FORTH-ICS / TR-155

Detection and location of moving objects using deterministic relaxation algorithms

N. Paragios and G. Tziritas
Institute of Computer Science
Foundation for Research and Technology-Hellas, and,
Department of Computer Science, University of Crete
P.O. Box 1470, Heraklion, Greece
E-mails: parag@csd.uch.gr, tziritas@csi.forth.gr

Abstract

Two important problems in motion analysis are addressed in this paper: change detection and moving object location. For the first problem, the inter-frame difference is modelized by a mixture of Laplacian distributions, a Gibbs random field is used for describing the label field, and ICM (Iterated Conditional Modes) or HCF (Highest Confidence First) algorithms are used for solving the resulting optimization problem. The solution of the second problem is based on the observation of two successive frames alone. Using the results of change detection an adaptive statistical model for the couple of image intensities is identified. Then the labeling problem is solved using either ICM or HCF algorithm. Results on real image sequences illustrate the efficiency of the proposed method.

1 Introduction

Detection and location of moving objects in an image sequence is a very important task in numerous applications of Computer Vision, including object tracking, fixation and 2-D/3-D motion estimation. For a stationary observer, detection is often based only on the inter-frame difference. Detection can be obtained by thresholding, or using more sophisticated methods taking into account the neighborhood of a point in a local or global decision criterion. For a moving observer, the problem is much harder, since everything in the image may be changed. In this case, egomotion should be estimated and compensated to be able to detect independent motion.

This paper deals with two related problems, change detection and moving object location. Indeed, complete motion detection is not equivalent to temporal change detection. Presence of motion usually causes three kinds of "change regions" to appear. They correspond to (1) uncoverd static background, (2) a covered background, and (3) an overlap of two successive object projections. Note also that regions of third class are difficult to recover by a temporal change detector, when the object surface intensity is rather uniform. All this implies that a complementary computation must be performed after change detection, to extract specific information about the exact location of moving objects.

The simplest change detector is obtained by thresholding the difference of two consecutive frames. Pixels with absolute value above a certain threshold are considered as moving, whereas the other pixels are considered stationary, as proposed by N. Diehl [7]. An extension of this model, using a mixture decomposition for the observed difference (usually Laplace or Normal distributions are used for this reason), is proposed by G. Tziritas and C. Labit [19], where the use of a Maxixum A Posteriori probability criterion, gives an adaptive determination of the decision threshold. In both methods the decision is taken independently from point to point. Another approach for change detection is suggested by K. Karmann, A. Brandt and R. Gerl [12]. They use Kalman filtering of certain reference frames in order to adapt to changing image characteristics.

A different class of algorithms is that based on a statistical modelization of the context, to which the algorithms proposed in this paper belong. Cafforio and Rocca [5] assumed that pixels in difference frame are statistically independent and each pixel is a zero-mean Laplacian stochastic variable. The segmentation field is modeled as a first order Markov chain along rows and the solution was obtained by solving a Maximum A Posteriori (MAP) probability problem using the Viterbi algorithm. An extension of this model into a two-dimensional order causal Markov field is proposed by Mori, Rocca and Tubaro [14] and also by Driessen, Biemond and Boekeee [8]. The two methods differ only in the probability distribution function assumed for the pixel in the difference frame. In both methods transition probabilities have to be empirically obtained.

More sophisticated models are suggested by Lalande and Bouthemy [13], Sivan and Malah [18], Bouthemy and Odobez [4] and also Bouthemy and Lalande [3]. They use a spatiotemporal Markov Random Fields (MRFs) model through Gibbs distribution, and construct a cost function which is minimized using a deterministic relaxation algorithm. The main difference between [13], [18], [4] and [3] appears in the cost function. Also different iterative algorithms for the minimization are used. In [13] a deterministic relaxation algorithm is used, which may be sensitive to the initial segmentation. In [18] a more complicated problem is under consideration, because they assume four possible states on change detection problem, $\{\text{Texture, Smooth}\}\times\{\text{Static,Mobile}\}$. They use a multiresolution approach on two levels, where the ICM algorithm is proposed for the minimization of a cost function in each level. Also many initial segmentations are used to avoid local minima. In [4] motion detection is achieved through a statistical regularization approach, where particular attention has been paid to the definition of the energy function, which seems to be complicated and also somehow expensive. There is a multiresolution scene and the approach deals also with the case of mobile camera. Finally, [3] deals with moving object location problem using three frames, where the solution is derived by minimizing an energy function using an iterative deterministic relaxation scheme, which is independent of the size, intensity distribution motion magnitude, and direction of the moving objects. The main idea of this approach is to consider three successive images at instants t_1, t_2, t_3 to recover the moving object location at time t_2. Two binary temporal change maps, between t_1 and t_2 and between t_2 and t_3 are determined, then a basic logical-AND operation is performed on these two maps.

The proposed here change-detection and moving object location algorithms, use a MRF model, through Gibbs distribution, to describe globally the labeling problem. A mixture of two Laplacian distribution functions is used to model the inter-frame difference, and Gaussian distribution functions are used to model the intensities in the moving object location problem. Cost functions are constructed, based on the above distributions, and a MAP problem is solved using Iterated Conditional Modes (ICM) and Highest Confidence First (HCF) algorithms. The process has two basic steps. First the change detection problem is solved, and then moving object location is performed using change detection results. The proposed algorithm deals also with the case of a mobile camera and the detection of independent motion. In this case dominant motion (camera's motion) is estimated and then compensated, resulting in a background that appears as static.

In order to check the efficiency and the robustness of the proposed algorithms experimental results are presented with real image sequences. In the so-called *Trevor White* image sequence (Figure 1(a)) the camera is stationary. Whereas, in the so-called *Interview* image sequence (Figure 1(b)), the camera is moving, and an independent moving object also exists. These two image sequences are used as input to check the proposed methods.

The remainder of this paper is organized as follows. In Section 2 we deal with the change detection problem, where camera's motion estimation (if the camera is moving), definition of energy function, and proposition of appropriate algorithms for the minimization of cost function take place. The moving object location problem appears in Section 3, while Section 4 contains concluding remarks and future work. Finally, in an Appendix is presented a Maximum Likelihood (ML) estimator for mixture decomposition.

