Detection and location of moving objects using deterministic relaxation algorithms

N. Paragios and G. Tziritas
Institute of Computer Science - FORTH, 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 HCF (Highest Confidence First) algorithm is 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 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. 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 pixel by pixel, or over block-wise intensities to avoid noise influence. Extension of this model, using a mixture decomposition for the observed difference, where the use of a MAP criterion, gives an adaptive determination of the decision threshold.

The use of first order Markov chains ([4]) along rows and two-dimensional order causal Markov field, to modelize the problem has also been proposed for the change detection problem. The statistical modelisation of the problem leads to more sophisticated models, to which the algorithms proposed in this paper belong, where MRFs and Gibbs distributions are used to modelise the problem ([9], [2], [11] and [1]). The detection map arises by minimizing of cost functions using deterministic relaxation algorithms.

The related work with location problem is limited. A similar aproach with the one which we adopt appears in ([8]), while a more complicate solution exists in ([3]). In this approach, three successive images at instants t_1, t_2, t_3 are considered to recover the moving object location at time t_2.

The proposed here change-detection and moving object location algorithms, use for both cases 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 HCF algorithm. The proposed algorithm deals also with the case of a mobile camera and the detection of independent motion. In order to check the efficiency and the robustness of the proposed algorithms experimental results are presented with synthetic and real image sequences.

The remainder of this paper is organized as follows. In Section 2 we deal with the change detection problem, while the moving object location problem appears in Section 3. Finally Section 4 contains concluding remarks and future work.

2. Change Detection

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 ([9]), in order to create a compensated sequence in which only the motion of independent moving objects is still valid. An affine motion-model with six parameters is considered to describe this motion.

Let $D = \{d_s\}$ denote the gray level difference image where $s \in S$ the set of sites in the image grid. 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, $\Theta_s = \alpha$, if s is a static pixel (hypothesis H_0), and $\Theta_s = \kappa$, if s is a mobile pixel (H_1). Under these assumptions, for each pixel it can be written

$$H_0 : \Theta_s = \alpha, \; H_1 : \Theta_s = \kappa \tag{1}$$

Let $p_{D|\alpha}(d|\alpha)$ (resp. $p_{D|\kappa}(d|\kappa)$) 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 we assume that they are under Laplacian law with zero mean. Let P_α (resp. P_κ) be the *a priori* probability of hypothesis H_0. Thus the probability density function is given by

$$p_D(d) = P_\alpha \frac{\lambda_\alpha}{2} e^{-\lambda_\alpha |d|} + P_\kappa \frac{\lambda_\kappa}{2} e^{-\lambda_\kappa |d|} \tag{2}$$

In this mixture distribution $\{P_l, \lambda_l; l \in \{\alpha, \kappa\}\}$ are unknown parameters and principle of Maximum Likelihood is used to obtain an estimation of these parameters ([6]) iteratively. In Figure 1 is given the histogram and the approximated probability density function (dashed line) for *Trevor White* sequence.

The change detection problem can be formulated as a scene labeling with contextual information. In this framework, the problem is to assign a label to each site in such a way, that the solution is consistent with the constraints which arise from a neighborhood relation, G, over the sites.

Let ω be a labeling form of the image a realization of the set of random variables $\Theta = \{\Theta_s, \; s \in S\}$, and ω_s represents the label attached to the site s according to ω. The Static-Mobile decision field as it appears in our aproach is modeled as a MRF with a 8-pixel neighborhood. To

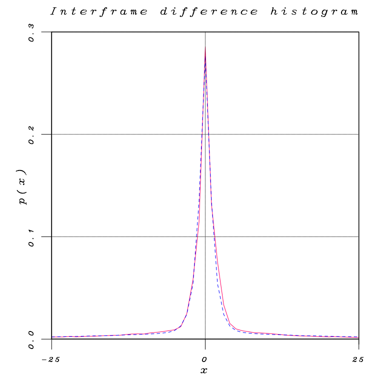


Figure 1. Mixture decomposition for inter-frame difference

describe $p(\omega)$ a Gibbs distribution is used, where only two-pixel cliques are considered. Using the local characteristics of the MRF, and the equivalence with Gibbs distributions, $p(\omega) = \frac{1}{Z} e^{-\frac{1}{T}U(\omega)}$, where the energy function is given by,

$$U(d, \omega) = U_1(\omega) + U_2(d, \omega) \tag{3}$$

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

  $$U_1(\omega) = \sum_{c \in C} V_c(\omega) \tag{4}$$

  where C is the set of all two-pixel cliques in the whole frame, and $V_c(\omega)$ potential function,

  $$V_c(\omega_s, \omega_n) = \begin{cases} -\alpha_s & \text{if } \omega_s = \omega_n = \alpha \\ -\alpha_m & \text{if } \omega_s = \omega_n = \kappa \\ \alpha_d & \text{if } \omega_s \neq \omega_n \end{cases} \tag{5}$$

  n being a neighboor of s (according to G definition). 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 ($0 < \alpha_d < \alpha_s, \alpha_m$).

