Adaptive Detection and Localization of Moving Objects in Image Sequences

N. Paragios and G. Tziritas
Institute of Computer Science - FORTH,
P.O. Box 1385, Heraklion, Greece
and,
Department of Computer Science, University of Crete
P.O. Box 1470, Heraklion, Greece
E-mails: tziritas@csi.forth.gr nparagio@sophia.inria.fr

Abstract

In this paper we address two important problems in motion analysis: the detection of moving objects and their localization. A statistical approach is adopted in order to formulate these problems. For the first, the inter-frame difference is modelized by a mixture of two zero-mean generalized Gaussian distributions, and a Gibbs random field is used for describing the label set. A new method to determine the regularization parameter is proposed, based on a voting technique. This method is also modelized using a statistical framework. The solution of the second problem is based on the observation of only two successive frames. Using the results of change detection an adaptive statistical model for the couple of image intensities is identified. For each problem two different multiscale algorithms are evaluated, and the labeling problem is solved using either ICM (Iterated Conditional Modes) or HCF (Highest Confidence First) algorithms. For illustrating the efficiency of the proposed approach, experimental results are presented using synthetic and real video sequences.

1 Introduction

Detection and localization of moving objects in an image sequence is a crucial issue of moving video [31], as well as for a variety applications of Computer Vision, including object tracking [8], fixation and 2-D/3-D motion estimation. In the case of a static camera, 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. In many real world cases, this hypothesis is not valid because of the existence of ego-motion (*i.e.*, visual motion due to the movement of the camera). This problem can be avoided by computing the camera motion and creating a compensated sequence.

This paper deals with two related problems, change detection and moving object localization. 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) *the uncovered static background*, (2) *the covered background*, and (3) *the 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. This implies that a complementary computation must be performed after temporal change detection, to extract specific information about the exact location of moving objects.

Simple approaches to motion detection consider thresholding techniques pixel by pixel [14], or block-wise difference to improve robustness to noise [33]. More sophisticated modelings have been considered within a statistical framework, where the inter-frame difference is modeled as a mixture of Gaussian or Laplacian distributions [32]. The use of Kalman filtering for certain reference frames in order to adapt to changing image characteristics has been investigated also [23]. The use of first order Markov chains [10] along rows and of two-dimensional causal Markov fields [16] has also been proposed to model the problem for the motion detection problem.

Spatial Markov Random Fields (MRFs), through Gibbs distribution have been widely used for modeling the change detection problem [1], [2], [7], [23], [26] and [29]. These approaches are based on the construction of a global cost function, where interactions (possibly nonlinear) are specified among different image features (*e.g.*, luminance, region labels). Besides, multiscale approaches have been investigated in order to reduce the computational complexity of the deterministic cost minimization algorithms [26] and to get estimates of improved quality. Finally, in the presence of ego-motion, this motion is estimated before the change detection problem is solved [26].

The existing work related to the localization problem is limited. A similar approach to the one adopted in this paper appears in [24], where an adaptive statistical model using Gaussian distributions is used to create a term which relates the observation set (luminances) with the label set. A more complicated solution exists in [9]. In this last approach, three successive images at instants t_1, t_2, t_3 are considered to recover the moving object location at time t_2.

Here, we propose a motion detection method based on a MRF model, where two zero-mean generalized Gaussian distributions are used to model the inter-frame difference. For the localization problem, Gaussian distribution functions are used to model the couple of the intensities at the same site in two successive frames. In each problem, a cost function is constructed based on the above distributions along with a regularization of the label map. The associated Maximum A Posteriori (MAP) estimator is determined by using multiscale techniques, in order to decrease the large computational cost. Two deterministic relaxation algorithms, ICM and HCF, are used for the minimization of the cost function at each level. The proposed approach can be extended to motion detection problems in the case of a mobile camera.

A new *vote* technique to dynamically determine the regularization parameter(s) in the cost function of the motion detection problem, is proposed. The estimation of the detection map and the estimation of the optimal regularization parameter(s) are alternated. The current solution to the one leads to a more robust estimation for the other. Thus, the current detection map is used to provide an update of parameter values, while these values hopefully lead to better detection maps at the next step. The criterion used is also modeled in a statistical framework.

In order to check the efficiency and the robustness of the proposed method, experimental results are presented both on synthetic and real image sequences. Sequences with stationary camera, as well as sequences with moving camera and independent moving objects, are used to test the method. The remainder of this paper is organized as follows. In Section 2 we make a brief introduction to MRFs, which are used to model the examined problems. Also this section contains the multiscale techniques and the deterministic relaxation algorithms. Motion detection problem and the regularization parameter estimation problem are examined in Section 3, while the moving object localization problem appears in Section 4. Finally, Section 5 contains comments concerning the comparison of the different techniques and algorithms used, as well as conclusions of our work.

2 The Label Field Model and its Estimation

2.1 Introduction to MRFs

Many problems in image analysis can be formulated as a scene labeling with contextual information. In such a statistical framework, there are

- a set of sites $S = \{s_1, s_2, ..., s_n\}$;
- a set of possible labels for each site $\Omega_i \subset \Omega = \{l_1, l_2, ..., l_q\}$, $i = s_1, s_2, ..., s_n$;
- a set of observations $D = \{d_1, d_2, ..., d_n\}$, accosiated with S.
- a neighborhood relation, G, over the sites, which defines a graph where the vertices represent sites, and the edges represent the constraints on the label assignment of the neighboring sites (weighted edges).

The problem is to assign a label to each site in such a way that the solution is consistent with the constraints and the available observations set.

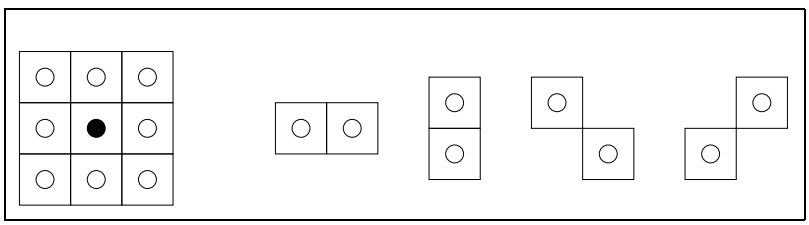


Figure 1: Second order neighborhood: a possible choice of effective cliques

Let c denote a clique of the graph G, and let $C_s = \{c | s \in c\}$ be the set of cliques containing the site s (Figure 1). A global discrete labeling ω assigns one label $\omega_i \in \Omega_i$. An MRF [19], is defined by the so-called clique potentials $V_c(\omega)$, for every possible c and every possible labeling. Following the *Hammersley-Clifford* theorem and the equivalence between MRFs and Gibbs distributions [5], the propability of a labeling ω is given by the following formula [19]

$$P(\omega) \triangleq \frac{1}{Z} e^{-\frac{1}{T}U(\omega)} \tag{1}$$

where T is a regularizing constant, and

$$U(\omega) \triangleq \sum_{c \in C} V_c(\omega), \qquad Z = \sum_{\omega} e^{-\frac{1}{T}U(\omega)} \tag{2}$$

In the above formula C denotes the set of totally connected cliques with respect to the neighborhood definition G and Z is a normalizing constant, called partition function. In statistical terms, U is the *energy (cost)* function of the system, while $V_c()$ is called potential function and corresponds to the contribution of the local interactions to the global energy. A very crucial issue in this process is to incorporate the prior knowledge with the available observations, in order to create a new form for the energy function. This form is a combination between the expected spatial properties (*homogeneity*) of the label field and adequacy between observations and labeling decisions. Under this hypothesis, the energy function is given by

$$U(\omega, d) \triangleq \frac{1}{T} \sum_{c \in C} V_c(\omega) + \sum_{s \in S} \delta(\omega_s, d_s) \triangleq U_1(\omega) + U_2(\omega, d) \tag{3}$$

where the term $\delta()$ expresses this adequacy demand.

A very common technique in such problems is to consider the Maximum A Posteriori (MAP) criterion, *i.e.* the maximization of the *a posteriori* distribution of the labels given the observations, which is equivalent to the minimization of the cost function. This minimization may be performed using either stochastic relaxation algorithms [19], or deterministic relaxation algorithms [11], [6].

2.2 Multiscale Techniques

Defining global *energy (cost) functions* is a powerful tool for specifying non-linear interactions between observed and hidden variables in image restoration problems, where the restoration is provided by the minimization of these functions. The main drawback of this process is usualy the huge solution space (sometimes infinite in terms of computations). Thus, even the simplest restoration schemes demand considerable amount of computations. At the same time, the cost function usually prohibits many local minima, and a very common result is that the final restoration corresponds to one of these minima.

On the other hand, it has been shown that the *multiscale* techniques reduce to a significant ratio the required computational cost of the restoration operation [30] and perform a smooth operation in the observed cost functions, which eliminates a large percentage of local minima. These techniques have been widely used in image analysis problems with a positive influence in the restoration process, as well as in the computational complexity.

