REGION-LEVEL MOVING OBJECT SEGMENTATION BY GRAPH LABELING

I. Grinias and G. Tziritas

Department of Computer Science, University of Crete
P.O. Box 2208, 71409, Heraklion, Greece
phone: + (30) 2810 393517, fax: + (30) 2810 393501, email: {grinias, tziritas}@csd.uoc.gr

ABSTRACT

In this paper we propose a method for the detection and localization of moving objects. The change detection problem in the pixel domain is formulated by two zero mean Laplacian distributions. Furthermore, the image is split in homogeneous colour regions and their inter-frame mean absolute difference is used to describe the change detection problem in the region level by two *Gamma* distributions. The pixel and region based change detection statistics are used to classify the colour regions as "changed" or "unchanged" with high confidence. These initially labeled regions constitute the "seeds" of the "changed"/"unchanged" classes. The remaining unlabeled regions are classified as belonging to one of them using a growing algorithm, which has been modified to refer to the labeling of regions (instead of pixels). Class growing is accomplished using the change detection and boundary information of unlabeled regions. The interconnection between region-nodes is represented by a region adjacency graph.

1. INTRODUCTION

Video segmentation is a key step in determining the motion features, as well as the position and the 2D shape of the scene objects. Such a description may be used either for coding purposes in order to reduce storage and transmition requirements or for indexing and retrieval purposes in order to improve the content description and storage reduction of visual databases. The development of the corresponding international standards MPEG-4 for coding and MPEG-7 for visual content description, which both rely on the concept of audio/visual objects, has raised the importance of these methods.

Several approaches have been proposed for spatio-temporal video segmentation. A recent overview of segmentation tools as well as the object-oriented video description are presented in [1]. In [2], we proposed an object localization algorithm in which change detection is based on Bayesian tests that are applied on the inter-frame difference, while object localization is achieved using the object colour information. In a number of methods, object extraction is applied on the spatial partition of the image in homogeneous regions (region-level instead of pixel-level based extraction) in order to reduce the spatio-temporal redundancy of video images and to speed up and robustify the computations [3] [4] [5]. Video object extraction could be based then on change detection and moving objects localization or on motion field segmentation of the spatial regions.

We follow the region-level approach in the proposed segmentation system which is depicted in Fig. 1. The system is divided in three layers of computation. In the first one (top-down order), the basic segmentation characteristics are evaluated by the corresponding modules in the order implied by the arrows in Fig. 1. Hence, (i) first the change detection mixture parameters are computed using the pixel inter-frame difference over all the image pixels, (ii) the colour regions of the current image are extracted and after that (iii) the region-based change detection statistics are computed using a region-based change detection feature.

The second level in Fig. 1 serves as the intermediate level between the first and the third one, since it uses the change detection parameters and colour regions, which have been extracted by the first level in order to produce an initial labeling, which will be expanded by the third. The top middle level task is to split the colour regions obtaining sub-regions which can be labeled as "changed" with high confidence. The pixel-based change detection statistics are used for this splitting. The second module involves the labeling of the remaining (sub-) regions as "unchanged" using the region-based statistics of the previous level.

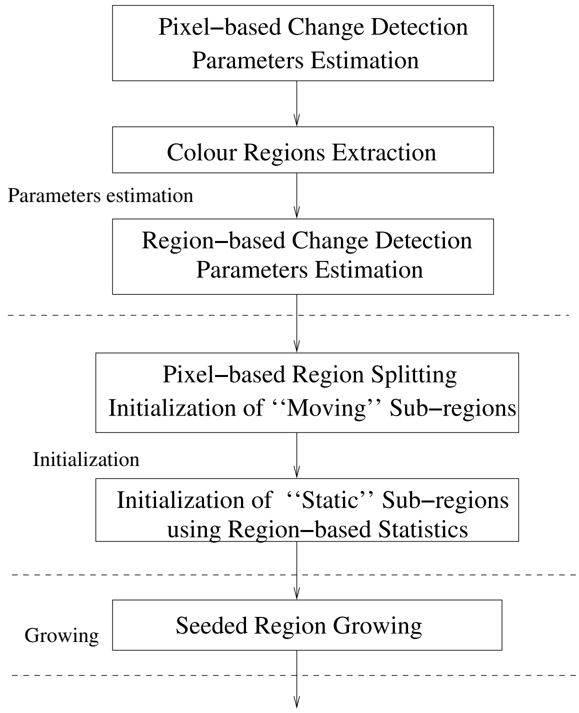

Figure 1: System framework.