- Energy U_2 expresses the adequacy between observed temporal differences and corresponding labels according to $p(D = d_s | \Theta_s = \omega_s)$, thus

  $$U_2(d, \omega) = -\sum_s \ln[p(D = d_s | \Theta_s = \omega_s)] \tag{6}$$

The solution of the labeling problem is derived using a 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 $U(d, \omega)$. To minimize $U(d, \omega)$ an iterative deterministic relaxation algorithm is used, Highest Confidence First (HCF)([5]) . This algorithm

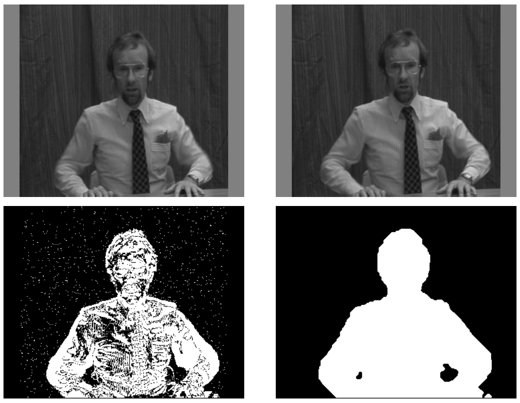


Figure 2. Change Detection for *Trevor White*

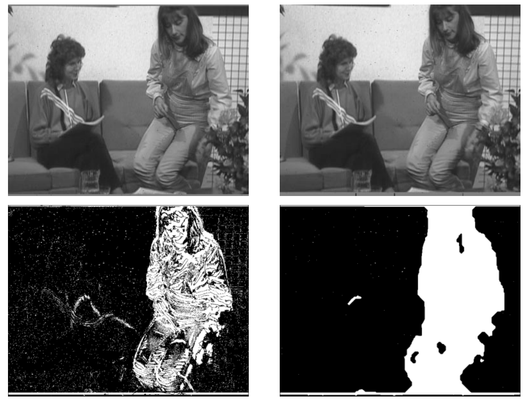


Figure 3. Change Detection for *Interview*

is suboptimal, that may converge to local minima, but it induces drastically less computational cost than a stochastic relaxation scheme (*i.e.* simulated annealing [7]).

In Figures 2 and 3 are given results for *Trevor White* and *Interview* (mobile camera) sequences, including the ML test result (bottom left image).

3. Moving Object Location

The modelization of moving object location problem is similar with the one we adopted in change detection. In this case the goal is to characterize the situation that holds in both frames, for each pixel. Any pixel in any frame either belongs to the background, 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 $(\Theta_s(t), \Theta_s(t+1)) \in U \times U$. 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,

$$\begin{array}{l} H_{00} : (\Theta_s(t), \Theta_s(t+1)) = (B, B) \\ H_{01} : (\Theta_s(t), \Theta_s(t+1)) = (B, O) \\ H_{10} : (\Theta_s(t), \Theta_s(t+1)) = (O, B) \\ H_{11} : (\Theta_s(t), \Theta_s(t+1)) = (O, O) \end{array} \quad (7)$$

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

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

be the conditional density function for case (α, β), where $(\alpha, \beta) \in U \times U$ and $I_s(t)$ the grey level value at pixel s. In case of $\alpha \neq \beta$ the problem is easier (cases $(B, O), (O, B)$), since the two events are completely independent and the densities funcion can be extracted straightly by the use of one-dimensional density function, thus

$$p(x_0, x_1 | \alpha, \beta) = p(x_0 | \alpha) p(x_1 | \beta) \quad (8)$$

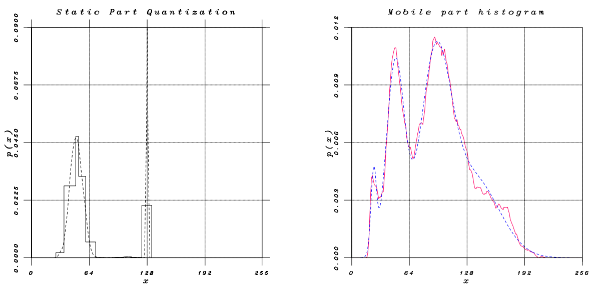
3.1. Gaussian mixture decomposition of the probability density function

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. Thus the density function of the gray level value, for each case may be decomposed in a mixture of Gaussians,

$$p(x|\alpha) = \sum_{i=1}^{c_\alpha} \frac{P_{\alpha i}}{\sigma_{\alpha i}\sqrt{2\pi}} e^{-\frac{(x-\mu_{\alpha i})^2}{2\sigma_{\alpha i}^2}} \quad (9)$$

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. 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 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. Results are given in Figure 4(b).