The main idea is to solve the restoration problem in many different label spaces, which are subsets of the original one. The label decision corresponds to a set of pixels in the original space. A label process in many different levels is evaluated. Using a coarse to fine pyramid, an extrapolation of the label decisions from levels with low resolution to levels with finer label configurations takes place. This extrapolation scheme is used as initial labeling and a new relaxation process is performed.

Thus, the necessity of real time implementation in our cases leads to multiscale techniques. Two different types of multiscale models are proposed (Figure 2). In the first one, a Gaussian pyramid of images is built upon the full resolution image and similar cost functions to be minimized are defined through the different levels [32]. This *multiresolution* structure is then utilized according to a coarse-to-fine strategy (Figure 2(a)). Another more sophisticated approach consists in defining a consistent *multigrid* label model by using detection maps which are constrained to be piecewise constant over smaller and smaller pixel subsets [21]. The cost function which is considered at each level is then automatically derived from the original finest scale energy function. Also, full observation space is used at each label level and there is no necessity for constructing a multiresolution pyramid of the data (Figure 2(b)).

2.3 Relaxation Algorithms

The modeling of restoration problems by the use of MRFs, leads to a minimization operation process. This process even in the case of multiscale techniques must be performed separately in each level. During the last decade two different types of minimization schemes have been proposed. The stochastic relaxation algorithms [19] demand a considerable amount of computations, but converge under conditions to the optimal solution. On the other hand, it has been shown that the use of deterministic relaxation algorithms ([11] and [6]) could perform a restoration very close to the optimal one, with much less cost in comparison with the stochastic relaxation algorithms.

In this paper the minimization operations are performed by the use of two well known and slightly

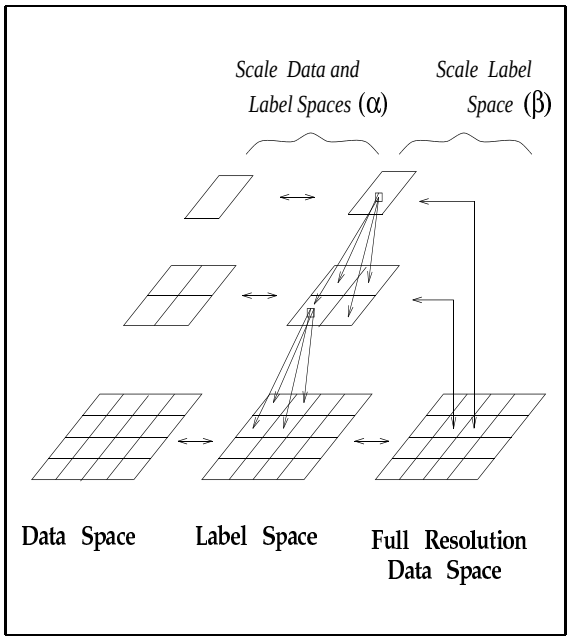


Figure 2: Multiscale Techniques: (a) Multiresolution, (b) Multigrid

modified deterministic relaxation algorithms: the Iterated Conditional Modes (ICM) and Highest Confidence First (HCF), which are both iterative deterministic relaxation techiques. These algorithms are suboptimal, thus they might converge to a local minimum, but they induce drastically less computational cost and time than a stochastic relaxation scheme (*i.e.*, simulated annealing [19]).

In the ICM algorithm [6], as we used it, an initial estimation of labels is provided by the Maximum Likelihood (ML) criterion. Then, the labels are computed iteratively and in parallel for the whole frame. The ICM algorithm is slightly modified here, since 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 certain from the case where a decision is somehow ambiguous. In case of decision, a plausible choice is the label which has maximum conditional probability given the observations and the current labels in the neighborhood of each pixel. In order to avoid redundant computations, the label process is performed only in the pixels where a change in the label of one of their neighbours occurs in the previous iteration. In HCF algorithm [11], the minimization is performed as follows. At each site, a label is selected if it provides the greatest local decrease of the energy function. Computational cost can be drastically 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 (in the case of change detection problem), or four labels (in the case of moving object localization problem), 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 $U_2(\omega, d)$ and the "instability" measure. Convergence is reached, when the stack is empty.

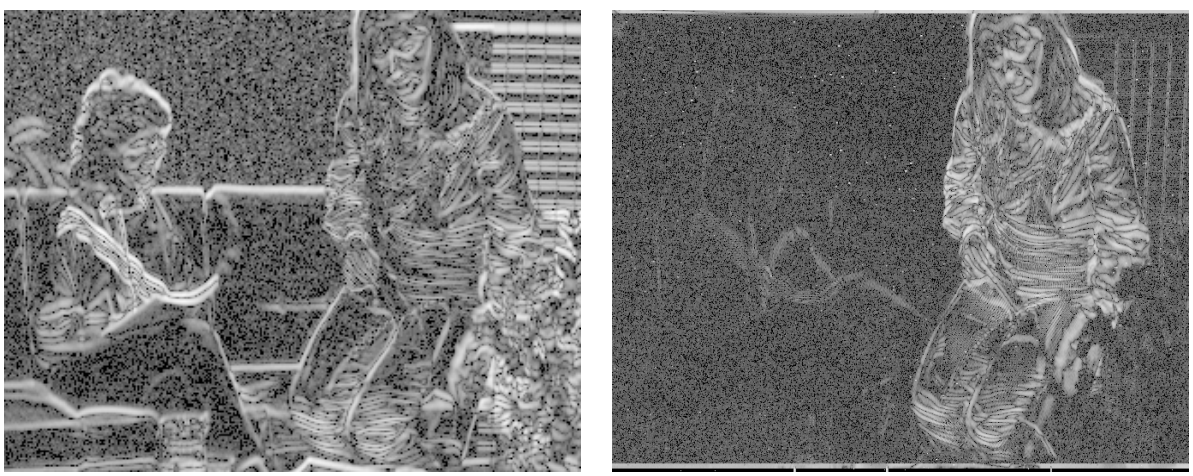
3 Detection of moving objects

3.1 Dominant motion estimation

A very common hypothesis in the change detection problem is the static camera, which holds in a large number of the 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 ([22], [12] and [26]), 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:

$$\begin{pmatrix} u(i,j) \\ v(i,j) \end{pmatrix} = \begin{pmatrix} a_{1,0} + a_{1,1}i + a_{1,2}j \\ a_{2,0} + a_{2,1}i + a_{2,2}j \end{pmatrix} \tag{4}$$

The use of 2-D parametric models to describe the dominant motion is not proper for any case of 3-D motion. Thus the affine model can be used for motions with only translational components, where the overall depth of the scene is much greater than the variation of the depth within the scene, which is equivalent to parallel projection. In addition, this model is proper for rotational motion only about the optical axis (γ rotation) and for 2-D scaling.



(a) (b)

Figure 3: Dominant motion estimation for *Interview* Sequence (a) Interframe Difference, (b) Displaced Interframe Difference (after dominant motion compensation)

The estimation of the set of unknown parameters $\Theta = \{a_{ij}; i = 1,2, \text{ and } j = 0,1,2\}$ between frames at time instant t and $t+1$ is extracted as follows:

$$\hat{\Theta} = \min \left\{ \sum_{(i,j)} \Delta^2 \rho \left(\frac{r(i,j)}{\Delta} \right) \right\} \tag{5}$$

where $r(i,j)$ is the displaced frame difference,

$$r(i,j) = I(i + u(i,j), j + v(i,j); t+1) - I(i,j;t) \tag{6}$$

and $\rho()$ is given by

$$\rho(x) = \begin{cases} x^2, & \text{αν } |x| < 1 \\ 1, & \text{αν } |x| \geq 1 \end{cases} \tag{7}$$

The parameters estimation is performed in an incremental method using a Gaussian pyramid according to Equation (5). Parameter Δ is also an unknown variable and can be estimated iteratively, according to current parameters estimation, by the following formula:

$$\hat{\Delta} = \frac{\sum_{|r(i,j)|<\Delta} |r(i,j)|}{\sum_{|r(i,j)|<\Delta} 1} \tag{8}$$

The minimization is performed using a simple method, Iteratively Reweighted Least Squares (as proposed in [26]), 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, and the percentage of outliers, that is the area of the independently moving object, is relatively small. The resulting motion field is used to compute a compensated image sequence, in which the background then appears as static.

In Figure 3, results of applying the above method to the *Interview* sequence are given, where the motion of the camera is only translational. The inter-frame difference after camera's motion compensation indicates the presence of independent motion.