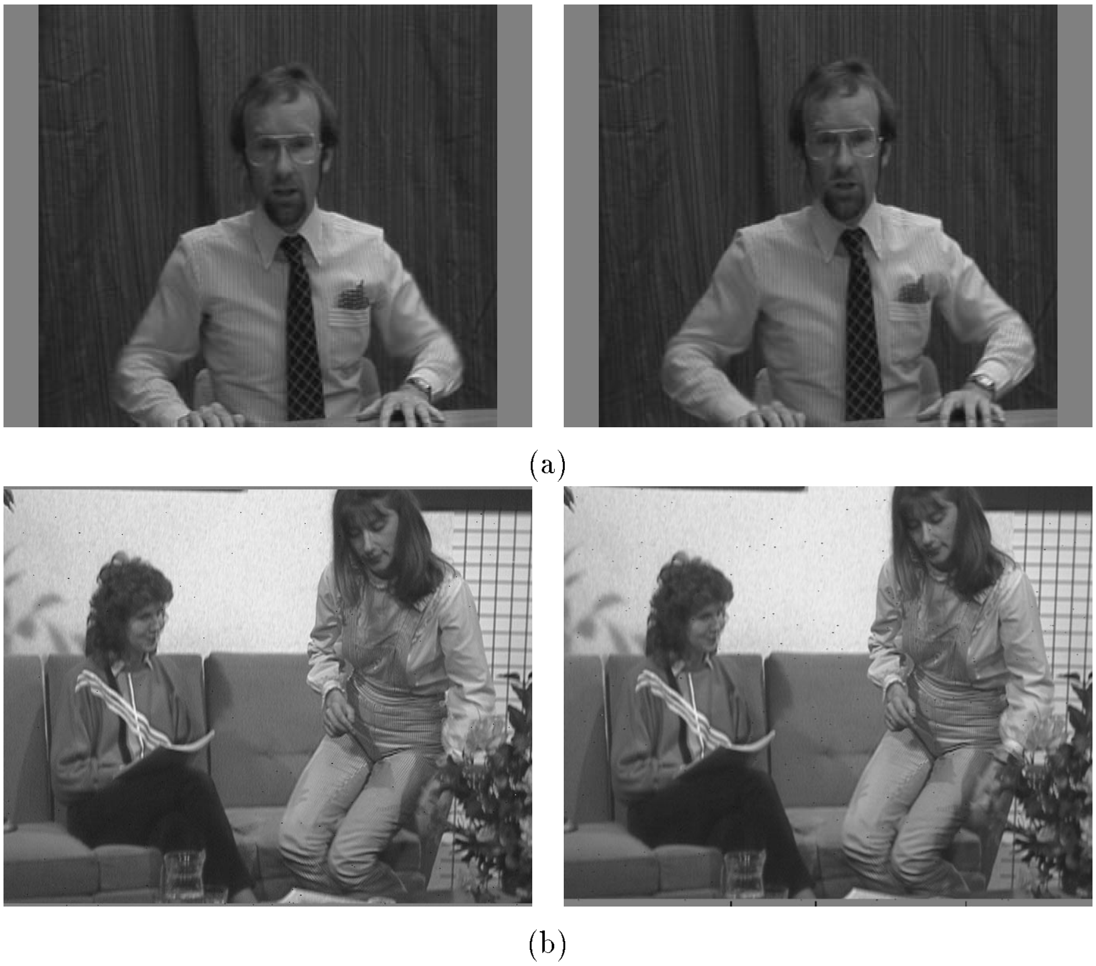


Figure 1: (a) *Trevor White* image sequence. (b) *Interview* image sequence

2 Change Detection

2.1 Dominant Motion Estimation

A very common hypothesis in change detection problem is the static camera, which holds in a large number of proposed solutions. An expected result is that these solutions cannot be used when they deal with a mobile camera. This constraint is raised, computing the dominant motion, using a gradient-based robust estimation method proposed by J. Odobez and P. Bouthemy [15], in order to create a compensated sequence in which only the motion of independent moving objects is still valid. An affine motion-model is considered defined by:

$$\begin{pmatrix} u(x,y) \\ v(x,y) \end{pmatrix} = \begin{pmatrix} a_{11} + a_{12}x + a_{13}y \\ a_{21} + a_{22}x + a_{23}y \end{pmatrix} \tag{1}$$

The estimation of the set of unknown parameters $\Theta = \{a_{ij}; i = 1, 2, \text{ and } j = 1, 2, 3\}$ between frames at time instant t and $t + 1$ is extracted as follows

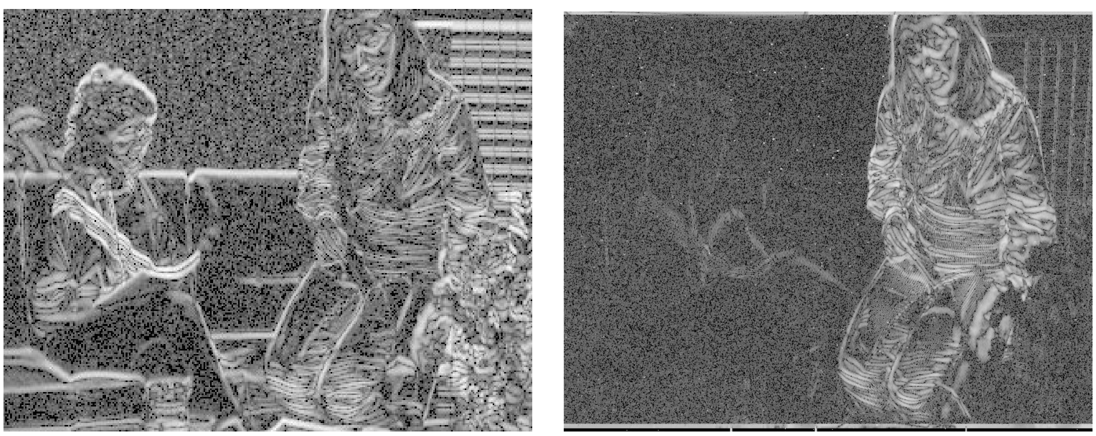
$$\hat{\Theta} = \arg\min \sum_{(i,j)} \rho(I(i + u(i,j), j + v(i,j); t + 1) - I(i,j;t)) \tag{2}$$

where $I(i,j;t)$ is the gray level of the observed image at pixel location (i,j) and time t, and

$$\rho(x) = \begin{cases} x^2, & \text{if } |x| \leq T \\ T^2, & \text{if } |x| > T \end{cases} \tag{3}$$

where T is a threshold determined experimentally, and in any case depending on the amplitude of the motion and the intensity gradient. The minimization is performed using a simple method, Iteratively Reweighted Least Squares (as proposed in [15]), with a binary weight, determined by the above mentioned threshold. This estimator allows getting a good estimation of the dominant motion (*i.e.* background apparent motion), if the affine motion model is sufficiently accurate, which is used to compute a compensated image sequence in which the background then appears as static.