Finally, the last level consists of the initial labels expansion task. The overall system is explained in detail below beginning with the colour regions extraction method, since the computation of the pixel-based change detection statistics is that of [2].

2. REGION INTER-FRAME DIFFERENCE

The first step involves the partitioning $\mathbf{R}$ of the image in N homogeneous colour components, in the *YCbCr* colour space. The well known k-means algorithm is employed to compute the dominant *YCbCr* colours, which are then used to extract the independent colour regions R_i of the image. Denoting by   the overall image points, the following equations hold for the final partition:

$$R_i \cap R_j = \emptyset \text{ and } \quad = \bigcup_{i=1}^{N} R_i \tag{1}$$

The segmentation algorithm is mainly based on *change detection*. The appropriate statistics involve not only the inter-frame difference, as was the case in [2], but also the mean of absolute differences of colour regions R_i:

$$d(R_i) = \frac{1}{|R_i|} \quad _{(x,y)\in R_i} |I_t(x,y) - I_{t+\ t}(x,y)| \tag{2}$$

where $|R_i|$ denotes the cardinality of region R_i and I_t (resp. $I_{t+\ t}$) is the intensity frame at time t (resp. $t+\ t$).

As it is described in [2], the two classes of "changed"/"unchanged" *pixels* are modeled by two *Laplace* distributions. Experimental results have shown that in the case of the mean absolute difference of colour regions, the two classes of "changed"/"unchanged" regions follow the *Gamma* distribution. Let $D = \{d(R_i), 1 \leq i \leq N\}$ denote the mean of gray level differences of each color region. The change detection problem consists of determining a binary label $\ (R_i)$ for each region R_i of the image. We associate the random field $\ (R_i)$ with two possible events, $\ (R_i) = static$ ("unchanged" region), and $\ (R_i) = mobile$ ("changed" region). Let $p_{D|static}(d|static)$ (resp. $p_{D|mobile}(d|mobile)$) be the probability density functions of the observed mean absolute inter-frame region difference under the H_0 (resp. H_1) hypothesis. These probability density functions are assumed to be *Gamma* for both hypotheses ($l = 0; 1$):

$$p(d(R_i)|\ \ (R_i) = l) = \frac{d(R_i)^{a_l} e^{-\frac{d(R_i)}{\ l}}}{(a_l + 1)\ \ _l^{a_l+1}}$$

Let P_0 (resp. P_1) be the a priori probability of hypothesis H_0 (resp. H_1). Thus the probability density function is given by

$$p_D(d) = P_0\ \ p_{D|0}(d|static) + P_1\ \ p_{D|1}(d|mobile) \tag{3}$$

In this mixture distribution $\{P_l, a_l,\ \ _l, l \in \{0,1\}\}$ are unknown parameters.

The experimental results in our effort for a robust region-based mixture decomposition, shown that it is sufficient to investigate only integer values of a_l ($l = 0,1$). Nevertheless, in most cases holds that $a_0 \geq a_1$. These observations lead us to a straightforward method for the estimation of mixture distribution parameters, by evaluating the $\ ^2$ criterion between the histogram of D and the mixture distribution (eq. (3)) that is obtained for a finite set $\mathbf{a} = \{0, 1, \ldots, A\}$ of integer values of a_l, under the restriction that $a_0 \geq a_1$. Furthermore, the principle of Maximum Likelihood is used to obtain an estimate for P_l and $\ _l$, ($l = 0,1$), for each investigated pair of a_0, a_1. The set of parameters $\{\hat{P}_l, \hat{a}_l, \hat{\ }_l, l \in \{0,1\}\}$ which minimizes the $\ ^2$ metric is selected as the better estimate for the mixture distribution of eq. (3).