3.2 Change Detection Model

Let $D = \{d_s, s \in S\}$ denote the gray level difference image with

$$d_s = I_s(t+1) - I_s(t) \tag{9}$$

The change detection problem consists of a "binary" label Θ_s for each pixel on the image grid. We associate the random field Θ_s with two possible events, $\Theta_s = static$ (*static: background pixel*), if the observed difference d_s supports the hypothesis for static pixel (H_0), and $\Theta_s = mobile$ (*mobile: moving pixel*), if the observed difference supports the alternative hypothesis H_1, for mobile pixel. Under these assumptions, for each pixel it can be written

$$\begin{aligned} H_0 &: \Theta_s = static \\ H_1 &: \Theta_s = mobile \end{aligned} \tag{10}$$

Let $p_{D|static}(d|static)$ (resp. $p_{D|mobile}(d|mobile)$) be the probability density function of the observed inter-frame difference under the H_0 (resp. H_1) hypothesis. These probability density functions are assumed to be homogeneous, *i.e.* independent of the pixel location, and usually they are under Laplacian or Gaussian law. We use here a zero-mean generalized Gaussian 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

$$p(d_s|\Theta_s = l) = \frac{c_l}{2\sigma_l\, \Gamma\left(\frac{1}{c_l}\right)} \; e^{-\left(\frac{|d_s|}{\sigma_l}\right)^{c_l}} \tag{11}$$

Let P_{static} (resp. P_{mobile}) be the *a priori* probability of hypothesis H_0. Observed difference values are assumed to be obtained by selecting a label $l \in \{static, mobile\}$ with probability P_l and then selecting a d according to the probability low $p(d|l)$. Thus the probability density function is given by

$$p_D(d) = P_{static}\; p_{D|static}(d|static) + P_{mobile}\; p_{D|mobile}(d|mobile) \tag{12}$$

In this mixture distribution $\{P_l, \sigma_l, c_l; l \in \{static, mobile\}\}$ are unknown parameters. The principle of Maximum Likelihood is used to obtain an estimation of these parameters ([17] and [25]). The unknown

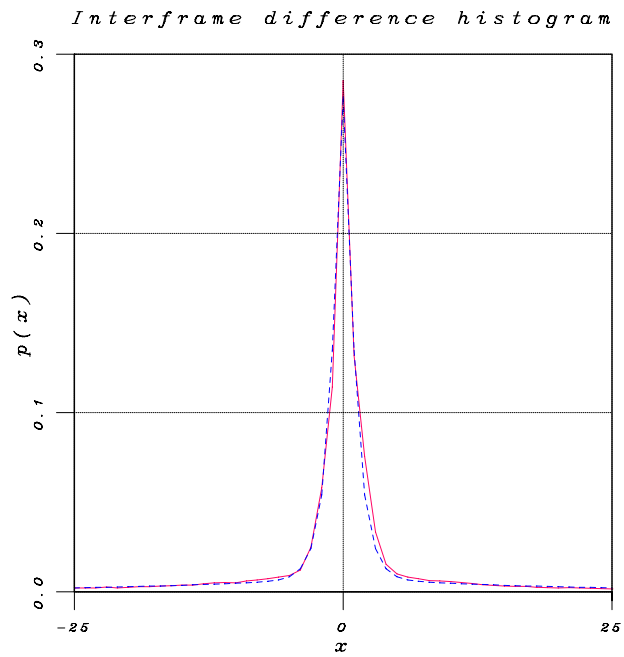


Figure 4: Mixture decomposition for inter-frame difference for *Trevor White*

parameters are iteratively estimated using the observed distribution of gray level inter-frame differences. An initial estimation is calculated using first, second and third order moments of the variable considered. The values of c parameter come from a discrete set: $\{0.5,\ 1.0(:\mathit{Laplace}),\ 1.5,\ 2.0(:\mathit{Gauss})\}$. In Figure 4, the histogram and the approximated probability density function (dashed line) for a test sequence is given, where the Laplacian case is finally selected.

The Static-Mobile decision field, as it appears in our approach, is modeled as a MRF with a 8-pixel neighborhood (Figure 1). In a second step, a MRF model where only two-pixel cliques are considered, in order to reduce the number of necessary parameters and the computational cost, is built to incorporate a *smoothing prior* about the detection map, and a temporal coherence with the final map estimate in the previous frame. A Gibbs posterior distribution $p(\omega|d,\tilde{\omega})$ results with the following energy:

$$U(\omega, d, \tilde{\omega}) \triangleq U_1(\omega) + U_2(\omega, d) + U_3(\omega, \tilde{\omega}) \tag{13}$$

where $\tilde{\omega}$ denotes the detection map estimated at time $t-1$, and:

- $U_1(\omega)$ is the prior term which accounts for the expected spatial properties (homogeneity) of the label field:

$$U_1(\omega) \triangleq \sum_{\{s,u\}\in C} V_{s,u}(\omega_s, \omega_u) \tag{14}$$

  where C is the set of two pixel cliques for the second order neighborhood system, and clique potentials are given by:

$$V_{s,u}(\omega_s, \omega_u) \triangleq \begin{cases} -\alpha_s & \text{if } \omega_s = \omega_u = \mathit{static} \\ -\alpha_m & \text{if } \omega_s = \omega_u = \mathit{mobile} \\ \alpha_{diff} & \text{if } \omega_s \neq \omega_u \end{cases} \tag{15}$$

  $\alpha_{diff} > 0$ is the cost to pay to get neighbors with different labels, while $\alpha_s > 0$ and $\alpha_m > 0$ balances the relative proportions of the two labels.

- $U_2(\omega, d)$ expresses the adequacy between observed temporal variations and current labels according to $p(d_s|\omega_s)$ likelihoods:

$$U_2(\omega, d) \triangleq \sum_{s \in S} \underbrace{-\ln[p(d_s|\omega_s)]}_{\triangleq \delta(\omega_s, d_s)} \tag{16}$$

- Finally $U_3(\omega, \tilde{\omega})$ has a conservative role and expresses a temporal coherence with respect to the labeling at time $t-1$:

$$U_3(\omega, \tilde{\omega}) \triangleq \sum_{s \in S} \rho(\omega_s, \tilde{\omega}_s) \tag{17}$$

where

$$\rho(\omega_s, \tilde{\omega}_s) \triangleq \begin{cases} -\zeta & \text{if } \omega_s = \tilde{\omega}_s \\ 0 & \text{if } \omega_s \neq \tilde{\omega}_s \end{cases} \tag{18}$$

We consider the Maximum A Posteriori (MAP) estimation problem, *i.e.*, the maximization of the *a posteriori* distribution of the labels given the observations, which is equivalent to the minimization of the energy function $U(\omega, d, \tilde{\omega})$,

$$U(\omega, d, \tilde{\omega}) = \sum_{c \in C_s} V_c(\omega) + \sum_{s \in S} (\rho(\omega_s, \tilde{\omega}_s) - \ln[p(d_s|\omega_s)]) \tag{19}$$

3.3 Regularization Parameter Estimation

In this subsection we consider a simpler model, with the smoothing prior U_1 only depending on a single *regularization* parameter λ ($\alpha_s = \alpha_m = \alpha_{diff} = \lambda$), and without temporal coherence term U_3. The aim is to determine the value of the potential λ.

During the last decade, many researchers have investigated the problem of regularization parameter determination, often for problems of image restoration [4]. A common conclusion is that λ may have significant influence on the MAP estimate of the label field [15]. Simple approaches for λ estimation use statistical analysis, where the optimal solution is derived through a pseudo-likelihood criterion [13]. Cross-validation methods have been investigated [28] as well. Finally, an error analysis based on an objective mean square error criterion has also been used to motivate the regularization [18]. In [18], two methods for choosing the regularization parameter are proposed, based on the absence or not of knowledge for the noise model. All the above methods attempt to solve the label field estimation simultaneously with the regularization parameter estimation. Their main drawback is their large computational cost. Another significant drawback is that, in some cases, a prior knowledge of the noise model is required. In this paper, a different method for the regularization parameter estimation is proposed. The general idea is to use the detection map computed for a given parameter value, together with the observations set, in order to extract, with a voting technique, a new λ value which increases the "optimality" of the current map, which, in turn, is re-estimated.