In Figure 2 are given results of applying the above method to the *Interview* sequence, where camera's motion is only translational. The inter-frame difference after camera's motion compensation indicates the presence of independent motion.



(a) (b)

Figure 2: Estimated dominant motion on *Interview* sequence: $(\alpha_{11}, \alpha_{21}) = (5.75, -0.3)$. (a) Inter-frame difference. (b) Inter-frame difference image after camera's motion compensation.

2.2 Change Detection Algorithm

Let $D = \{d(i,j)\}$ denote the gray level difference image with

$$d(i,j) = I(i,j;t+1) - I(i,j;t)$$

The change detection problem consists of a "binary" label $\Theta(i,j)$ for each pixel on the image grid. We associate the random filed $\Theta(i,j)$ with two possible events, $\Theta(i,j) = s$ (Static), if the observed difference $d(i,j)$ supports the hypothesis for static pixel (H_0), and $\Theta(i,j) = m$ (Mobile), if the observed difference supports the alternative hypothesis H_1, for mobile pixel. Under these assumptions, for each pixel it can be written

$$\begin{aligned} H_0 &: \Theta(i,j) = s \\ H_1 &: \Theta(i,j) = m \end{aligned} \tag{4}$$

Let $p_{D|s}(d|s)$ (resp. $p_{D|m}(d|m)$) be the probability density function of the observed inter-frame difference under the H_0 (resp. H_1) hypothesis. These probability density functions are supposed homogeneous, *i.e.* independent of the pixel location and usually they are under Laplacian or Gaussian low. We use here a Laplacian distribution function to describe the statistical behavior of the pixels for both hypotheses, thus the conditional probability density function of the observed difference values is given by

$$p(D = d|\Theta(i,j) = l) = \frac{\lambda_l}{2} e^{-\lambda_l |d|};\ l \in \{s, m\} \tag{5}$$

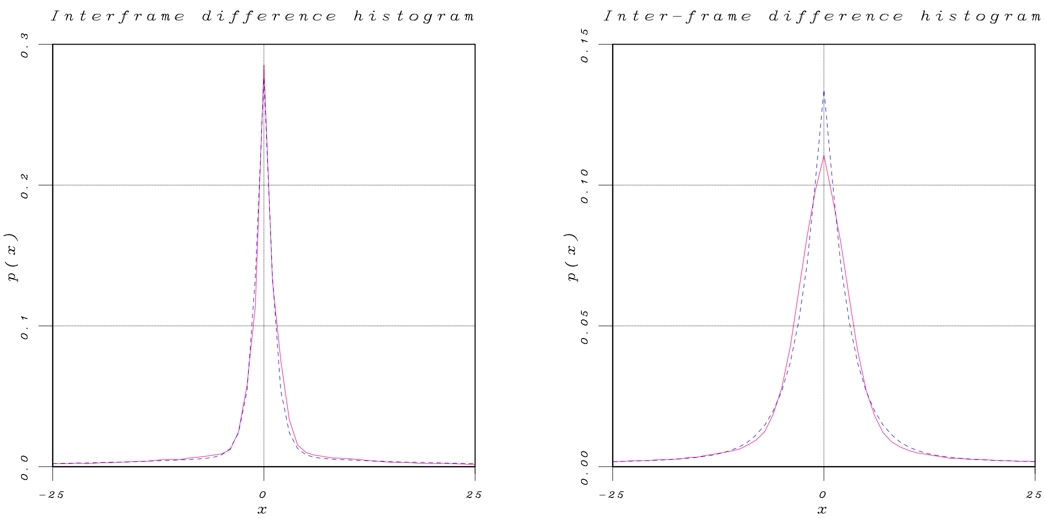
The parameter λ_l is related to the standard deviation σ_l by $\lambda_l = \frac{\sqrt{2}}{\sigma_l}$. Let P_s (resp. P_m) be the *a priori* probability of hypothesis H_0. Observed difference values are assumed to be obtained by selecting a label $l \in \{s, m\}$ with probability P_l and then selecting a d according to the probability low $p(D = d|l)$. Thus the probability density function is given by

$$p_D(d) = P_s p_{D|s}(d|s) + P_m p_{D|m}(d|m) \tag{6}$$

In this mixture distribution $\{P_l, \lambda_l; l \in \{s, m\}\}$ are unknown parameters. The principle of Maximum Likelihood is used to obtain an estimation of these parameters ([9], [17]). The unknown parameters are iteratively estimated using the observed grey level inter-frame differences. An initial estimation is calculated using first, second and third order moments of the variable considered (cf. Appendix). In Figure 3 are given the histogram and the approximated probability density function (dashed line) for both test sequences.

Under the above assumptions the change detection problem can be formulated as a scene labeling with contextual information. In such a framework, there are

- a set of sites $S = \{(i,j)\}$
- a set of possible labels for each site (here the same for each site)
  $A = \{s, m\}$
- a neighborhood relation, G, over the sites, which defines a graph where the vertices represents the sites, and the (weighted) edges represent the constraint on the label assignment of the neighboring sites.



(a) (b)

Figure 3: Mixture decomposition in Laplace distributions for inter-frame difference: (a) *Trevor White* sequence, (b) *Interview* sequence

The problem is to assign a label to each site in such a way, that the solution is consistent with the constraints.

Let ω be a labeling form of the image with respect to A as a realization of the set of random variables $\Theta = \{\Theta(i,j),\ (i,j) \in S\}$, and $\omega(i,j) \in A$ represents the label attached to the site (i,j) according to the labeling ω. The Static-Mobile decision field as it appears in our aproach is modeled as a MRF with a 8-pixel neighborhood (Figure 4). To describe $Pr(\omega)$ a Gibbs distribution is used, where only two-pixel cliques are considered in order to reduce the number of necessary parameters and the computational cost.

