optical flow. The method we propose consists in extraction of object shape by interactive colour segmentation with topological constraints and then estimation of optical flow inside the object area. The paper is organized as follows. In Sect-2 the interactive spatial segmentation tool is presented. The method of estimation of object-based optical flow is described in Sec-3. Results and conclusions are outlined in Sec-4 and Sec-5.

II. INTERACTIVE SPATIAL SEGMENTATION

The goal of interactive spatial segmentation of a video scene is to extract a precise object mask from image plane limiting user interaction. To do this, an unsupervised region-based segmentation is fulfilled first. Then user-selected regions with fine boundary are regrouped into an object mask (see figure 1-c). Our segmentation predicate is a low level predicate based on a measure of the homogeneity of regions and only depending on the geometrical, topological and statistical properties of the regions. The construction of the region partition is based on the most general region based segmentation approach which is the split and merges approach.

The region homogeneity predicate is based on the minimization of the mean squared error of merged region, which is computed from the statistical moments. For a region the predicate fails when it is higher than a threshold function and merging stops when all the regions are homogeneous enough. The shape of the regions is presented by an inter-pixel boundary and the topology of the region partition by a set of combinatorial maps. This model allows an efficient extraction of geometrical and topological features needed by split and merge algorithms [4]. The object of interest can be easily obtained by interactively selecting the desired region or set of regions (see figure 1-b).
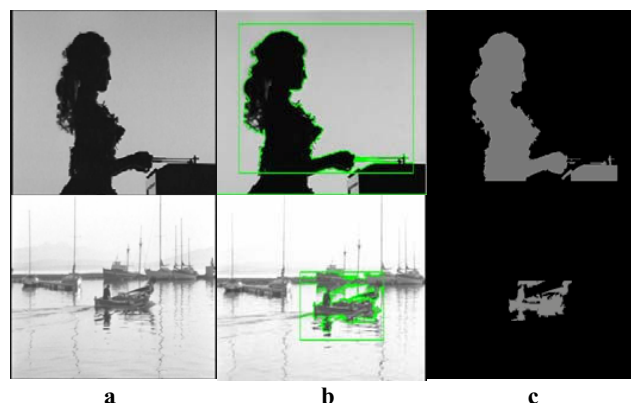

a   b   c
Fig1. Results of interactive segmentation: a) original frame, b) region-based segmentation, c) object mask.

III.   OPTICAL FLOW ESTIMATION FOR OBJECT-BASED SEGMENTATION

In this work, we are interested in low-level motion descriptors, such as the dense optical flow and develop a fine method of its estimation. The optical flow is a good initialisation for a global motion estimation, e.g. in the case of linear affine models [2].

Pel-recursive methods presented in fundamental works [5,6], allow computation of optical flow by iterative/recursive techniques. The error functional is usually represented by a squared DFD.

$$DFD(p,t) = I(\vec{p},t) - I(\vec{p} + \vec{d}, t-1) \rightarrow (1)$$

Here $\vec{p}$ is a coordinate vector of pixels in the considered area and $\vec{d}$ is the elementary displacement vector. The Steepest gradient descent method proved to be efficient in case of model-based motion estimation [2,8,9]. The basic scheme of gradient descent can be expressed as

$$\vec{d}^{\,i} = \vec{d}^{\,i-1} - \frac{\varepsilon}{2} \nabla((DFD)^2(\vec{d}^{\,i-1})) \rightarrow (2)$$

Here ε is a constant gain matrix. The adaptive gain [7], allows grey level gradient to be taken in to account in order to reduce the descent step in the neighborhood of contours. Nevertheless, the scheme with an adaptive gain remains slow and particularly sensitive to local minima of the cost function. These is why we considered two schemes of descent methods, accelerated and conjugate gradient and adopted them for optical flow estimation [5].

The accelerated gradient scheme consists in computation of gradient of the functional at the initial point and then keeping the same gradient while the error functional decreases. In the case of non-linearity of error functional, the conjugate gradient optimisation method can improve the estimation quality. At each iteration, the conjugate gradient method seeks to move in a new direction orthogonal to all preceding ones [8]. Then the iterative equation is:

$$\vec{d}^{\,i} = \vec{d}^{\,i-1} - \alpha \, (\overrightarrow{dir})^{i-1} \rightarrow (3)$$

The orthogonal condition is satisfied when $\left(dir^{i-1}\right)^T .H.dir^i = 0$, (H is the Hessian matrix). However, the cost function is not convex (i.e. H=0) in several positions in image and at several iterations. For this reason, we proposed an adaptive scheme where α is constant for all situations when H=0 and it can be computed when $H \neq 0$. This scheme requires a re-initialization of the direction by the gradient computation rather often (every 2-3 iterations). Nevertheless, it remains sufficiently interesting as it ensures

the descent along the optimal direction when possible. In order to reduce the risk of local minima we also use a classical multi-resolution approach with Gaussian filtering and sub sampling.

IV.   TEST RESULTS

The whole method: interactive color based segmentation and motion estimation were tested on 100 images of video sequences "Aquaculture", "La Joueuse de Tympanon" and "L'homme de Tautavel"® SFRS in format CCIR601.

