MOVING OBJECT DETECTION IN COLOR IMAGE SEQUENCES USING REGION-LEVEL GRAPH LABELING

Ronan Fablet[1], *Patrick Bouthemy*[2] *and Marc Gelgon*[2]

[1]IRISA/CNRS
[2]IRISA/INRIA
Campus Universitaire de Beaulieu,
35042 Rennes Cedex, France
e-mail : rfablet@irisa.fr
Tel : (33) 2.99.84.25.23 Fax : (33) 2.99.84.71.71

Proc. 6th IEEE Int. Conf. on Image Processing, ICIP'99, Kobe, October 1999

ABSTRACT

We aim at detecting moving objects in color image sequences acquired with a mobile camera. This issue is of key importance in many application fields. To accurately recover motion boundaries, we exploit a fine spatial image partition supplied by a MRF-based color segmentation algorithm. We introduce a region-level graph modeling embedded in a Markovian framework to detect moving objects in the scene viewed by a mobile camera. This is stated as the binary segmentation into regions conforming or not conforming to the dominant image motion assumed to be due to the camera movement. The method is validated on real image sequences.

1. PROBLEM STATEMENT

Extracting moving objects from image sequences is of major interest in numerous applications : target tracking, video surveillance, vehicle navigation, video indexing, ... A complete motion-based segmentation is often not required, but only the extraction of some meaningful moving entities. Motion detection remains an important and difficult task to cope with when the camera is itself mobile.

Motion detection techniques usually rely on pixel-level classification schemes exploiting local motion-related information (typically, the DFD, Displaced Frame Difference, or the normal flow). The classification step is achieved either using thresholding techniques, [6, 7, 13], or a Bayesian framework, [11, 14]. Besides, as far as motion-based segmentation is concerned, pixel-level and region-level labeling are often exploited, [1, 4, 15, 9, 16, 17]. The computation of a primary layer of spatial regions is processed either relying on motion-based criterion, [16, 17], or on intensity, texture and color information, [1, 4, 15]. The second type of techniques usually supplies a better localization of motion boundaries, likely to correspond to intensity, texture or color contours. In fact, starting from this initial spatial partition, a 2D parametric motion model, generally an affine one, is attached to each spatial region, and spatial regions are merged according to motion properties. To this end, usual techniques rely either on clustering schemes in motion parameter space, [15], or on MDL criterion, [17], or on Markovian graph labeling approach, [4]. The major drawback of these approaches is that they prevent from processing a very fine spatial segmentation, since parametric motion estimators generally require a significiently large estimation support to be reliable. Thus, it may result in the loss of motion boundaries.

In this paper, we describe a region-level approach with a view to directly detecting moving objects in the scene from a color image sequence acquired by a mobile camera. As in [4], we determine a primary spatial color-based partition. Nevertheless, our method does not require to attach a parametric motion model to each extracted region. We only compute an estimation of the dominant image motion, and we benefit from the integration of local motion-related measures to determine the relevance of the estimated dominant motion in each spatial region.

Thus, our motion detection scheme involves three steps. First, we compute the 2D affine motion model accounting for the dominant image motion, (which is usually the case). Second, a spatial graph, whose nodes correspond to spatial regions, is derived from the color-based segmentation ; third, a Markovian framework is introduced to assign to each node of the graph a binary label stating if a region is conform or not to the dominant motion. If the latter is due to camera motion, the set of regions labeled as non-conform includes moving objects in the scene. We design an appropriate energy function to tackle this issue : it involves the integration of local motion-related measures in the compensated sequence. Moreover, it enables to make use of a very fine spatial partition in order to accurately recover motion boundaries.

The sequel is organized as follows. In Section 2, the principle of our region-level Markovian graph labeling approach is introduced. Section 3 describes the different steps of our motion detection method. Finally, we report experimental results in Section 4, and Section 5 provides concluding remarks.