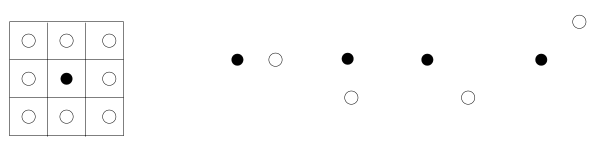


Figure 4: 8-pixel neighborhood : a possible choice of effective cliques

Using the local characteristics of the MRF, $p(d,\omega)$ is given by

$$p(d,\omega) = \frac{e^{-\frac{1}{T}U(d,\omega)}}{Z} \tag{7}$$

$U(d,\omega)$ is the energy function and Z is a normalizing constant. The function $U(d,\omega)$ can be decomposed into two terms

$$U(d,\omega) = U_1(\omega) + U_2(d,\omega) \tag{8}$$

- The first term, $U_1(\omega)$ accounts for the expected spatial properties (homogeneity) of the label field:

$$U_1(\omega) = \sum_{c \in C} V_c(\omega) \tag{9}$$

where C is the set of all two-pixel cliques in the whole frame, and the potential of a clique c, $V_c(\omega)$, is given by

$$V_c(\omega(i,j), \omega(i+k, j+l)) = \begin{cases} -\alpha_s & \text{if } \omega(i,j) = \omega(i+k, j+l) = s \\ -\alpha_m & \text{if } \omega(i,j) = \omega(i+k, j+l) = m \\ \alpha_d & \text{if } \omega(i,j) \neq \omega(i+k, j+l) \end{cases} \tag{10}$$

$(i+k, j+l)$ being a neighboor of (i,j) (here, $0 < k^2 + l^2 \leq 2$). The potential α_d is the cost to pay to get neighboors having different labels, α_s is a potential value which facilitates the selection of Static label, and α_m facilitates the selection of Mobile label ($0 < \alpha_d \ll \alpha_s, \alpha_m$).

- Energy U_2 expresses the adequacy between observed temporal differences and corresponding labels. The relation between the observation and the label given by $p(D = d(i,j) | \Theta(i,j) = \omega(i,j))$, the following expression is obtained

$$U_2(d, \omega) = -\sum_{(i,j)} \ln \left[p(D = d(i,j) | \Theta(i,j) = \omega(i,j)) \right] \tag{11}$$

The solution of the labeling problem is derived using a Maximum A Posteriori (MAP) criterion, *i.e.* the *a posteriori* distribution of the labels given the observations is maximized, which is equivalent with the minimization of the energy function

$$U(d, \omega) = \sum_{c \in C} V_c(\omega) - \sum_{(i,j)} \ln \left[p(D = d(i,j) | \Theta(i,j) = \omega(i,j)) \right] \tag{12}$$

To minimize $U(d, \omega)$ two different types of iterational algorithms are used, Iterated Conditional Modes (ICM) and Highest Confidence First (HCF), which are both iterative deterministic relaxation techiques. These algorithms are suboptimal, that may converge to local minima, but they induce drastically less computational cost and time than a stochastic relaxation scheme (*i.e.* simulated annealing [10]).

In the ICM algorithm [2], as we used it, an initial estimation of labels is provided by the Maximum Likelihood criterion, that is the minimization of $U_2(d, \omega)$. We also use an Undecision label, and in consequence a threshold on the decision function is used to discriminate the case where a decision seems to be almost sure from the case where a decision is somehow ambiguous. Then, in case of decision, a plausible choice is the label which has maximum conditional probability given the observation of $d(i,j)$ and the current labels in neighborhood of (i,j).

In HCF algorithm [6], the minimization is performed as follows. et each site, a label is selected if it provides the greatest local decrease of the energy function. Computational cost can drastically be

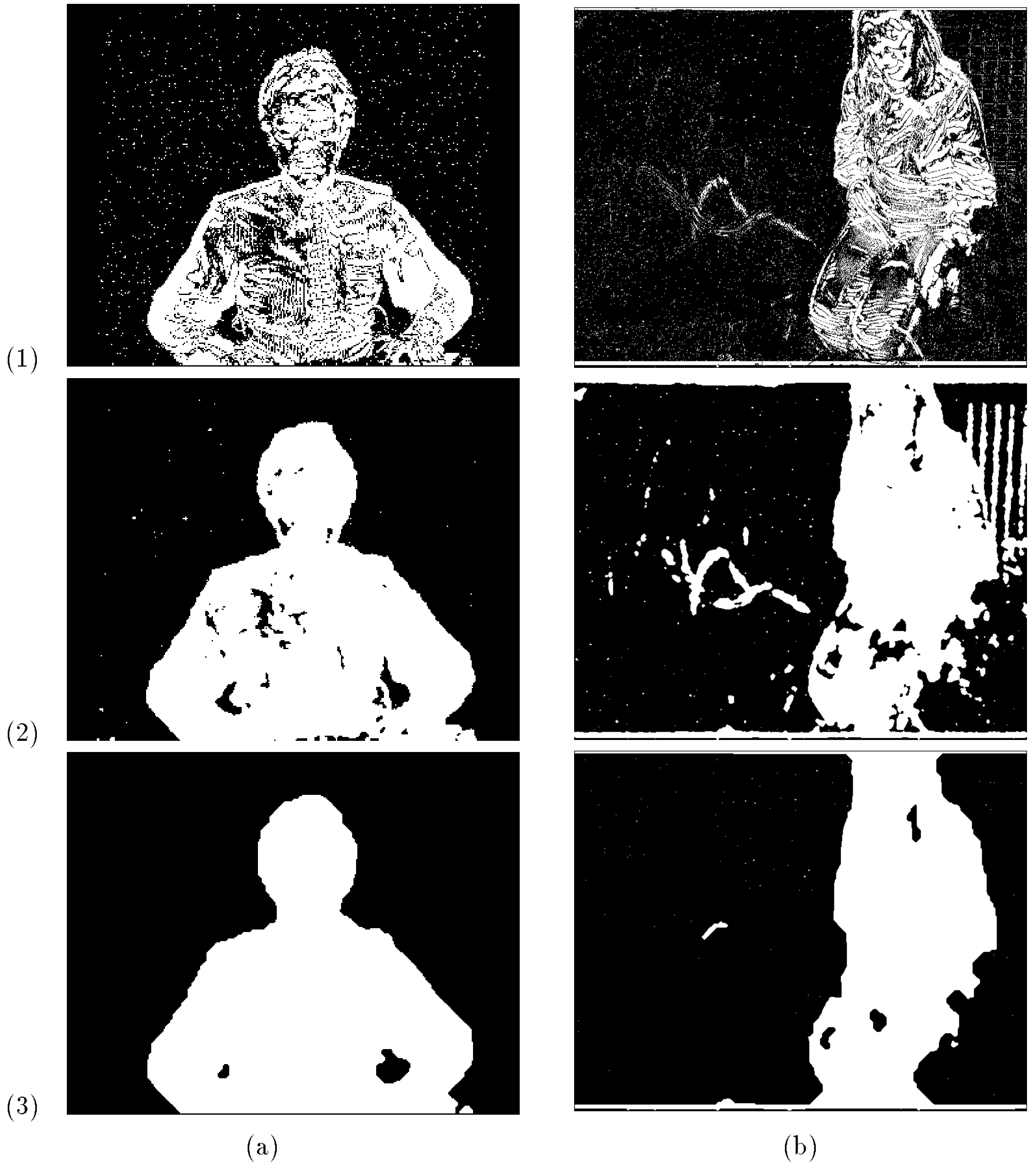


Figure 5: Change detection maps: (1) Maximum Likelihood, (2) ICM, (3) HCF, for (a) *Trevor White* sequence, (b) *Interview* sequence