3. CHANGE DETECTION USING SRG ON REGIONS

In what follows, we describe an extension of the well known *Seeded Region Growing (SRG)*[6] algorithm. In the extended algorithm the classes that are to be grown, are classes of regions instead of pixels and the same holds for the initially unlabeled items which are regions and not pixels. This modified algorithm is used to segment the image in the two classes of "changed"/"unchanged" regions, as it is described below.

3.1 Initialization

The growing algorithm requires a number of initial correctly labeled items. In our case, these are the colour regions which may be considered "static" or "moving" with high confidence. The confidence measurements are performed in both the pixel and region-based change detection statistics.

The first observation is that some "static" regions may contain a number of subregions with high inter-frame difference due to their overlapping with "mobile" regions and thus have to be split further. The splitting is performed using the pixel-based statistics of change detection. As in [2], the pixels that may be considered "changed" with high confidence are determined using the decision threshold:

$$T_1 = \frac{1}{\ _0} \ln \frac{1}{P_F},$$

where P_F is the given small false alarm probability and $\ _0$ is the estimated laplacian parameter of "unchanged" pixels. Then, the connected "changed" pixels of each region are grouped to form new regions, which constitute the "changed" sub-regions of high confidence. The remaining region pixels are also grouped in connected sub-regions, leading to a new image partition

$$\mathbf{R}' = \{R'_i, 1 \leq i \leq M\},$$

where M is the number of image regions. Next, the "unchanged" regions with high confidence are determined among the non "changed" sub-regions, using the decision criterion $d(R'_i) \leq T_0$, where T_0 satisfies the equation

$$P_{ND} = Pr\{d \leq T_0 | mobile\},$$

for a given small probability P_{ND} of not detecting a "changed" region.

3.2 Growing

The modified *SRG* algorithm is applied on the initial labeled regions in order to "grow" them. *Growing* refers now to regions instead of pixels and its effort is to assign the label "changed" or "unchanged" to the initially unlabeled regions. Each one of the two labels is grown according to dissimilarity criteria which are based on the label, the mean absolute difference and the boundary information of regions.

A label-depended term is set according to the *a-posteriori* probability principle. Assuming that the change detection statistics of each label follow the Gamma distribution, the dissimilarity of a colour region R from a label l is measured as

$$DIS_l(R) = \frac{1}{Pr(l(R)|d(R))} \tag{4}$$

Using the Bayes rule

$$Pr(l(R)|d(R)) = \frac{p(d(R)|l(R))Pr(l(R))}{\ _k\, p(d(R)|k(R))Pr(k(R))}$$

which gives

$$DIS_l(R) = 1 + \frac{\ _{k\neq l}\, p(d(R)|k(R))Pr(k(R))}{p(d(R)|l(R))Pr(l(R))}$$

Ignoring the constant term in the last equation and taking the logarithm of the second term gives

$$\begin{aligned} dcd_l(R) &= \ln(\ \ _{k\neq l}\ p(d(R)|k(R))Pr(k(R))) \\ &\quad - \ln(p(d(R)|l(R))Pr(l(R))) \end{aligned}$$

In our case of change detection the metric for label 0 becomes

$$dcd_0(R) = ln(p(d(R)|1)Pr(1)) - ln(p(d(R)|0)Pr(0))$$

and under the Gamma distribution assumption this gives

$$\begin{aligned} dcd_0(R) &= (a_1 - a_0)\ln d(R) + d(R)(\frac{1}{\ _0} - \frac{1}{\ _1}) \\ &\quad + \ln\ \ (a_1+1)Pr(1) + (a_1+1)\ln\ \ _1 \\ &\quad - \ln\ \ (a_0+1)Pr(0) - (a_0+1)\ln\ \ _0 \end{aligned}$$

and $dcd_1(R) = -dcd_0(R)$. Since $Pr(l)$, $(l = 0,1)$ are only estimates and not a-priori knowledge, they have been set to 0.5 in the current implementation of criterion dcd_0.