2. REGION-LEVEL GRAPH LABELING

Assuming that an initial fine spatial partition of the image has been determined, we aim at grouping regions based on color or motion criteria. To this end, we consider a Markovian labeling approach applied to the adjacency graph $\mathcal{G} = (\mathcal{N}, \mathcal{A})$ where $\mathcal{N}$ refers to the set of regions of the partition, and $\mathcal{A}$ to the set of arcs which relate two neighboring connected regions, [4]. We define a region-level MRF model the sites of which are the nodes of the graph. A two-site clique neighborhood system is then straightforwardly derived from the set of arcs relating two nodes. Adopting a MAP criterion and using the equivalence between Markovian and Gibbsian fields, [5], the grouping procedure comes to determine the label field $\hat{e}$ which verifies :

$$\hat{e} = \arg\min_{e} U(e, o) \tag{1}$$

where $U(e,o) = U^a(e,o) + U^b(e)$, with o the set of observations attached to each node of the graph, U^a the data-driven energy term, and U^b the regularization term. Both energy terms are split in the sum of local potentials V^a and V^b :

$$\left\{ \begin{array}{l} U^a(e,o) = \displaystyle\sum_{N \in \mathcal{N}} V^a(e_N, o_N) \\ \\ U^b(e) = \displaystyle\sum_{(N_1,N_2) \in \mathcal{A}} V^b(e_{N_1}, e_{N_2}) \end{array} \right. \tag{2}$$

The regularization potential V^b tends to favor identical labels for two neighboring regions. It takes into account their "degree" of adjacency through the computation of two geometrical features. It is expressed as follows :

$$V^b(e_{N_1}, e_{N_2}) = -\beta \frac{\alpha_{N_1 N_2}}{\alpha_{N_1 N_2} + D_{N_1 N_2}} \delta(e_{N_1} - e_{N_2}) \tag{3}$$

where β is a pre-set constant, $\alpha_{N_1 N_2}$ is the length of the common border of regions N_1 and N_2, and $D_{N_1 N_2}$ the Euclidean distance between the gravity centers of the two regions.

In fact, we will exploit this approach twice, to achieve region grouping with respect to color information (subsection 3.2) and to perform moving object detection (subsection 3.3).

3. MOVING OBJECT DETECTION

3.1. Motion estimation

The first step of our motion detection scheme consists in computing the dominant inter-frame motion represented by a 2D affine model. We assume that it is due to the camera movement. The velocity $w_\Theta(s)$, at a pixel s, related to the affine motion model parameterized by Θ is given by :

$$w_\Theta(s) = \left(\begin{array}{c} a_1 + a_2 x + a_3 y \\ a_4 + a_5 x + a_6 y \end{array} \right) \tag{4}$$

with $s = (x, y)$ and $\Theta = [a_1\ a_2\ a_3\ a_4\ a_5\ a_6]$. The computation is achieved with the gradient-based multiresolution incremental estimation method described in [10]. The following minimization problem is solved :

$$\widehat{\Theta} = \arg\min_{\Theta} \sum_{s} \rho\left(DFD(s, \Theta)\right) \tag{5}$$

where $DFD(s,\Theta) = I_{t+1}(s + w_\Theta(s)) - I_t(s)$ and $\rho()$ is Tukey's biweight function. The use of a robust estimator ensures the motion estimation not to be sensitive to secondary motions due to mobile objects in the scene. Criterion (5) is minimized by means of an iterative reweighted least-square technique embedded in a multiresolution framework and involving appropriate successive linearizations of the DFD expression.

3.2. Color-based spatial partitioning

Color information is an appropriate cue to recover accurate object boundaries in real dynamic scenes. As in [1, 4], stating that motion boundaries refer also to color contours, we first aim at determining an initial color-based partition of the image. We introduce a Markovian framework associated with a Gaussian modeling of the color distribution in each region. After a first stage involving a usual pixel-level segmentation, [8], we apply a region grouping step with a view to suppressing redundancies between color distributions.

The considered pixel-level procedure consists in iteratively estimating the Gaussian model attached to each label, standing for region number, by using the empirical moments, and in updating the label field by means of a Markovian regularization. We make use of a MAP criterion, and we define local potentials v^a and v^b relative respectively to the data-driven energy term and the regularization term. At each site s, the observation is supplied by the color component c expressed in the space described in [12]. It is given by $c = (c_1, c_2, c_2)$, where $c_1 = r - v$, $c_2 = 2b - r - g$ and $c_3 = r + g + b$, with (r, g, b) the color coordinates in the (red, green, blue) color space. Then, the data-driven potential v^a at each site s is defined by :

$$v^a(e_s, c(s)) = \eta_s(M_{e_s}, \Sigma_{e_s}) \qquad (6)$$

where $c(s)$ is the color vector at site s, (M_{e_s}, Σ_{e_s}) the Gaussian model attached to label e_s, and $\eta_s()$ the Gaussian error evaluated at site s and expressed by :

$$\eta_s(M_{e_s}, \Sigma_{e_s}) = (c(s) - M_{e_s})^t \, \Sigma_{e_s}^{-1} \, (c(s) - M_{e_s}) \qquad (7)$$

On the other hand, the term v^b favors the spatial homogeneity of the region partition :

$$v^b(e_r, e_s) = \mu(1 - \delta(e_r - e_s)) \qquad (8)$$

where μ is positive constant, (e_r, e_s) forms a second-order clique and δ is the Kronecker symbol.