reduced, if the visit stragegy (for image sites) is optimized. Thus, according to the HCF algorithm the sites are not visited in turn, and we are able to constantly focus on illabeled sites, by introducing an "instability" measure according to which sites are ordered in a stack. Because we are dealing here with only two labels, this "instability" measure can be easily computed. The site to be visited is the one at the top of the stack. On the other hand supplementary computations are required to construct and to maintain the stack. Thus, due to the initialization step, all sites are pushed to the stack according to the energy term $U_2(d, \omega)$ and "instability" measure. Convergence is reached when the stack is empty. The main advandage of HCF is that all computations are very local (apart from the stack updating process) and can be easily parallelized. A significant difference between our implementation and the original HCF, is that we use it two times, in order to enforce $Mobile$ points. Usually in the first pass the potential function $V_c(\omega)$, for both labels $(Static, Mobile)$, has equal values ($\alpha_s = \alpha_m$), something which does not hold in the second pass, where α_s is decreaced while α_m is increased. The initial labeling guess in the second pass, is the labeling result from the first pass.

In Figure 5 are given on the two test sequences results of applying firstly a threshold Maximum Likelihood estimator, and then that of applying the ICM and HCF algorithms, as described above.

3 Moving Object Location

The modelization of moving object location problem is similar with the one we adopted in change detection. The labeling problem in this case is more complicated because the goal is to characterize the situation that holds in both frames, for each pixel in the image grid. Any pixel in any frame either belongs to the background pixel, or it belongs to some moving object. Let $U = \{B, O\}$ be the set of the two possible labels, where B means "background" and O means "object". In the moving object location problem a couple of labels should be estimated $(\Theta(i, j; t), \Theta(i, j; t+1)) \in U \times U$. This notation is equivalent with given label $\Theta(i, j; t)$ (resp. $\Theta(i, j; t+1)$) for the situation that holds on frame at time instant t (resp. $t+1$) at pixel loaction (i, j). We have four possible label events,

$$\begin{aligned} H_{00} &: (\Theta(i, j; t), \Theta(i, j; t+1)) = (B, B) \\ H_{01} &: (\Theta(i, j; t), \Theta(i, j; t+1)) = (B, O) \\ H_{10} &: (\Theta(i, j; t), \Theta(i, j; t+1)) = (O, B) \\ H_{11} &: (\Theta(i, j; t), \Theta(i, j; t+1)) = (O, O) \end{aligned} \tag{13}$$

The available observation set is composed of change detection map, and gray level values for both frames. The first problem we deal with is the computation of conditional density functions. Let

$$p((I(i, j; t), I(i, j; t+1)) = (x_0, x_1) | (\Theta(i, j; t), \Theta(i, j; t+1)) = (\alpha, \beta))$$

be the conditional density function for case (α, β), where $(\alpha, \beta) \in U \times U$ and $I(i, j; t)$ the grey level value at pixel (i, j). In case of $\alpha \neq \beta$ the problem is easier, since the two events are completely independent, thus

$$p(x_0, x_1|\alpha, \beta) = p(x_0|\alpha)p(x_1|\beta) \tag{14}$$

Under the above hypothesis we are not obliged to calculate the two-dimensional density functions for cases $(B, O), (O, B)$ because their values can be extracted by the use of one-dimensional density functions.

3.1 Gaussian mixture decomposition of the probability density function

The Static, as well as the Mobile, part of change detection map may be composed of many different populations according to their gray level values. Under this hypothesis the density function of the gray level value, for both object and background, may be decomposed in a mixture of Gaussians,

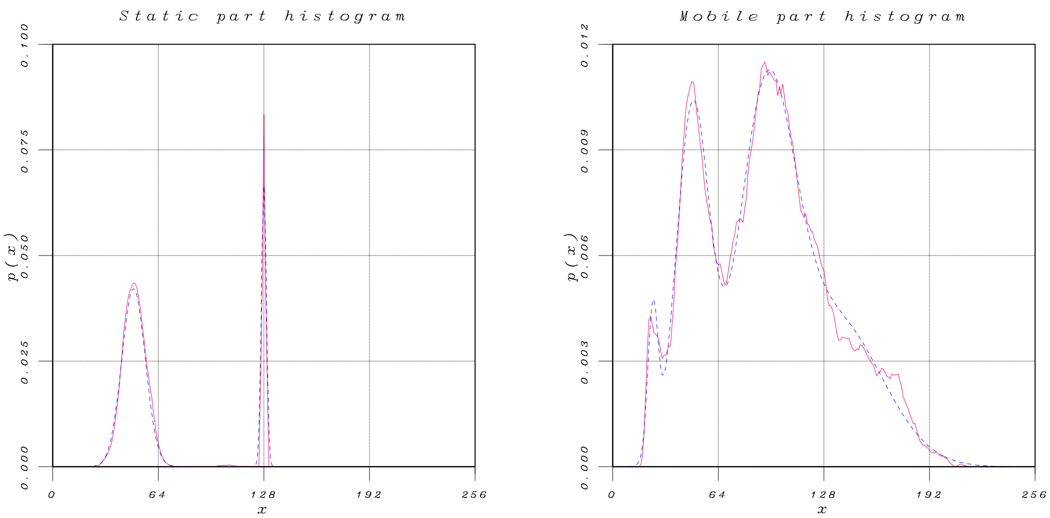
$$p(x|\alpha) = \sum_{i=1}^{c_\alpha} \frac{P_{\alpha i}}{\sigma_{\alpha i}\sqrt{2\pi}} e^{-\frac{(x-\mu_{\alpha i})^2}{2\sigma_{\alpha i}^2}} \tag{15}$$

Using change detection map and pixels labeled as unchanged, we are able to evaluate the histogram for the gray level values of the background. The problem now is to estimate the parameters of the mixture decomposition. An additional problem is that the number of populations, c_α, is unknown. The number of populations is extracted empirically using the observed histogram. To avoid the influence of noise first we perform a smooth operation on the observed histogram and then we are looking for local maxima, according to their probability; that is we are seeking for the modes of this distribution. Then using the ML estimator for mixture decomposition, we can compute the unknown parameters $(P_{Bi}, \sigma_{Bi}, \mu_{Bi})$ for each population. The same approach is used for the estimation of $p(x|O)$. The only difference is that pixels labeled as changed, and presenting an important inter-frame difference, are excluded from the Object as considered to belong to the occluding regions. The evaluation of the histograms for both cases is performed only on the first frame, because we assume the temporal stationarity of the corresponding variables. Results are given in Figure 6.