Furthermore, a boundary term dbd_l has been added to the label growing criterion:

$$dbd_0(R) = -\frac{b_0 - b_1}{\sqrt{|R|}}$$

and $dbd_1(R) = -dbd_0(R)$, where b_0 (resp. b_1) is the common boundary length between R and the regions that have been labeled as "unchanged" (resp. "changed") while $|\cdot|$ denotes the cardinality of its argument. The effect of the boundary term is to bypass the difficulties that arise in uniform regions which are parts of moving objects although their mean inter-frame difference is low. By minimizing dbd_l locally, the total common boundary between the two classes tends to be minimized.

The total dissimilarity $_l(R)$ is then defined as

$$_l(R) = f_{dcd}dcd_l(R) + f_{dbd}dbd_l(R)$$

where f_{dcd} is defined as:

$$f_{dcd} = \begin{cases} \frac{|R|}{100}, & \text{if } |R| \leq 5 \\ 1, & \text{otherwise} \end{cases}$$

and is used to decrease the effect of the "change detection" measurement in small regions, where mean difference estimation is often insufficient. By contrary, f_{dbd} is a binary decision factor:

$$f_{dbd} = \begin{cases} 1 & \text{if } |R| \leq 500 \\ 0, & \text{otherwise} \end{cases}$$

which implies that large enough regions cannot be treated in the same way that boundary pixels are used in order to enforce the smoothness of the boundary between the growing classes. It should be noticed that when $f_{dbd} = 0$, the overall criterion is solely based on "change detection" statistics, since the measurements in that case can be considered accurate. Apparently from the limitations that are imposed in the size of regions above, the criterion $_l(\cdot)$ tends to give more emphasis to the "change" detection statistics as the size of regions becomes larger, since the boundary term $dbd_l(\cdot)$ decreases with size. This is an admirable property of the overall criterion which is achieved without any further tuning.

Furthermore, the distance $a(R)$ between the center of mass of region R and the boundary of the previous "change" mask has been introduced in metric $_1(R)$ which after all becomes:

$$_1(R) = dcd_1(R) + dbd_l(R) - a(R).$$

This "memory" term is used only for regions that their area was exclusively included in the "moving" objects of the previous "change" mask. The distinction between the two labels is justified by the fact that in frames, which undergo a small motion and contain large uniform background areas, a large part of the "unchanged" area is labeled at the initialization stage. By contrary, the "changed" label is initialized in small regions leading to a mismatch for that label at the "growing" stage. Labels are expanded using the *SRG* algorithm for regions instead of pixels. For the implementation of *SRG* a list that keeps its members ordered according to the dissimilarity criterion $(\cdot)$ is used, traditionally referred to as Sequentially Sorted List (SSL). In addition, $\mathbf{R}'$ is represented as a set of nodes in a connected undirected graph called *region adjacency graph* (RAG). Two nodes g_i and g_j of the graph are connected by an edge, if and only if, the corresponding regions R_i and R_j are adjacent. Finally, we define the set of indices $\mathbf{L} = L_0, \ldots, L_M$ to the class $L_i = l$ whose statistics give the minimum $_l(\cdot)$ value for the region R_i', $(1 \leq i \leq N)$. The complete SRG algorithm is as follows:

S1 Label the initial colour regions of classes 0 and 1 (initialization stage).

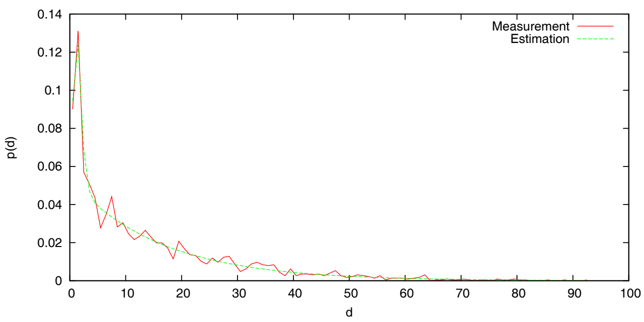

Figure 2: Region-based mean absolute inter-frame difference statistics for frame 42 of the sequence *"Mother"*.

S2 Insert all the unlabeled spatial neighbors of the initial regions into the SSL. If they adjoin both the two classes use as $_l(\cdot)$ that with the minimum value. Update properly their L value.

S3 While the SSL is not empty:

S3.1 Remove the first region y from the SSL and label it according to its L label.