In a second step, the approach described in Section 2 is exploited with a view to grouping regions which present similarities in terms of color distributions. Let us introduce the set of labels Λ, which initially refer to the different region numbers, and the associate Gaussian models $(M_\lambda, \Sigma_\lambda)_{\lambda \in \Lambda}$. Then, considering a given node graph N, the data-driven potential V^a_{coul} computed at N quantifies the ability of a Gaussian model, associated to label λ, to describe the color distribution relative to N. If N is also labeled λ, the potential V^a_{coul} is expressed as :

$$V^a_{coul}(e_N = \lambda, o_N) = \sum_{s \in \mathcal{R}_N} \eta_s(M_\lambda, \Sigma_\lambda) \qquad (9)$$

where $\eta_s()$ is the Gaussian error at site s introduced in equation (7). When evaluating a new label λ' at site N currently labeled by λ, we compute in fact the loss of information when considering the model $(M_{\lambda'}, \Sigma_{\lambda'})$ instead of the model $(M_\lambda, \Sigma_\lambda)$. Besides, we introduce an additional information in the regularization term. It consists in favoring identical labels for neighboring regions which present a weak color contrast on their common boundary. We weigh the regularization potential (equation 3) by a coefficient related to this color contrast. Finally, when visiting a given node N, we only take into account the labels present in its neighborhood, and after assigning N with the best current label $\hat{\lambda}$, the model associated to $\hat{\lambda}$ is re-estimated.

The color-based segmentation procedure is embedded in a multiscale framework. We first assign a different label to each block at the coarsest scale of the pyramid. Hence, performing a coarse-to-fine strategy, we iterate at each scale the pixel-level regularization stage and the region grouping step. In both cases, the minimization is driven using the HCF algorithm [2]. Since we put no a priori on the number of color regions, our algorithm is unsupervised, which enables to handle a large variety of real situations. Besides, in order to reliably and accurately extract all the motion boundaries, which are assumed to correspond to color contours, we deal with a very fine spatial partition (typically, down to 50 pixels per region).

3.3. Motion detection

The motion detection stage consists in determining a binary labeling of the color-based partition in terms of regions conforming or not to the estimated dominant motion model $\widehat{\Theta}$. To this end, we exploit the graph labeling approach presented in Section 2.

This scheme first requires to define a data-driven potential V^a_{mvt} in order to quantify the relevance of the estimated dominant motion model $\widehat{\Theta}$ in a given region. We consider local motion-related measurements in the compensated image sequence. In that context, the DFD and the normal flow have been broadly used in order to detect outliers to the dominant motion distribution. Nevertheless, these quantities reveal really sensitive to the noise attached to the computation of the spatio-temporal derivatives of the intensity function. As a consequence, we prefer to consider a more robust local motion-related measurement, already used in [7, 11], which remains straightforwardly derived from intensity gradients.

The region-level observation o_N is a set $(\epsilon_s)_{s \in R_N}$ of pixel-level motion-related measurements while we still take into account color information. At each site s, the observation is given by a vector $\epsilon_s = (\epsilon^i_s)$, whose components are :

$$\epsilon^i_s = \frac{\displaystyle\sum_{p \in \mathcal{W}(s)} |DFD^i(p, \widehat{\Theta})| \cdot \|\nabla I^i(p)\|}{\max\left(G_m^2, \displaystyle\sum_{p \in \mathcal{W}(s)} \|\nabla I^i(p)\|^2 \right)} \qquad (10)$$

where i refers to the color component c_i, $\mathcal{W}(s)$ is a 3×3 window centered on s and G_m a predefined constant which prevents from dividing by zero in regions poorly textured and accounts for noise level. The quantity DFD^i is the DFD for the color component c_i. In [11], lower and upper interpretation bounds l_s and L_s