The problem remains with cases $(B, B), (O, O)$ and two solutions are proposed. The simplest one is the use of a global correlation coefficient ρ_α, for both cases. Then using this coefficient and assuming that it is valid separately for the populations composing the distribution of the gray levels, we can write

$$p(x_0, x_1|\alpha, \alpha) = \sum_{i=1}^{c_\alpha} P_{\alpha i} p_{G2}(x_0, x_1; \mu_{\alpha i}, \sigma_{\alpha i}, \rho_\alpha) \tag{16}$$

where $p_{G2}(x_0, x_1; \mu_{\alpha i}, \sigma_{\alpha i}, \rho_\alpha)$ is a two-dimensional Gaussian probability density with parameters $(\mu_{\alpha i}, \mu_{\alpha i}, \sigma_{\alpha i}, \sigma_{\alpha i}, \rho_\alpha)$.



(a) (b)

Figure 6: Mixture decomposition in Gaussian distributions for *Trevor White* sequence: (a) Static (background) part, (b) Mobile part.

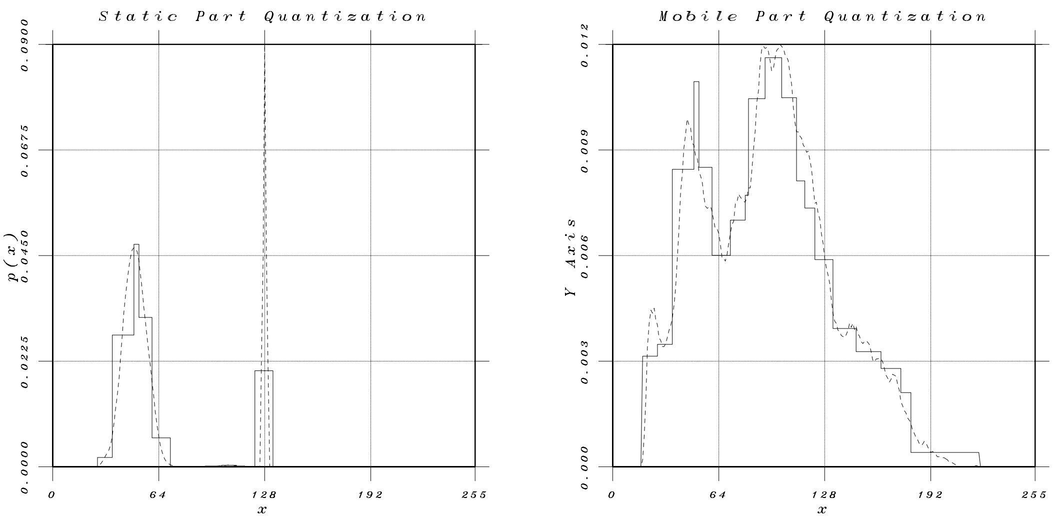
A more robust and reliable approach is the estimation of two-dimensional normal density functions. Using as initial guess all the possible combinations between the observed populations of Background and Object hypotheses and their parameters, and the proposed ML estimator for mixture decomposition, we can compute the unknown parameters of this model. During processing, some classes are rejected because their probability is very small (almost zero). This approach demands a considerable amount of computations, but it has a significant beneficial infuence on the extracted results.

3.2 Piecewise uniform probability density function

An alternative method to determine the values of the energy term U_2 is the quantization of all variables, obtaining thus a piecewise uniform model for the probability density functions. This technique is proposed, in order to avoid the great computational cost of mixture decomposition.

The general idea is to divide the set of possible grey level values in non-overlapping intervals, in such a way that the four probability density functions could use the same orthogonal division of the two-dimensional space of possible values for the couple of intensities on the two frames and for all possible labels of this couple. As the division should be orthogonal for covering the two cases of independent distribution of the two variables, quantization can be simply one-dimensional. The change detection being available, and the necessity to have a good representation of both background and mobile part, independently of their relative size, leads to the construction of two different quantizers, one for each population. The two quantizers are then unified to one having as set of decision levels the union of the two sets of decision levels.

A key problem with quantizers is the determination of the number of decision levels. This problem is solved using the observed histograms and a criterion on the mean squared quantization error. So at the beginning, a number of prevailing values is selected according to the observed histogram, and it composes the set of initial quantization levels. Then the Lloyd-Max algorithm [11] is performed until the convergence is reached. If the global mean square error is above the given threshold, the level with biggest mean square error is subdivided and a new pass of Lloyd-Max algorithm is performed. This operation holds until the global mean square error is above the given threshold.



(a) (b)

Figure 7: Trevor Sequence - Quantization Approach
Quantizations : (a)Static part , (b)Mobile part

Then, according to the final set of decision levels and the observed histograms, the probability for each level for both cases (Static, Mobile) is evaluated. Such a result on *Trevor White* sequence is given in Figure 7. The two-dimensional observed histograms for the couple of pixels with identical labels (both Static or both Mobile) is used on the orthogonally divided set of values to obtain the two-dimensional distribution of the respective couple of variables, again piecewise uniform.

3.3 MAP labeling

Using the same neighborhood definition as it appears in change detection part, we can modelize the problem as a MRF with second order neighborhood, where Gibbs distribution is used to describe the *a posteriori* probability of a global labeling form ω,

$$p(\omega) = \frac{e^{-\frac{U(I_t, I_{t+1}, \omega)}{T}}}{Z} \tag{17}$$

where

$$U(I_t, I_{t+1}, \omega) = U_1(\omega) + U_2(I_t, I_{t+1}, \omega) \tag{18}$$

The definition of U_1, U_2 is similar to those presented in change detection. A more sophisticated definition of potential function is required now

$$V_c(\omega(i,j)) = \xi e_k^\top \begin{bmatrix} -\alpha_s & 1 & 1 & 1 \\ 1 & -\alpha_d & \alpha_{dd} & 1 \\ 1 & \alpha_{dd} & -\alpha_d & 1 \\ 1 & 1 & 1 & -\alpha_s \end{bmatrix} \begin{bmatrix} n_{BB} \\ n_{BO} \\ n_{OB} \\ n_{OO} \end{bmatrix} \tag{19}$$