Let $U_l(\omega_s, \omega_{g_s}, d_s)$ be the local energy for label ω_s in the pixel location s, given labels in its neighborhood g_s and the data d_s associated with this location:

$$U_l(\omega_s, \omega_{g_s}, d_s, \lambda) \;=\; \delta(\omega_s, d_s) + \sum_{u \in g_s} V_{s,u}(\omega_s, \omega_u) \;\triangleq\; \delta(\omega_s, d_s) + \lambda \phi(\omega_s, \omega_{g_s}) \tag{20}$$

where

$$V_{s,u}(\omega_s, \omega_u) = \begin{cases} -\lambda, & \text{if } \omega_s = \omega_u \\ +\lambda, & \text{if } \omega_s \neq \omega_u \end{cases} \tag{21}$$

The current label field estimate ω is a sitewise local minimum of the global energy function with the previous value of regularization parameter. We look at λ values for which this still holds, *i.e.*:

$$U_l(\omega_s, \omega_{g_s}, d_s, \lambda) - U_l(\overline{\omega}_s, \omega_{g_s}, d_s, \lambda) \leq 0 \tag{22}$$

where $\overline{\omega}_s$ is the opposite label to ω_s. Let N_s be the number of neighbors of s, and let $n_s(\omega_s)$ be the number of those neighbors with the same label ω_s as s. Using the above notation, the local energy is:

$$U_l(\omega_s, \omega_{g_s}, d_s, \lambda) = \delta(\omega_s, d_s) + \lambda[N_s - 2n_s(\omega_s)] \tag{23}$$

Since $n_s(\overline{\omega}_s) = N_s - n_s(\omega_s)$, constraint (22) becomes:

$$\delta(\omega_s, d_s) - \delta(\overline{w}_s, d_s) + 2\lambda[N_s - 2n_s(\omega_s)] \leq 0 \tag{24}$$

From the above relation we can extract some restrictions about admissible λ. In addition, there are values of λ for which the current map ω is a "better" energy minimum, *i.e.*, the above local energy differences are larger in average than those with the previous parameter value. To determine the new λ, a weighted vote technique is adopted in order to take into account this fact. First, the computational cost of the vote technique is reduced by *quantizing* the parameter search space. Then, according to the above relation at each site, a vote is given to each admissible value of the finite search space. The votes are weighted, according to their contribution in minimizing local energies, *i.e.* in maximizing differences in left-hand side of (24). Also, in order to avoid over-smoothing which too large λ values would favor, a method for balancing the two terms of the energy function is required. For this purpose, the spatial mean value ($\mathsf{E}(.)$) and variance ($\mathsf{var}(.)$) of the energy term $\delta(\omega_s, d_s)$, $s \in S$, are used in the vote weighting. For each $s \in S$, each admissible λ value receives a vote weighted according to:

$$\frac{\delta(\overline{\omega}_s, d_s) - \delta(\omega_s, d_s) - 2\lambda[N_s - 2n_s(\omega_s)]}{\mathsf{var}(\delta(\omega, d)) + \left[\sum_{u \in g_s} V_{s,u}(\omega_s, \omega_u) + \mathsf{E}(\delta(\omega, d))\right]^2} \tag{25}$$

where the mean value and the variance of $\delta(w, d)$ are computed on the image grid, according to the current detection map. The value with larger votes sum, is taken to be the center of the new reduced search space with finer quantization. This hierarchical quantization search procedure allows for quickly obtaining a robust estimate of λ.

This method can be easily extended to problems with a number of states greater than two. It can also be used for regularization models with more than one potential value parameter. In such cases the candidates are vectors. In order to avoid the large computational cost, the quantization in each parameter could be done differently, based on its importance. Thus, for parameters of vital importance, one can use a finer step quantization and a coarser one for parameters of less relevance.

An example of this approach is given in Figure 5, which concerns the motion detection for the *Trevor White* sequence. A detection map arises at the coarsest level using the ML estimator, and the proposed method is applied, and an optimal regularization value arises. This value is used and a new detection map is determined. The method for λ determination is applied again and the results with the current detection map are used as initial values at the next level.

The proposed approach could be modeled and exploited in a global statistical criterion. Let X and Y to be random variables, which denote the difference in the image grid, between the values of energy terms for the observed label and the opposite label, thus

$$\begin{aligned} Y(u_s, u_{g_s}) &= \phi(\overline{u}_s) - \phi(u_s) \\ X(u_s, d_s) &= \delta(\overline{u}_s, d_s) - \delta(u_s, d_s) \end{aligned} \tag{26}$$

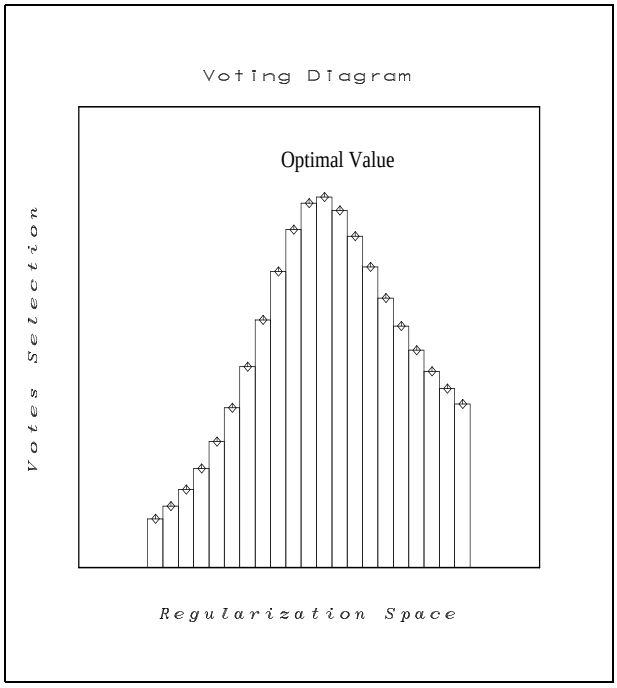


Figure 5: Voting diagram for λ determination

The λ parameter is determined by maximizing the mean distance between the energies for the two opposite labels. At the same time the mean of the energy for the observed label should be minimized, thus minimizing the energy variance and balancing the two terms of the energy. The proposed global statistical criterion is the following

$$f(\lambda) = \frac{(\mathsf{E}\,(X + \lambda Y))^2}{\mathsf{E}\,(\delta + \lambda\phi)^2} \tag{27}$$

If we consider the current detection map, the only unknown variable in this function, which has to be maximized is λ, and $f(\lambda)$ is linear according to this parameter. Thus we search for the λ value which gives the maximum value for $f(\lambda)$. A possible drawback of this approach is that pixels with positive contribution of the cost function (increase) are taken also into account. For the optimal λ the following equation holds

$$\frac{d}{d\lambda} f(\lambda) = 0 \tag{28}$$

and since according to the previous form, a second degree equation results, where only two roots exist (a positive and a negative), it is easy to select the proper one (positive). This method can be easily and directly extended for cases with more than two possible label events.

In Figure 6, results obtained by the use of the vote technique for the change detection problem are presented. Concerning the comparison between the vote approach and the method based on the statistical criterion of the Equation (27), both methods give very satisfactory similar results. The main advantage of the vote approach is that only sites providing confirmation of the decision map are used, while in the global statistical criterion opposite decisions are also taken into consideration. On the other hand, the vote technique requires a very larger amount of computations.

4 Moving Object Localization

The modeling of moving object localization 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