have been derived from the local spatial intensity gradient distribution to evaluate the information provided by this measurement *w.r.t.* a minimal residual motion magnitude Δ to be detected. This can be extended to color image as follows :

$$\left\{ \begin{array}{ll} \text{if } \epsilon_s^i < l_s^i(\Delta) & \text{then } \|w(s) - w_{\widehat{\Theta}}(s)\| < \Delta \\ \text{if } \epsilon_s^i > L_s^i(\Delta) & \text{then } \|w(s) - w_{\widehat{\Theta}}(s)\| > \Delta \end{array} \right. \quad (11)$$

where $w(s)$ is the real (unknown) motion at site s. By exploiting these bounds, the data-driven potential V_{mvt}^a for the region R_N corresponding to a node N in the graph is expressed by :

$$\left\{ \begin{array}{l} V_{mvt}^a(conf) = \sum_{s \in R_N} \frac{\max_i \left[A(\epsilon_s^i, l_s^i(\Delta)) \right]}{|R_N|} \\ V_{mvt}^a(nconf) = \sum_{s \in R_N} \frac{\max_i \left[1 - A(\epsilon_s^i, L_s^i(\Delta)) \right]}{|R_N|} \end{array} \right. \quad (12)$$

where α is a constant, *conf* and *nconf* respectively refer to "conform" and "non-conform" labels, and the function $A()$ is a smooth version of a step.

Since the number of nodes is small (compared to the size of the image), the minimization procedure can be efficiently and properly achieved using a HCF algorithm, [2], which consists in visiting nodes of the graph according to their rank in an unstability stack. This unstability at a given node refers to the difference of the potentials computed when considering each of the two binary labels. The initial detection map is obtained by considering only the data-driven term.

4. EXPERIMENTAL RESULTS

For all experiments carried out, the parameters of the algorithm were set as follows. For the MRF-based color segmentation, $\mu = 0.045$, $\beta_{coul} = 0.1$. For the motion detection stage, $G_m = 15.0$, $\alpha = 200.0$, $\beta = 100.0$ and $\delta = 1.0$. Factor Δ is typically a user-defined parameter, and has obviously a great importance in the number of regions detected as non-conform to the dominant motion. Since it has a straightforward physical meaning, it is easy to set it according to the application at hand. Besides, it is explicitly taken into account in a well posed manner in the method (potential V_{mvt}^a), and it confers an attractive flexible feature to the proposed method.

Satisfactory results have been obtained. In the first example, reported in Fig.4$(a-b)$, the camera moves to the top right corner and the car is driven forward while turning to the right. The use of color information allows us to extract quite relevant and accurate motion boundaries in spite of illumination effects. The second example, reported in Fig.4$(c-d)$, is a complex dynamic scene : the camera tracks a tennis player during the game. Thus, it involves large non-rigid motions which are difficult to handle. Again, the motion boundaries of the body are accurately recovered.

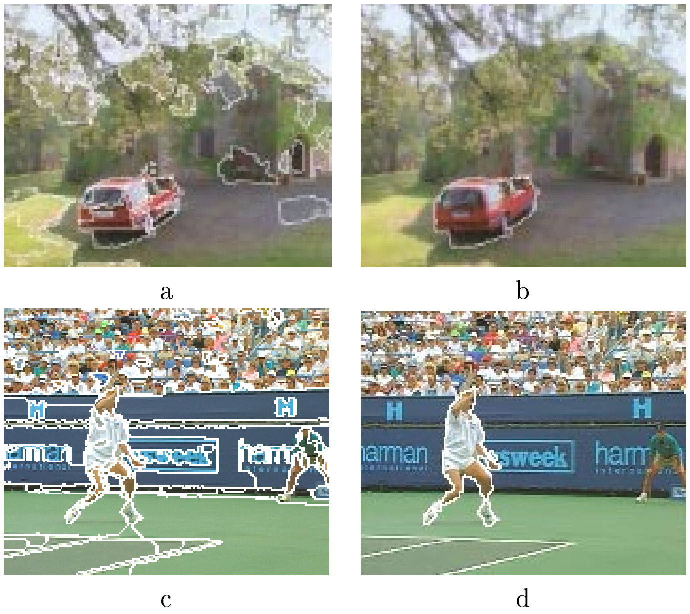
a
b
c
d

Figure 1: Results of our motion detection method on the sequences "Car" (a-b) "Stefan" (c-d) : color-based spatial boundaries (a-c), contours in white of the detected moving regions (b-d).