S3.2 Test the neighbors of y and update the SSL:

S3.2.1 Add to the SSL neighbors of y which are neither already labeled nor already in the SSL, according to their value of $(\cdot)$. If they adjoin and the other class, use as $_l(\cdot)$ that with the minimum value. Update properly their L value.

S3.2.2 Test for neighbors of y which are already in the SSL and promote them accordingly in the SSL:

S3.2.2.1 if they border on and the other class, insert them in the SSL using as $_l(\cdot)$ that with the minimum value,

S3.2.2.2 otherwise, insert them using the $_l(\cdot)$ of y' s label.

Update properly their L value.

Each step of the modified SRG algorithm labels the minimum element-region of SSL y and a number of tests on yet unlabeled neighbors of y are performed followed by a constant number of insertions and deletions in SSL. Although in the first implementation [6] the unlabeled items were inserted only once in the SSL and their $(\cdot)$ value was not updated until their labeling, the computational cost of the modified SRG algorithm that we present still remains low since (a) SSL is implemented using AVL trees in which the computational cost of insertions and deletions for M items is $O(M \log M)$, (b) the number of colour regions M is small –a few hundreds– compared to that of pixels and the number of unlabeled regions is even smaller, (c) the number of the neighbors of each region is usually in the same order of the eight neighbors of pixels, when 8-connectivity is considered and (d) the criterion value $dcd_l(\cdot)$ can be computed only once per each unlabeled region and kept in memory, since it remains unchanged during SRG iterations and the same holds for the "memory" term $a(\cdot)$ and finally, (e) the dynamically updated local boundary term $dbd_l(\cdot)$ is computed sufficiently in low cost.

4. EXPERIMENTAL RESULTS

In what follows, we present the results that were obtained by the object detection system for the image sequences *"Mother"* and *"HallMonitor"*, which have been included in the COST testing data set. The camera in both sequences is static, while in *"HallMonitor"* the background is known.

As we can see in the localization result of Fig. 3, *"Mother"* is characterized by low spatial detail while the two objects move very slowly or do not move at all during a large number of frames. The large homogeneous regions together with the low objects movement

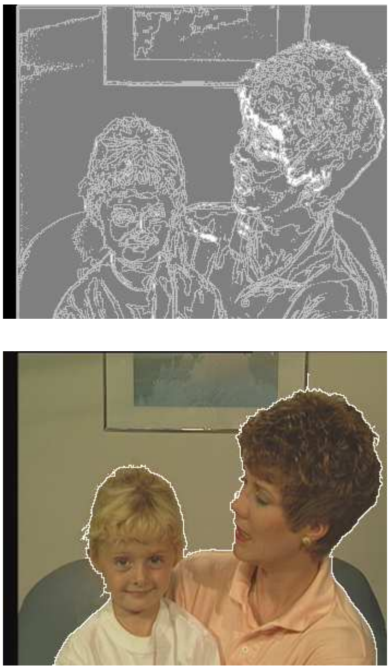

Figure 3: Initialization (up) and moving objects extraction result (down) for frame 42 of the sequence *"Mother"*.

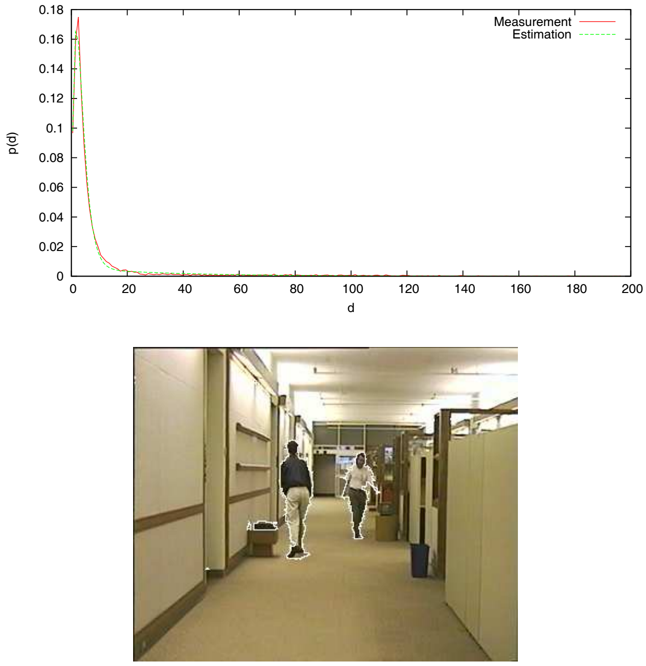

Figure 4: Inter-frame regions mean absolute difference statistics (up) and moving objects extraction result (down) for frame 148 of the sequence *"HallMonitor"*.