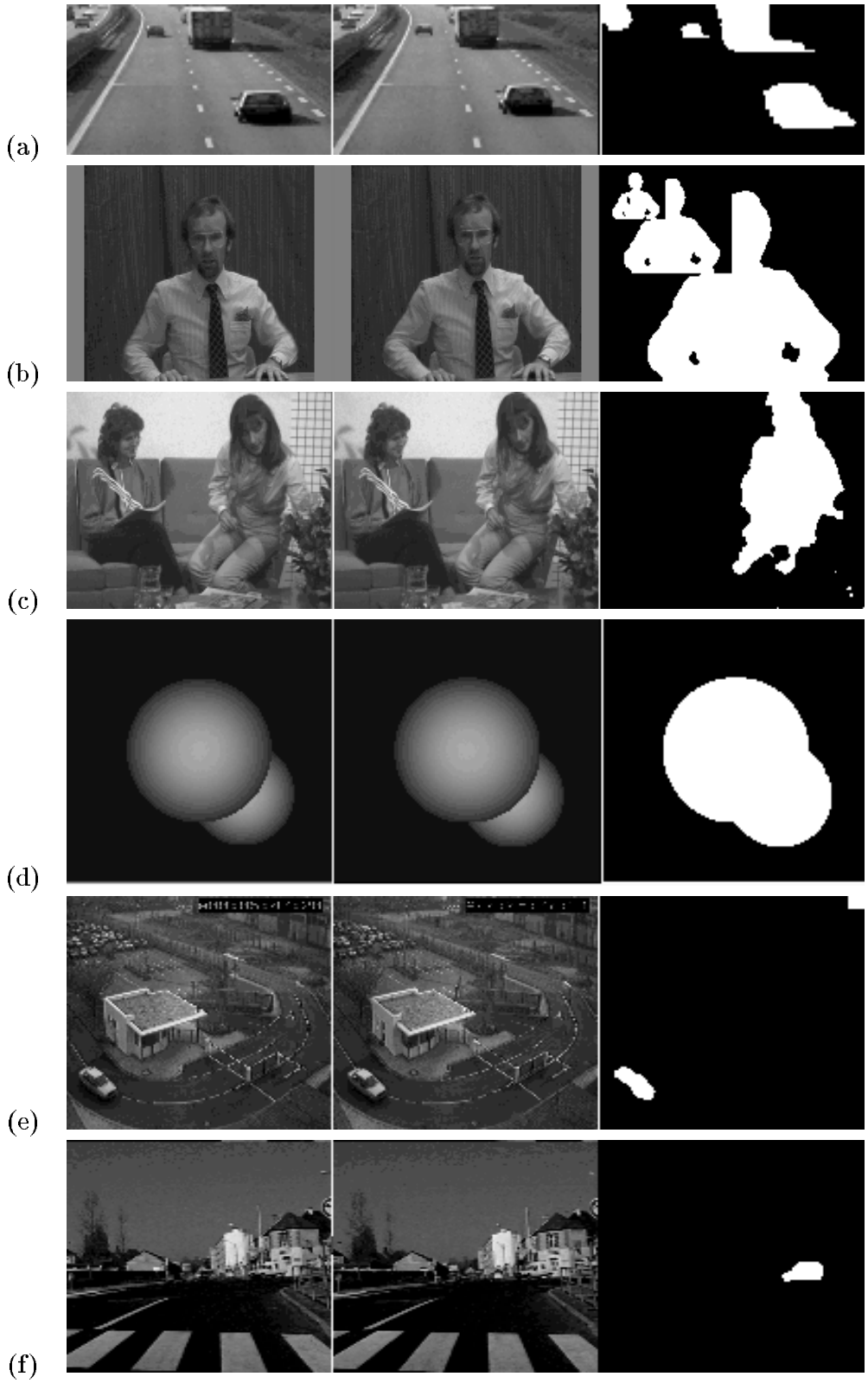


Figure 6: Detection of Moving Objects: *Static Camera*: (a), (b), (d), (e), (f); *Mobile Camera*: (c) **Multigrid** approach: (b) *Trevor White*, ICM; (d) *Sphere*, HCF. **Multiresolution** approach: (a) *Highway*, ICM; (c) *Interview*, HCF; (e) *Kollnig*, HCF; (f) *Van*, ICM. **Automatic λ determination**: a (*1.0125*), c (*0.8125*), e (*0.8625*) and f (*0.9875*).

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 localization problem a couple of labels should be estimated $(\Theta_s(t), \Theta_s(t+1)) \in U \times U$. This notation is equivalent with given label $\Theta_s(t)$ (resp. $\Theta_s(t+1)$) for the situation that holds on frame at time instant t (resp. $t+1$) at pixel location s. We have four possible label events,

$$\begin{aligned} H_{00} &: (\Theta_s(t), \Theta_s(t+1)) = (B, B) \\ H_{01} &: (\Theta_s(t), \Theta_s(t+1)) = (B, O) \\ H_{10} &: (\Theta_s(t), \Theta_s(t+1)) = (O, B) \\ H_{11} &: (\Theta_s(t), \Theta_s(t+1)) = (O, O) \end{aligned} \tag{29}$$

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

$$p(x_0, x_1 | (\Theta_s(t), \Theta_s(t+1)) = (\alpha, \beta))$$

be the conditional *pdf* for the couple of the intensity values $(I_s(t), I_s(t+1))$ at pixel s, where $(\alpha, \beta) \in U \times U$. In case of $\alpha \neq \beta$ the problem is easier, since the two events are independent, thus

$$p(x_0, x_1 | \alpha, \beta) = p(x_0 | \alpha) p(x_1 | \beta) \tag{30}$$

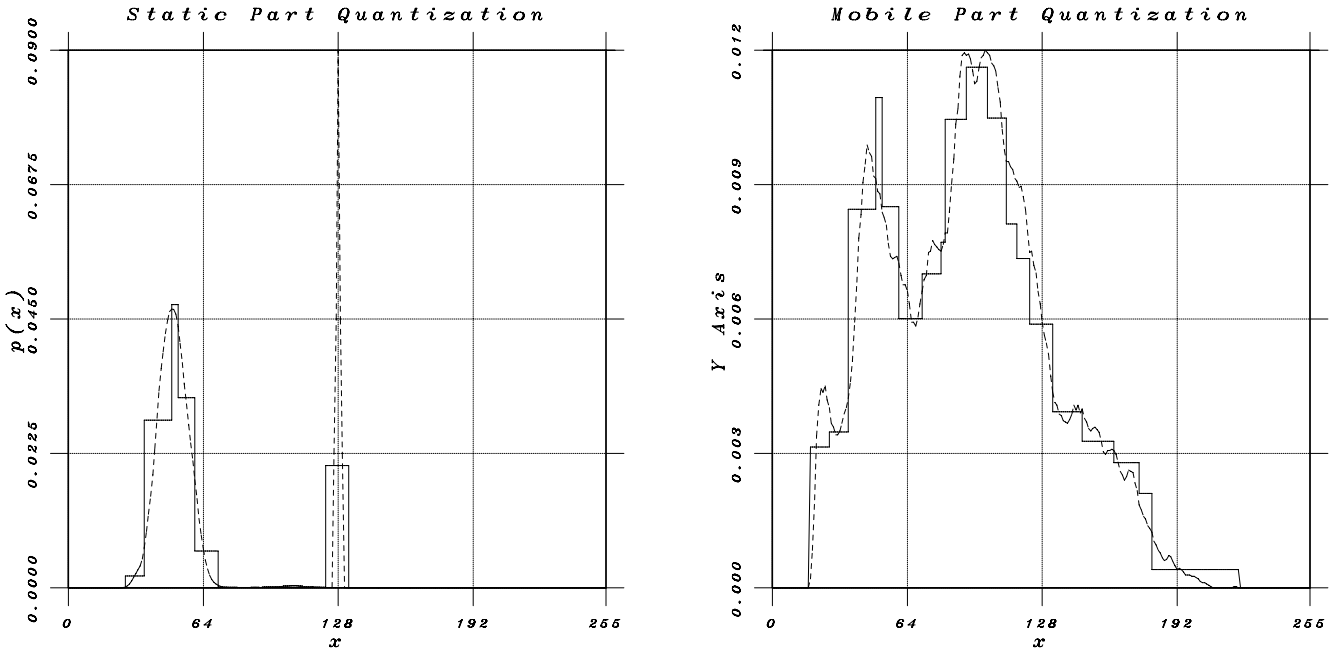
Under the above hypothesis we are not obliged to calculate the two-dimensional probability density functions for cases (B, O) and (O, B), because their values can be extracted by the use of one-dimensional *pdf*s.

Using the *change detection* map, from pixels labeled as unchanged, we are able to evaluate the histogram for the gray level values of the *background*, as well as for the *mobile part*. 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 random variables. These histograms are used for the estimation of the unknown conditional probability density functions. The main drawback of this approach is that the *static part* of the change detection map might have much bigger area than the *mobile part*, which causes problems in the statistical representation of the *covered* and *uncovered areas*, as well as in *mobile area*. Also, in cases with more than one moving object the evaluated histograms are not valid and reliable for each object. To avoid these problems, these histograms are evaluated for each object in a rectangular area around the object, where the two cases, changed and unchanged, have approximately the same area. This process demands a connected component labeling operation in the *change detection* map, before the examination of the localization problem.

According to the observed histograms, the *static*, as well as the *mobile*, part of the change detection map may be composed of many different populations according to their gray level values, and the decomposition of these maps is the key problem.

4.1 Piecewise uniform probability density function

The simplest method to determine the values of the energy term U_2 (Equation (3)) is the *quantization* of all variables, obtaining thus a *piecewise uniform model* for the probability density functions. This technique demands a reasonable amount of computations.

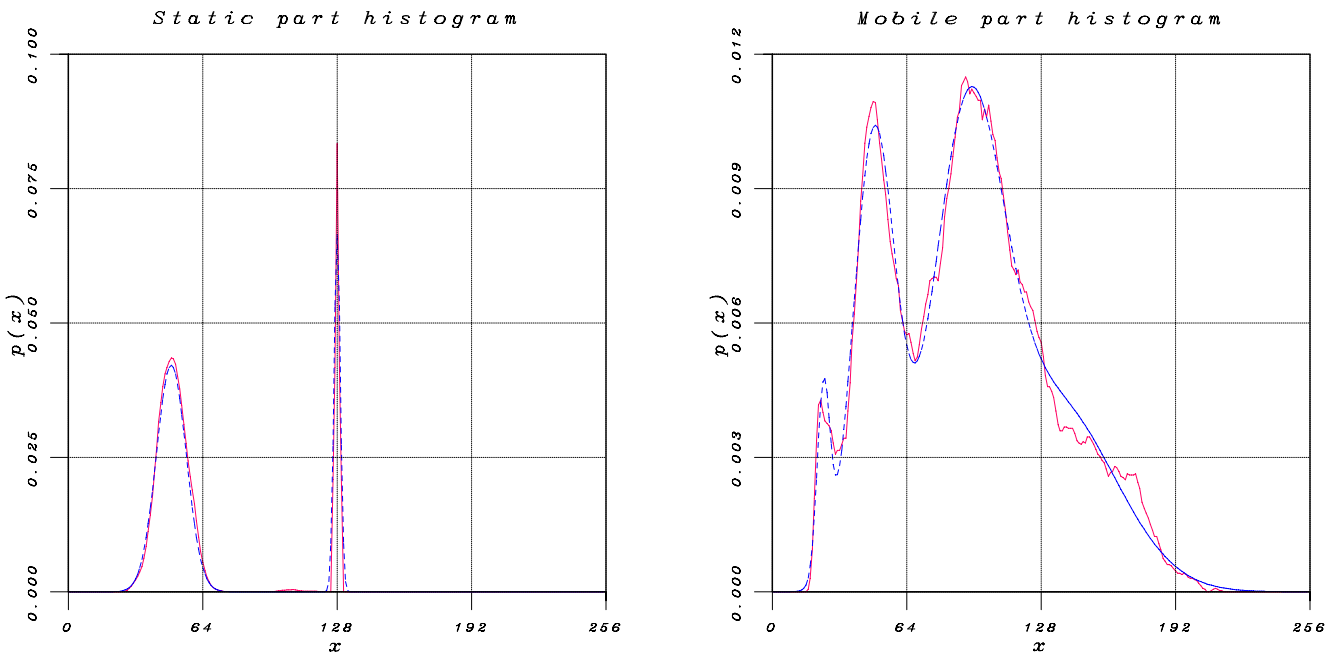


(a) (b)