The computational time associated to the color-based segmentation stage is about one minute for a 512 × 512 image, whereas the region-level motion detection stage requires a few seconds for a graph containing about one hundred spatial regions (for a Sun Creator workstation 360MHZ).

5. CONCLUSION

We have presented in this paper a method to detect moving objects in color image sequences acquired with a mobile camera. We achieve an appropriate labeling of the adjacency graph of regions resulting from a spatial partition of the image based on a color criterion. Thus labeling separates regions conforming or not to the dominant image motion, represented by a 2D parametric motion model.

The use of a color-based criterion improves the accuracy of the localization of motion boundaries. Thanks to region-level approach, we can exploit contextual information at a higher-level than in classical pixel-level techniques, which enables us to be closer to the notion of "object" as demonstrated in the reported results.

In future work, in order to take into account perspective effects not handled by the 2D dominant image motion model, we aim at computing this method with recent developments described in [3].

Acknowledgments

The authors are thankful to INA, Département Innovation, Direction de la Recherche, for providing the MPEG-1 sequence "Ajax".

6. REFERENCES

[1] Y. Altunbasak, P. Ehran Eren, and A. Murat Tekalp. Region-based parametric motion segmentation using color information. *Graphical Models and Image Processing*, 60(1):13–23, January 1998.

[2] P.B. Chou and C.M. Brown. The theory and practice of Bayesian image modeling. *Int. Journal of Computer Vision*, 4:185–210, 1990.

[3] G. Csurka and P. Bouthemy. Direct identification of moving objects and background from 2D motion models. In *Proc. 7th IEEE Int. Conf. on Computer Vision, ICCV'99*, Kerkyra, Greece, September 1999.

[4] M. Gelgon and P. Bouthemy. A region-level motion-based graph representation and labeling for tracking a spatial image region. *Pattern Recognition.* To appear, 1999.

[5] S. Geman and D. Geman. Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 6(6):721–741, 1984.

[6] M. Irani and P. Anandan. A unified approach to moving object detection in 2D and 3D scenes. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 20(6):577–589, June 1998.

[7] M. Irani, B. Rousso, and S. Peleg. Detecting and tracking multiple moving objects using temporal integration. In *Proc. 2nd Eur. Conf. on Computer Vision, ECCV'92*, Santa Margherita, May 1992.

[8] Y.G. Leclerc. Constructing simple stable descriptions for image partitioning. *Int. Journal of Computer Vision*, 3:73–102, March 1989.

[9] L.Wu, J. Benois-Pineau, Ph. Delagnes, and D. Barba. Spatio-temporal segmentation of image sequences for object-oriented low bite rate image coding. *Signal Processing : Image Communication*, 8:513–543, 1996.

[10] J.M. Odobez and P. Bouthemy. Robust multiresolution estimation of parametric motion models. *Jal of Visual Communication and Image Representation*, 6(4):348–365, December 1995.

[11] J.M. Odobez and P. Bouthemy. Separation of moving regions from background in an image sequence acquired with a mobile camera. In *Video Data Compression for Multimedia Computing*, chapter 8, pages 295–311. H. H. Li, S. Sun, and H. Derin, eds, Kluwer Academic Publisher edition, 1997.

[12] M.J. Swain and D. Ballard. Color indexing. *Int. Journal of Computer Vision*, 7(1), 1991.

[13] W.B. Thompson, P. Lechleider, and E.R. Stuck. Detecting moving objects using rigidity constraint. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 15(2):162–165, February 1993.

[14] P.H.S. Torr and D.W. Murray. Statistical detection of independent movement from a moving camera. *Image and Vision Computing*, 11(4), May 1993.

[15] J.Y.A. Wang and E.H. Adelson. Representing moving images with layers. *IEEE Trans. on Image Processing*, 3(5):625–638, September 1994.

[16] W. Xiong and C. Graffigne. A hierarchical method for detection of moving objects. In *Proc. 1st IEEE Int. Conf. on Image Processing, ICIP'94*, pages 795–799, Austin, November 1994.

[17] H. Zheng and D. Blostein. Motion-based object segmentation and estimation using the MDL principle. *IEEE Trans. on Image Processing*, 4(9):1223–1235, September 1995.