The estimators were compared in terms of quality (PSNR and MSE of motion compensation) and performance (time) as shown in Figure (2). According to 2a, the conjugate gradient is the best (in terms of MSE/PSNR) compared with the steepest descent method. However it takes much time at each iteration because of the Hessian calculation at each descent step. The Figure 2c shows the efficiency of the accelerated method with almost ~50 % -reduction. By using the multi-resolution approach we obtain a strong reduction of functional error compared to the mono-resolution case (up to 50 % of reduction of MSE). In our actual work, we are interested in tracking based on discrete geometrical models and the optical flow estimators.
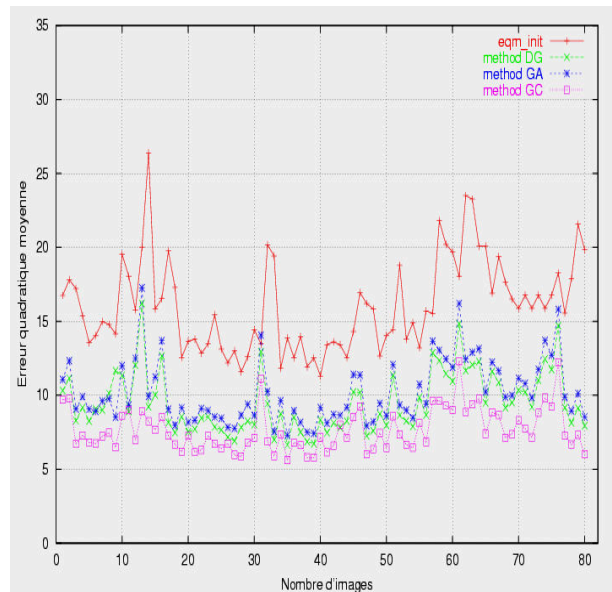
(a)
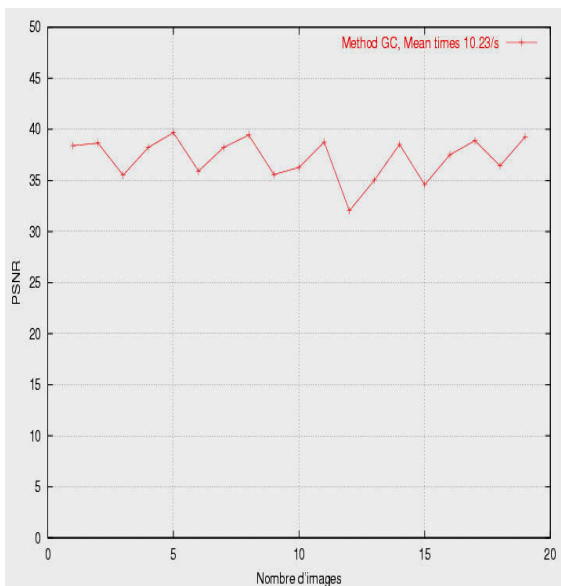
(b)
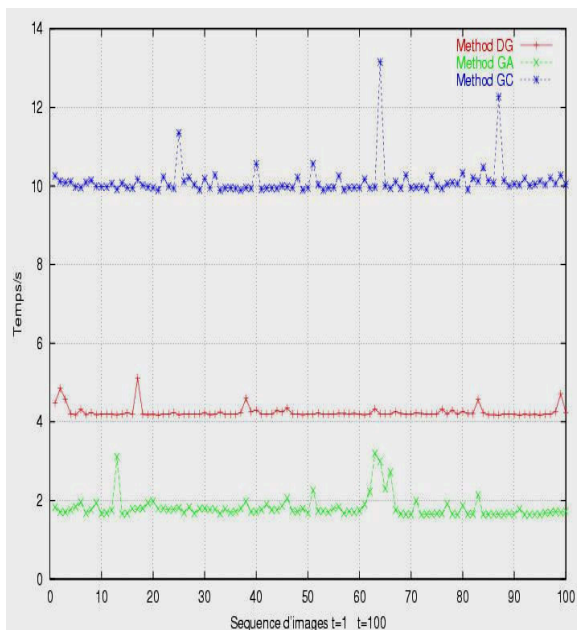
(c)

Fig2: Results of motion estimation. Sequence "Tympanon": a) MSE, b) PSNR of conjugate gradient, c) Computational time

V. CONCLUSION

The performance of the interactive color based segmentation method is estimated by comparing the quality (PSNR and MSE of motion compensation) and performance (time).

The conjugate gradient method has been used especially for a complex sequence (Boat) because it increased the speed of convergence. We can now conclude the results in terms of comparison:

1. Conjugate Gradient had a best optimization of the error functional (PSNR/MSE). Nevertheless, it is costly in terms of computation (time).
2. Accelarated Conjugate Gradient method had a strong reduction in time (~50%) but less optimization of MSE

REFERENCES

[1] A. M. Aree, *motion estimation by differential methods for MPEG2 video coding,* Journal of Dohuk University,2005, Vol.8, No.2.

[2] S. Jehan-Besson, M. Barlaud, G. Aubert, Video object segmentation using Eulerian region-based active contour, IEEE International Conference on ICCV, 2001, Vol. 1, pp. 353-360.

[3] W. Wei and N. King, Automatic Video Object Segmentation for MPEG-4. School of Computer Engineering, 2003, Nanyang Technological University, Singapore.

[4] S. Jehan, M. Barlaud, and G. Aubert, Detection and tracking of moving objects using a new level set based method, 2000, ICPR.

[5] M. Changick and H. Jenq-Neng, *Video Object Extraction for Object-Oriented Applications*. Journal on VLSI, 2001, 29(1), pp. 7-21.

[6] B. G. Shunck and B. P. Horn, *Determining optical flow*. Artificial Intelligence**,** 1981, Vol(17) :185-203.

[7] D. R. Walker and K. R. RAO, Improved pel-recursive motion compensation. Artificial Intelligence, 1981,Vol (32): 1128-1134.

[8] F. Cafforio, The differential method for image motion estimation". Image sequence processing , 1983, 104-124

[9] J. R. Shewchuk, An introduction to the conjugate gradient method. School of Computer Science, 1994.