Figure 7: Approach with piecewise uniform distributions: *Trevor White* Sequence - (a) Static part , (b) Mobile part

The general idea is to divide the set of possible gray 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, which composes the set of initial quantization levels. Then, the Lloyd-Max algorithm [20] 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. 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.



(a) (b)

Figure 8: Approach with mixture of Gaussian distributions: *Trevor White* Sequence - (a) Static part, (b) Mobile part

4.2 Gaussian mixture decomposition of the probability density function

A more complicated model to approximate the observed histograms is the *mixture of Gaussian distributions*. Under this hypothesis the density function of the gray level value, for both object ($\alpha = O$) and background ($\alpha = B$), may be decomposed in a mixture of Gaussians,

$$p(x|\alpha) = \sum_{i=1}^{c_\alpha} \frac{P_{\alpha i}}{\sigma_{\alpha i}\sqrt{2\pi}} e^{-\frac{(x-\mu_{\alpha i})^2}{2\sigma^2_{\alpha i}}} \tag{31}$$

The problem 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, firstly, a smoothing operation is performed on the observed histogram, and then its local maxima are searched; 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 $(P_{\alpha i}, \sigma_{\alpha i}, \mu_{\alpha i})$ for each population. Results of this approach are given in Fig. 8. The problem remains with cases (B, B) and (O, O), for which 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

$$p(x_0, x_1|\alpha, \alpha) = \sum_{i=1}^{c_\alpha} P_{\alpha i} p_{G2}(x_0, x_1; \mu_{\alpha i}, \sigma_{\alpha i}, \rho_\alpha) \tag{32}$$

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 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 could be rejected, because their probability is very small (almost zero). This approach demands a considerable amount of computations, but it has a significant beneficial influence on the extracted results.

4.3 MAP labeling

Using the same neighborhood definition as it appears in the change detection part, we can modelize the problem as a MRF with second order neighborhood, where a Gibbs distribution is used to describe the *a posteriori* probability of a global labeling form ω ($p(\omega) = \frac{1}{Z}e^{-\frac{1}{T}U(I_t, I_{t+1}, \omega)}$), where the cost function is decomposed in two terms,

$$U(\omega, I(t), I(t+1)) = U_1(\omega) + U_2(\omega, I(t), I(t+1)) \tag{33}$$

where the definition of U_1 and U_2 is similar to those presented in the change detection problem, and a more sophisticated definition is required for the potential function, thus

- $U_1(\omega)$ is the term which accounts for the expected homogeneity of the label field:

$$U_1(\omega) \triangleq \sum_{\{s,g_s\} \in C} V_{s,g_s}(\omega_s, \omega_{g_s}) \tag{34}$$

  where C denotes the relations between the pixels and their neighbors according to the given neighborhood system G, and the potential function is given by

$$V_{s,g_s}(\omega_s, \omega_{g_s}) \triangleq \xi e_k^\top \begin{bmatrix} -\alpha_m & 1 & 1 & 1 \\ 1 & -\alpha_d & \alpha_{dd} & 1 \\ 1 & \alpha_{dd} & -\alpha_d & 1 \\ 1 & 1 & 1 & -\alpha_m \end{bmatrix} \begin{bmatrix} n_{BB} \\ n_{BO} \\ n_{OB} \\ n_{OO} \end{bmatrix} \tag{35}$$

  where the following mapping $\{(B,B):1, (B,O):2, (O,B):3, (O,O):4\}$ is used for k, e_k is a vector with the k-th element equal to 1 and the other elements equal to zero, and n_{BB} (resp. n_{BO}, n_{OB}, n_{OO}) is the number of pixels with label (B,B) (resp. $(B,O),(O,B),(O,O)$). α_m 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 for getting neighbors with label (B,O) for pixels with label (O,B) (or the opposite), while the cost for getting 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.

- $U_2(\omega, d)$ expresses the adequacy between observed gray level values and current labels according to $p((I_s(t), I_s(t+1))|\omega_s)$ likelihoods:

$$U_2(\omega, I(t), I(t+1)) \triangleq -\sum_{s \in S} \ln[p((I_s(t), I_s(t+1))|\omega_s] \tag{36}$$

For solving the labeling problem the Maximum A Posteriori (MAP) criterion is considered, which is equivalent to the minimization of the energy function $U(I_t, I_{t+1}, \omega)$,

$$U(\omega, I(t), I(t+1)) = \sum_{c \in C_s} V_c(\omega) - \sum_{s \in S} \ln[p((I_s(t), I_s(t+1))|\omega_s] \tag{37}$$

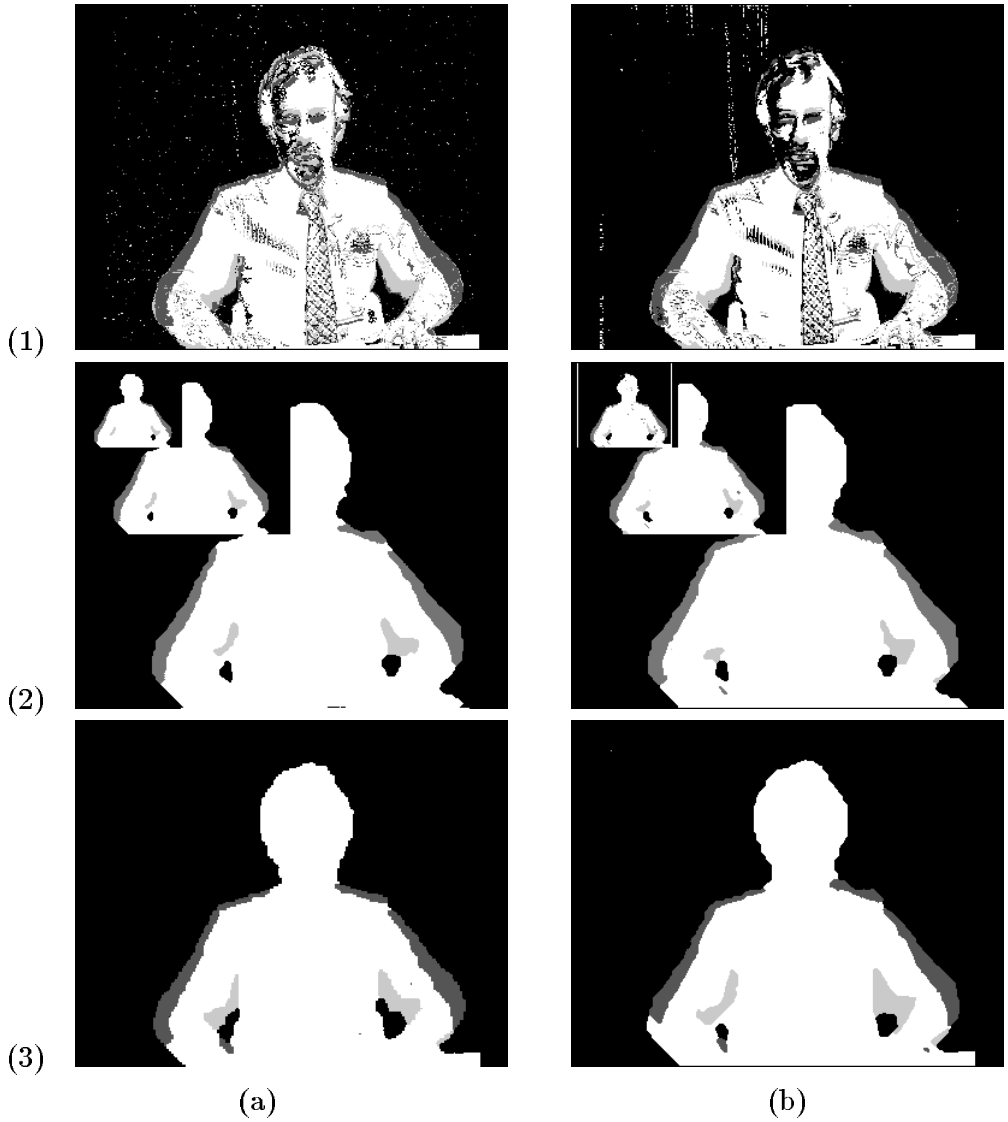


Figure 9: Localization for *Trevor White* sequence: (a) Histograms analysis with mixture of Gaussian distributions – (b) Histograms analysis with mixture of Uniform distributions, (1) Maximum Likelihood detection maps – (2) Detection maps with ICM algorithm and multiscaling in *label* and *data* spaces – (3) Detection maps with HCF algorithm and multiscaling in *label* space