lead to low discrimination of the statistics of classes "changed" and "unchanged", which in turn affects the efficiency of algorithms that are based on change detection. The proposed system overcomes this difficulty by using regions instead of pixels in order to initialize the "unchanged" regions under the Gamma distribution assumption. In Fig. 2, "Measurement" refers to the histogram of the mean absolute difference of the 2300 regions that were extracted by frame 42 of *"Mother"*, while "Estimation" is the computed mixture of eq. (3). The initially labeled regions for this frame are shown in black ("unchanged") and white ("changed") in the upper image of Fig. 3, while the gray regions are initially unlabeled. The light gray curves depict the boundary of regions. The memory term $a(\cdot)$ that has been introduced in the change detection part $dcd_1(\cdot)$ of the "changed" class growing criterion retains the moving objects classification to the "changed" class for a number of frames in which appear to be stationary. Thus, the objects are extracted efficiently, as it is shown by the bottom image of Fig. 3. The white curves in the images represent the boundary between the classes "changed" and "unchanged". However, since this method relies on change detection in order to determine the moving objects, the result has to be improved in the case of larger objects motion, in order to be able to cope with occlusions. For that purpose, the colour based objects localization method described in [2] may be applied on the output change detection map of our algorithm.

Finally, in the image of Fig. 4 we see the localization result for frame 148 of *"HallMonitor"*, while the plot of the figure refers to the inter-frame difference statistics of frame' s regions. Since the background of the sequence is known, the curve "Measurement" of the plot corresponds to the histogram of region differences between frames 148 and 0. For the same reason, the bag shown in the bottom result of Fig. 4, is bounded as "changed". The "growing" of classes is performed without using the "memory" term, because the shape and the position of the two humans of the sequence changes among frames.

5. CONCLUSION

We proposed a moving objects detection and localization method. The algorithm has been mainly based on change detection statistics. The image was over-segmented in homogeneous colour regions. Then, the regions that could be classified as "changed" or "unchanged" with high confidence were determined. The remaining regions were labeled by a class growing algorithm. Both the initialization and growing of the two classes were based on the change detection as well as the boundary information of regions. A region adjacency graph was used to represent the interconnection between colour regions. The algorithm is robust while it gives accurate, change detection based, moving object localization results.

REFERENCES

[1] P. Salembier and F. Marques, "Region-based representations of image and video: segmentation tools for multimedia services", IEEE Trans. on Circuits and Systems for Video Technology, 9:1147-1169, December 1999.

[2] E. Sifakis, I. Grinias, and G. Tziritas, "Video segmentation using fast marching and region growing algorithms", EURASIP Journal on Applied Signal Processing, pp. 379-388, April 2002.

[3] A.A. Alatan et al., "Image sequence analysis for emerging interactive multimedia services- the European COST 211 framework", IEEE Trans. on Circuits and Systems for Video Technology, 8:802-813, November 1998.

[4] I. Patras, E.A. Hendriks, and R.L. Lagendijk, "Video segmentation by map labeling of watershed segments", IEEE Trans. Pattern Analysis and Machine Intelligence, 22(3):326-332, March 2001.

[5] R. Fablet, P. Bouthemy, and M. Gelgon, "Moving object detection in color image sequences using region-level graph labeling", in Proc. 6th IEEE Int. Conf. on Image Processing, ICIP' 99, Kobe, October 1999.

[6] R. Adams and L. Bischof, "Seeded region growing", IEEE Trans. on Pattern Analysis and Machine Intelligence, PAMI-16:641-647, June 1994.