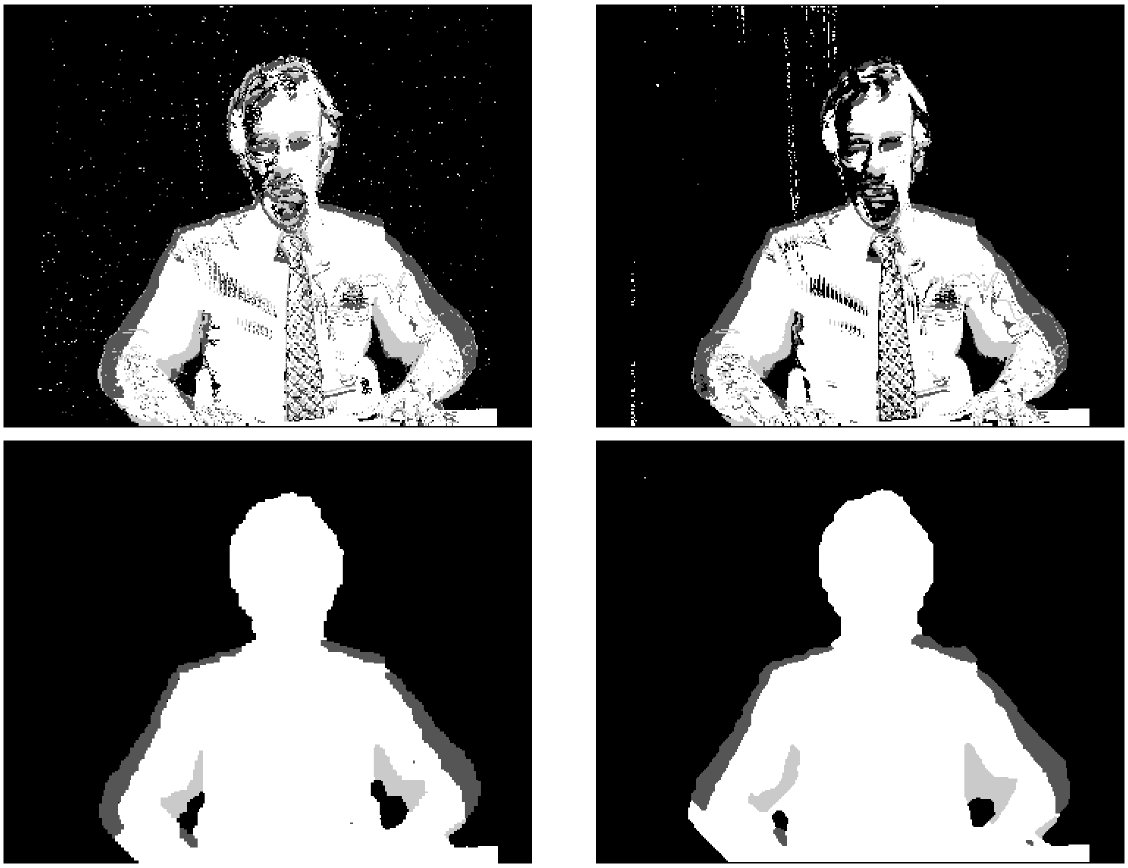
where the following mapping is used $\{(B,B):1, (B,O):2, (O,B):3, (O,O):4\}$, n_{BB} (resp. n_{BO}, n_{OB}, n_{OO}) is the number of pixels with label (B,B) (resp. $(B,O), (O,B), (O,O)$), and e_k is a vector with the k-th element equal to 1 and the others zero. α_s is a potential value that facilitates the selection of (B,B) and (O,O) label, α_d facilitates the selection of (B,O) and (O,B) labels and α_{dd} is the cost to pay to get neighbors with label (B,O) for pixels with label (O,B) (or the opposite), while the cost to pay to get neighbors with different label in any other case is 1.0. The exception value α_{dd} is used because facts (B,O) and (O,B) are mutually exlusive as neighbors. Finally ξ is a weight value.

The solution is derived using MAP criterion and the following energy function has to be minimized,

$$U(I_t, I_{t+1}, \omega) = \sum_{c \in C} V_c(\omega) - \sum_{(i,j)} \ln[p(I_t, I_{t+1} | \Theta(i,j) = \omega(i,j)] \tag{20}$$

HCF and ICM algorithms are proposed for the minimization of $U(I_t, I_{t+1}, \omega)$. An important point in HCF approach is that due to the initialization step, we give label (B,B), at pixels with Static decision on change detection map. This initialization decreases at a significant factor the required computational cost. HCF get more complicated, compared to the change detection case, since we are dealing with four labels, something which has negative effect in its computational time, especially in "instability" measure computations.

In Figure 8 are given the results of the labeling process on the *Trevor White* sequence for the two approaches of evaluation of the probability density functions presented above. The ML decision test result is given for illustrating the efficacy of these estimated probability distributions. In black is the backround, and in gray the covered and uncovered regions. The projection of this result on the two successive frames gives the location of the moving object at the two corresponding moments (Figure 9).



(1) (2)

(a) (b)

Figure 8: Moving object localization for *Trevor White* sequence: (1) ML labeling, (2) HCF labeling, using (a) Mixture decompostion for histogram analysis. (b) Quantization approach for histogram analysis.

(a) (b)

Figure 9: Moving region location

4 Conclusion

In this paper, we described methods and related algorithms for solving two interesting problems arising in motion detection.

Concerning the first problem, that is change detection, the main contribution of this paper is the use of a very efficient mixture decomposition of the distribution of the inter-frame difference. Thus the threshold for the ML decision test is adapted to the data. Introducing then a Gibbs random field model for the labels, we proposed the use of two known, and slightly modified, deterministic relaxation algorithms, ICM and HCF, for solving the resulting minimization problem. The reliable statistical model used enables to obtain good results on real image sequences, even if the camera is moving, in which case its motion is firstly estimated and compensated.

The image segmentation in changed and unchanged regions was then used for a further step in the segmentation process, which searchs for determining covered and uncovered regions as parts of the whole changed region. As a result we obtained the location of the moving object in the two frames. In our knowledge, the method described in this paper is the only method using only two frames for locating moving objects. At the first step of the proposed algorithm, the probability density function of the background and the moving object are evaluated by identifing an adaptive mixture decomposition, or by approximating them, for less computation cost, using a piecewise uniform distribution. Three solutions were proposed for the modelization and identification of the joint probability distribution of the couple of image intensities on the same site in two successive frames. The two first were an extension of the mixture decomposition of the respective one-dimensional distributions, and the other one was evaluated under a piecewise uniform probability distribution assumption. The efficacy of all these probability distributions was checked implementing the corresponding ML decision tests. The final labeling result were obtained using deterministic relaxation algorithms (as ICM or HCF) based on a Gibbs random field model. Very satisfactory results were obtained on a real image sequence for videoconference applications.