The ICM and HCF algorithms in a multiscale implementation are used for the minimization of the proposed cost function. An important point in this process is that due to the initialization step, the label (B, B) is given at pixels with Static decision on change detection map. This initialization decreases to a significant factor the required computational cost. In Figure 9 the results of the labeling process on the *Trevor White* sequence for the two approaches of evaluation of the probability density functions presented above are illustrated. The ML decision test result is given for illustrating the efficacy of these estimated probability distributions. In black is the background and in gray are the covered and uncovered regions. Also results from the multiscale approaches combined with the minimization algorithms are given. The projection of this result on the two successive frames provides the location of the moving object at the two corresponding moments (Figure 10).

Figure 10: Location for *Trevor White* Sequence

5 Comments and Conclusions

In this paper, we described methods and related algorithms for solving two interesting problems arising in motion detection, the *detection of moving objects and their localization.*

5.1 Comments

5.1.1 The regularization parameter estimation

In order to check the efficiency of the automatic estimation of λ, in the motion detection problem, we first choose $\alpha_s = \alpha_m = \alpha_{diff} = \lambda$, as already mentioned. In spite of this simplification, the adaptive determination of λ allows for obtaining very satisfactory motion detection maps.

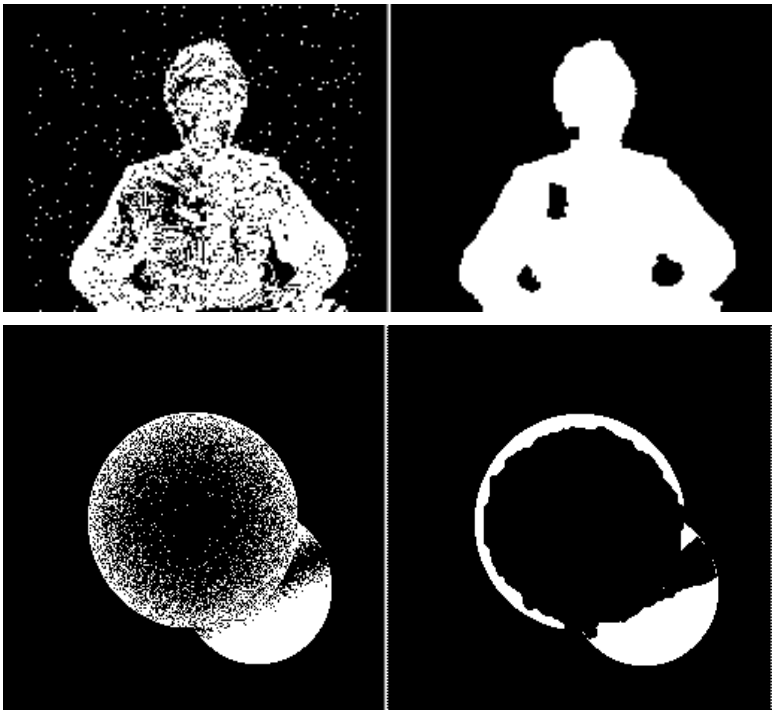


Figure 11: Automatic motion detection: *Trevor White, Sphere*

However, the method seems to fail for *Sphere* sequence (Figure 11). This can be explained by the fact that the initial ML labeling (left image) exhibits a very large and compact static region. Thus, a large λ value arises, which results in removing the isolated mobile labels. The best result for *Sphere* (Figure 6)

is obtained by using the complete model (13), with a reinforcement of mobile labeling through α_m. For *Trevor White* sequence a very satisfactory result is obtained with the simplified model.

5.1.2 ICM versus HCF

According to the experiments, ICM and HCF exhibit different behaviors. Three different aspects are examined: the computational cost, the sensitivity with respect to the regularization parameter, and the dependency on the initial labeling. As for the computational cost, ICM appears less expensive than HCF (Figure 12), due to the use of a sorted "instability stack" by the later. However, in multiscale approaches,

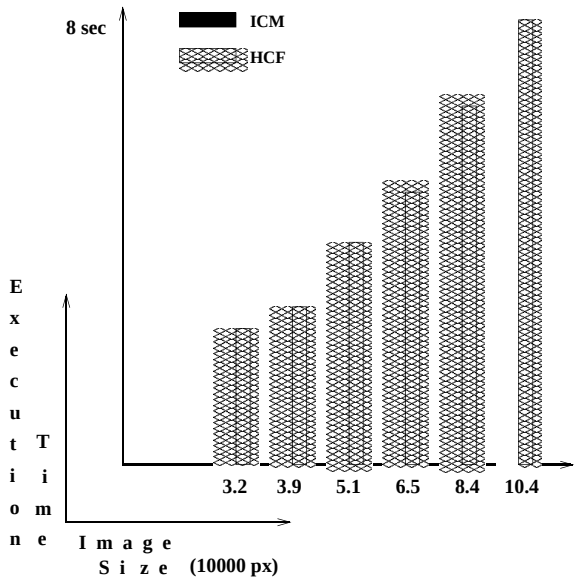


Figure 12: Computational cost for *ICM, HCF*

the cost for HCF is reduced significantly. Indeed, at the coarse levels the required cost for creating and maintaining the HCF stack is very small, and by the time the finer levels are reached, the stack operations become very few. On the contrary, the cost of ICM remains about the same, even with multiscale approach.

Another interesting aspect of the behavior of HCF is the sensitivity with respect to the regularization parameter. It turns out that it is quite high in the single-scale approach, especially around the "optimal" value, where small variations can produce completely different results. This can be explained by the fact that for many sites (especially at the beginning), the labeling decision is taken with an incomplete neighborhood labeling. However, ICM has the opposite behavior: large variations on the regularization parameter do not much influence the estimation (Figure 13). Finally, HCF seems to be more independent on the initial labeling. It produces estimates that can be significantly different from the initial ML labeling. On the other hand ICM has a significant dependency on the initial labeling (Figure 14).

A concluding comment is that, although ICM has less computational cost, it is not flexible and it cannot avoid strong noise influence (as it appears on the initial labeling). For cases with low noise level, however, it can quickly provide a good detection map. HCF is more flexible, thus compensating its significant computational cost. Especially in high level noise cases, it can produce a better result than ICM.

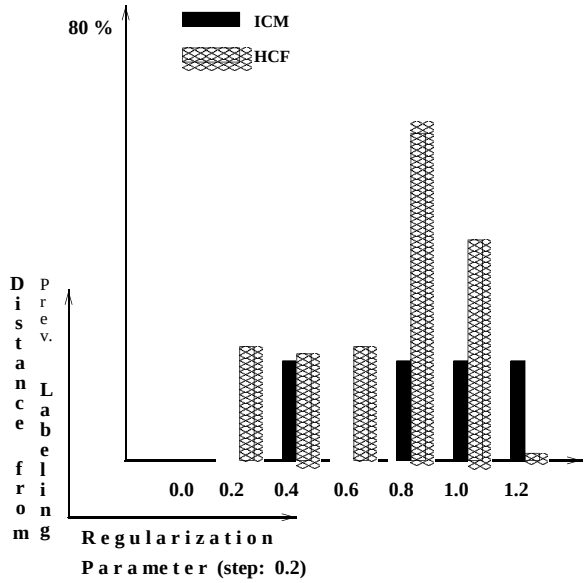


Figure 13: Sensitivity on Regularization for *ICM, HCF*

5.1.3 Multigrid versus Multiresolution

As for the comparison between the two hierarchical approaches, the one using a pyramid of images appears more flexible since parameters can be tuned independently at different resolutions. At the same time, this can be perceived as an increase of the model complexity in terms of parameter estimation. The second approach is by contrast simpler and is proven to be less sensitive to noise influence, since at the coarsest level blockwise data likelihoods are used. Both methods of multiscaling provide about the same computational cost.