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, estimated for both cases. Then using this coefficient and assuming that it is valid



(a) (b)

Figure 4. Mixture decomposition for *Trevor White*: (a) Quantization, (b) Gauss Mixting

separately for the populations composing the distribution of the gray levels, we can write

$$p(x_0, x_1|\alpha, \alpha) = \sum_{i=1}^{c_\alpha} P_{\alpha i} p_{G2}(x_0, x_1; \mu_{\alpha i}, \sigma_{\alpha i}, \rho_\alpha) \quad (10)$$

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

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, and the proposed ML estimator, we can compute the unknown parameters of this model. 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. The independent cases, can be estimated as the gaussian case, simply by one-dimensional distributions. 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 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 (Figure 4(a)). The two-dimensional observed histograms for the couple of pixels with identical labels 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 ($p(\omega) = \frac{1}{Z} e^{-\frac{U(\omega)}{T}}$) where the energy function is given by,

$$U(I_t, I_{t+1}, \omega) = U_1(\omega) + U_2(I_t, I_{t+1}, \omega) \quad (11)$$

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

$$V_c(\omega_s) = \xi e_k^\top \begin{bmatrix} -\alpha_s & 1 & 1 & 1 \\ 1 & -\alpha_d & \alpha_{dd} & 1 \\ 1 & \alpha_{dd} & -\alpha_d & 1 \\ 1 & 1 & 1 & -\alpha_s \end{bmatrix} \begin{bmatrix} n_{bb} \\ n_{bo} \\ n_{ob} \\ n_{oo} \end{bmatrix}$$

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. The value α_s facilitates the selection of (B,B) and (O,O), α_d facilitates the selection of (B,O) and (O,B) 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 energy function $U(I_t, I_{t+1}, \omega)$ has to be minimized. The minimization is performed by the use of HCF algorithm. 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.

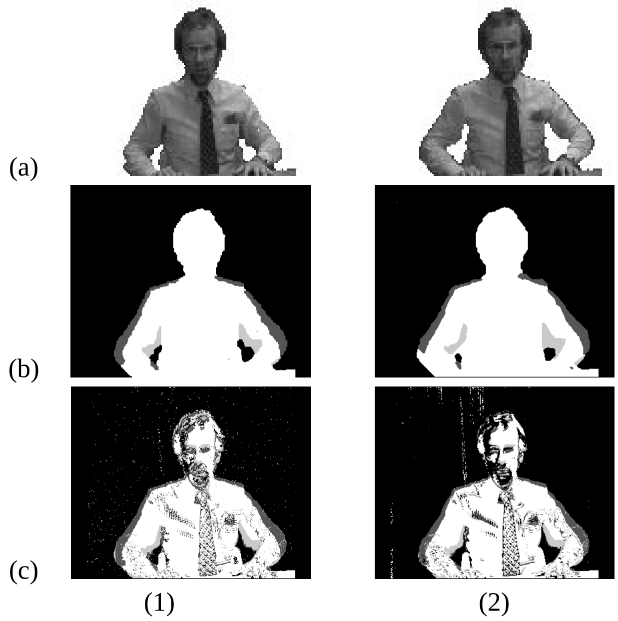


Figure 5. Moving object location for *Trevor White*

In Figure 5 are given the results of the labeling process on the *Trevor White* sequence for both approaches of evaluation of the probability density functions (Gaussian: 1, Quantization:2) presented above (b). The ML decision test result is also given for illustrating the efficacy of these aproaches (c). In black is the background, 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 is also given (a).

4. Conclusion

In this paper, we described methods and related algorithms for solving two interesting problems arising in motion detection. The main contribution of this paper in change detection, 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. The use of a Gibbs random field to model the labels, leads to a reliable statistical model. The solution of the resulting minimization problem is provided by the use of a well-known deterministic relaxation algorithms (HCF). Very satisfactory results were obtained on real image sequences, even if the camera is moving, in which case its motion is firstly estimated and compensated.

The second phase of process 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. 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 efficacy of all these probability distributions was checked implementing the corresponding ML decision tests. The final labeling result were obtained using HCF based on a Gibbs random field model. Very satisfactory results were obtained on a real image sequence for video-conference 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 ([10]). Of course the results we obtained could be further exploited for motion estimation, as the occluding boundaries could be considered as known.

References

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

[2] M. Bischel. Segmenting simply connected movin objects in a static scene. *IEEE Trans. on Pattern Analysis Machine Inteligence*, pages 1138–1142, Nov. 1994.

[3] P. Bouthemy and P. Lalande. Recovery of moving object masks in an image sequence using spatiotemporal contextual information. *Optical Engineering*, pages 1651–1664, June 1993.

[4] C. Cafforio and F. Rocca. Methods for measuring small displacements of television images. *IEEE Trans. on Information Theory*, pages 1973–1979, 1976.

[5] P. Chou and C. Brown. The theory and practise of bayesian image labeling. *Journal of Computer Vision*, pages 185–210, 1990.

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

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

[8] 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).

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

[10] N. Paragios, P. Perez, G. Tziritas, C. Labit, and P. Bouthemy. Adaptive detection of moving objects, using multiscale techniques. *To appear in* 3^{rd} *ICIP*, 1996 (Laussane).

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