Interesting questions for further investigation concern: the multiresolution implementation of the proposed algorithms for speeding up the computation process and the automatic data-dependent determination of the parameters of the Gibbs random field model. Of course the results we obtained could be further exploited for motion estimation, as the occluding boundaries could be considered as known.

Appendix: Mixture density estimation

A common problem in statistical analysis is mixture decomposition. To be more specific, the problem is to decompose observed samples in a known number of populations which could theoretically describe the data. It is assumed that the probability distribution for the observed data, except the values of some parameters, are known. Let c be known number classes, P_j be the *a priori* probability of class number j, and $p(x|\theta_j)$ be the probability density function for the same class, where θ_j is a vector of unknown parameters. The mixture of the c classes gives the following probability density function

$$p(x|\phi) = \sum_{j=1}^{c} P_j p(x|\theta_j), \quad \sum_{j=1}^{c} P_j = 1 \tag{21}$$

where ϕ is a vector made up of $\{\theta_j : j = 1, \ldots, c\}$ and $\{P_j : j = 1, \ldots, c\}$. The problem is to estimate the unknown parameters in ϕ. The maximum likelihood (ML) estimator is given by Duda and Hart [9] and R. Schalkoff [17]. Another method based on fuzzy ISODATA process, is proposed by J. Bedzek and J. Dunn [1]. Here we use the ML estimator, thus we present the general formula and its application in the case of Laplacian densities. The case of Gaussian densities is considered in [16] and [9].

Let us define the *a posteriori* probability of class i given an observation x

$$P_i(x|\phi) = \frac{P_i p(x|\theta_i)}{\sum_{j=1}^{c} P_j p(x|\theta_j)} \tag{22}$$

If $\{x_1, \ldots, x_n, \ldots, x_N\}$ is a data set, the *a priori* probabilities and the parameters of the probability density model must satisfy the following equations:

$$\hat{P}_i = \frac{1}{N} \sum_{n=1}^{N} \hat{P}_i(x_n|\phi) \tag{23}$$

and

$$\sum_{n=1}^{N} \hat{P}_i(x_n|\phi) \nabla_{\theta_i} \log p(x_n|\hat{\theta}_i) = 0 \tag{24}$$

where

$$\hat{P}_i(x_n|\hat{\phi}) = \frac{\hat{P}_i p(x_n|\hat{\theta}_i)}{\sum_{j=1}^{c} \hat{P}_j p(x_n|\hat{\theta}_j)} \tag{25}$$

For the case of a mixture of two Laplacian densities, the following iterative algorithm is obtained concerning parameters λ_0 and λ_1

$$\lambda_i(k) = \frac{N\hat{P}_i(k-1)}{\sum_{n=1}^{N} \hat{P}_i(x_n|\hat{\phi}(k-1))|x_n|} \tag{26}$$

Parameter values are initialized using the moment estimation method.

References

[1] J. Bedzek and J. Dunn, "Optimal fuzzy partitions: a heuristic for estimating the parameters in a mixture of normal distributions", *IEEE Transactions on Computers*, August 1975, pp. 835-839.

[2] J. Besag, "On the statistical analysis of dirty images", *J. R. Statist. Soc. B*, Vol. 48, No. 3, pp. 259-302, 1986.

[3] P. Bouthemy and P. Lalande, "Recovery of moving object masks in an image sequence using spatiotemporal contextual information", *Optical Engineering*, June 1993.

[4] P. Bouthemy and J.M. Odobez, "Detection of multiple moving objects using muliscale MRF with camera motion compensation", *1st IEEE Int. Conf. on Image Processing*, ICIP'94, Austin, Nov. 1994.

[5] C. Cafforio and F. Rocca, "Methods for measuring small displacements of television images", *IEEE Transactions on Information Theory*, Vol. IT-22, No. 5, 1976, pp 573-579.

[6] P. Chou and C. Brown, "The theory and practice of Bayesian image labeling", *International Journal of Computer Vision*, 4, pp. 185-210, 1990.

[7] N. Diehl, "Object-oriented motion estimation and segmentation in image sequences", *Signal Processing: Image Communication*, Vol. 3, No. 1, February 1991, pp. 1901-1904.

[8] J. Driessen, J. Biemond and D.E. Boekeee, "A pel recursive segmentation and estimation algorithm for motion compensated image sequence coding", *Proc Internat. Conf. Acoust. Speech Signal Processing*, 1989, pp. 1901-1904.

[9] R. Duda and P. Hart, *Pattern Classification and Scene Analysis*, New York, Willey-Interscience, 1973.

[10] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images", *IEEE Trans. Pattern Anal. Machine Intell.* , Vol. PAMI-6, No. 6, pp. 721-741, Nov. 1984.

[11] A. Gersho and R. Gray, *Vector quantization and signal processing*, Kluwer Academic Publishers, 1992.

[12] K. Karmann, A. Brandt and R. Gerl, "Moving object segmentation based on adaptive reference images", in:*Signal Processing, Vol. V:Theories and Applications*, 1990, pp. 951-954.

[13] P. Lalande and P. Bouthemy, "A statistical approach to the detection and tracking of moving objects in an inage sequence", in:*Signal Processing, Vol. V:Theories and Applications*, 1990, pp. 947-950.

[14] L. Mori, F. Rocca and S. Tubaro, "Motion compensated interpolation using foreground/background segmentation", *Digital Signal Processing*, 1991, pp. 379-384.

[15] J. Odobez and P. Bouthemy, "Robust multiresolution estimation of parametric motion models in complex image sequences", *Proc. 7^{th} EUSIPCO, Edinburgh, Scotland*, pp. 411-414, Sept. 1994.

[16] K. Pearson, " Contributions to the mathematical theory of evolution", *Phil. Trans. Roy. Soc. London*, vol. 185, pp. 71-100, 1894.

[17] R. Schalkoff, *Pattern Recognition : Statistical, Structural and Neural Approaches*, J. Wiley and sons, 1992.

[18] Z. Sivan and D. Malah, "Change detection and texture analysis for image sequence coding", *Signal Processing: Image Communication*, vol. 6 , pp. 357-376, 1994.

[19] G. Tziritas and C. Labit, *Motion analysis for image sequence coding*, Elsevier, 1994.