5.2 Conclusion

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, for solving the resulting minimization problem, in a single-scale or multiscale approaches. Also a new method for determining the regularization parameter is proposed resulting in a fully adaptive model for the moving object detection 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 searches for determining covered and uncovered regions as parts of the whole changed region. As a result, we obtained the localization of the moving object in the two frames. 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. This operation is done separately for each object in a rectagular area around the object, where the percentage of the populations are approximately equal. Three solutions were proposed for the modeling and identification of the joint probability distribution of the couple of image intensities on the same site in two successive frames. The first two were an extension

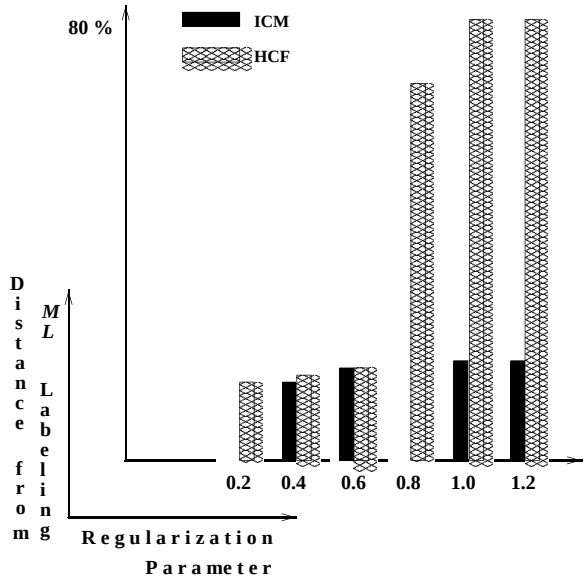


Figure 14: Dependency from *Maximum Likelihood* labeling for *ICM, HCF*

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 results were obtained using deterministic relaxation algorithms (HCF and ICM) based on a Gibbs random field model and multiscale techniques in order to reduce the required computational cost. Very satisfactory results were obtained.

A 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 K be a known number of classes, P_k be the *a priori* probability of class number k, and $p(x|\theta_k)$ be the probability density function for the same class, where θ_k is a vector of unknown parameters. The mixture of the K classes gives the following probability density function

$$p(x|\phi) = \sum_{k=1}^{K} P_k p(x|\theta_k), \qquad \sum_{k=1}^{K} P_k = 1 \tag{38}$$

where ϕ is a vector made up of $\{\theta_k : k = 1, \ldots, K\}$ and $\{P_k : k = 1, \ldots, K\}$. The problem is to estimate the unknown parameters in ϕ. The maximum likelihood (ML) estimator is given by Duda and Hart [17] and R. Schalkoff [27]. Another method based on fuzzy ISODATA process, is proposed by J. Bedzek and J. Dunn [3]. 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 [25] and [17].

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

$$P_i(x|\phi) = \frac{P_i p(x|\theta_i)}{\sum_{k=1}^{K} P_k p(x|\theta_k)} \tag{39}$$

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:

$$\hat{P}_i = \frac{1}{N} \sum_{n=1}^{N} \hat{P}_i(x_n|\phi) \tag{40}$$

and

$$\sum_{n=1}^{N} \hat{P}_i(x_n|\phi) \nabla_{\theta_i} \log p(x_n|\hat{\theta}_i) = 0 \tag{41}$$

where

$$\hat{P}_i(x_n|\hat{\phi}) = \frac{\hat{P}_i p(x_n|\hat{\theta}_i)}{\sum_{k=1}^{K} \hat{P}_k p(x_n|\hat{\theta}_k)} \tag{42}$$

For the case of a mixture of two generalized zero-mean Gaussian densities, the following iterative algorithm is obtained concerning parameters (σ_i, c_i)

- $$\sigma_i(t) = \frac{\sum_{n=1}^{N} \hat{P}_i(x_n|\hat{\phi}(t-1))|x_n|^{c_i(t-1)}}{N\hat{P}_i(t-1)} \tag{43}$$

- $$c_i(t) = \min_c \left\{ \sum_{n=1}^{N} \hat{P}_i(x_n|\hat{\sigma}_i(t), c) \left[\ln c - \ln \Gamma\left(\frac{1}{c}\right) - \left(\frac{|x_n|}{\sigma_i(t)}\right)^c \right] \right\} \tag{44}$$

where c takes values on a predefined finite set. Parameter values σ_i are initialized using the moment estimation method for Laplacian distributions ($c_i = 1.0$).

References

[1] T. Aach and A. Kaup. Bayesian algorithms for adaptive change detection in image sequences using markov random fields. *Signal Processing: Image Communication*, 7:147–160, 1995.

[2] T. Aach, A. Kaup, and R. Mester. Statistical model-based change detection in moving video. *Signal Processing*, 31:165–180, 1993.

[3] J. Bedzek and J. Dunn. Optimal fuzzy partitions: A heuristic for estimating the parameters in a mixture of normal distributions. *IEEE Trans. on Computers*, pages 835–839, August 1972.

[4] M. Bertero, T. Poggio, and V. Torre. Ill-posed problems in early vision. *Proceedings of the IEEE*, 76:869–889, August 1988.

[5] J. Besag. Spatial interaction and the statistical analysis of latice systems (with discussion). *Journal of Royal Statistics Society*, B:36:192–326, 1974.

[6] J. Besag. On the statistical analysis of dirty images. *Journal of Royal Statistics Society*, 48:259–302, 1986.

[7] M. Bischel. Segmenting simply connected moving objects in a static scene. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 16:1138–1142, November 1994.

[8] P. Bouthemy and P. Lalande. Detection and tracking of moving objects based on a statistical regularization method in space and time. *Proceedings in ECCV*, pages 307–311, 1990 (Antibes).

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

[10] C. Cafforio and F. Rocca. Methods for measuring small displacements of television images. *IEEE Trans. on Information Theory*, IT-22:1973–1979, 1976.

[11] P. Chou and C. Brown. The theory and practice of bayesian image labeling. *International Journal of Computer Vision*, 4:185–210, 1990.

[12] P. Davies. Aspects of robust linear regression. *The Annals of Statistics*, 21:1843–1889, 1993.

[13] H. Derin and H. Elliot. Modelling and segmentation of noisy and textured images using gibbs random fields. *IEEE Trans. on Pattern Analysis Machine Inteligence*, pages 39–55, 1987.

[14] N. Diehl. Object-oriented motion estimation and segmentation in image sequences. *IEEE Trans. on Image Processing*, 3:1901–1904, February 1990.

[15] J. Dinten, X. Guyon, and J. Yao. On the choise of the regularization parameter: the case of binary images in the bayesian restoration framework. *Spatial Statistics and Imaging (Editor: A. Posolo)*, pages 55–77, 1991.

[16] J. Driessen, J. Biemond, and D. Boekee. A pel-recursive segmentation algorithm for motion compensated image sequence coding. *Proceedings of ICASSP*, pages 1901–1904, April 1989 (New York).

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

[18] N. Galatsanos and A. Katsaggelos. Methods for choosing the regularization parameter and estimating the noise variance in image restoration and their relation. *IEEE Trans. on Image Processing*, 1:322–336, 1992.

[19] S. Geman and D. Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Trans. on Pattern Analysis Machine Inteligence*, 6:721–741, 1984.

[20] A. Gersho and R. Gray. *Vector quantization and signal processing.* Kluwer Publishers, 1990.

[21] F. Heitz, P. Pérez, and P. Bouthemy. Multiscale minimization of global energy functions in some visual recovery problems. *CVGIP: Image Understanding*, 59:125–134, 1994.

[22] P. Huber. *Robust Statistics.* Wiley, 1981.

[23] K. Karmann, A. Brandt, and R. Gerl. Moving object segmentation based on adaptive reference images. *Signal Processing: Theories and Applications*, V:951–954, 1990.

[24] A. Kuria'nski and M. Nieniewski. Hidden mrf detection of motion of objects with uniform brightness. *Proc. on ICIAP*, pages 657–662, 1995 (San Remo).

[25] G. McLachlan, D. Peel, and W. Whiten. Maximum likelihood clustering via normal mixture model. *Signal Processing: Image Communication*, 8:105–111, 1996.

[26] J-M. Odobez and P. Bouthemy. Robust multiresolution estimation of parametric motion models. *Visual Communication and Image Representation*, pages 348–365, December 1995.

[27] R. Schalkoff. *Pattern recognition: statistical, structural and neural approaches.* J. Wiley and sons, 1992.

[28] B. Silverman. A fast efficient cross-validation method for smoothing parameter choise in spline regression. *Journal of the American Statistical Association*, pages 584–589, 1984.

[29] Z. Sivan and D. Malah. Change detection and texture analysis for image sequence coding. *Signal Processing: Image Communication*, 6:357–376, 1994.

[30] D. Terzopoulos. Image analysis using multigrid relaxation methods. *IEEE Trans. on Pattern Analysis Machine Inteligence*, pages 129–139, 1986.

[31] R. Thoma and M. Bierling. Motion compensating interpolation considering covered and uncovered background. *Signal Processing: Image Communication*, 1:191–212, October 1989.

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

[33] O. Wenstop. Motion detection from image information. *Proceedings in Scandinavian Conf. on Image Analysis*, pages 381–386, 1983.