

# Motion

# Hidden MRF Detection of Motion of Objects with Uniform Brightness

Adam Kuriański<sup>1</sup> and Mariusz Nieniewski<sup>1,2</sup>

<sup>1</sup> Institute of Fundamental Technological Research, PAS, Warsaw,

<sup>2</sup> Dept. of Fundamental Research in Electrical Engineering, PAS, Warsaw.

e-mail: akurian@ippt.gov.pl

mnieniew@ippt.gov.pl

**Abstract.** The hidden Markov Random Field (MRF) model for motion detection in image sequences is described. A typical MRF model uses two observations: the difference in brightness between two consecutive images, and the value obtained from the mask of temporal changes. The performance of this model can be improved by including the third observation: the brightness at a given pixel. The paper gives the necessary equations and presents an example of motion detection of the object with uniform brightness.

## Notation

$f_k(i, j)$  – brightness value at pixel  $(i, j)$  of the  $k$ -th image,

$o_k(i, j)$  – brightness difference  $o_k(i, j) = f_{k+1}(i, j) - f_k(i, j)$ ,

$\bar{o}_k(i, j)$  – value at pixel  $(i, j)$  of the mask of temporal changes between  $k$ -th and  $(k + 1)$ -th image,

$C = \{1, 0\}$  – set of values which can be assigned to a particular pixel  $(i, j)$  by the mask of temporal changes,

$e_k(i, j)$  – label assigned to pixel  $(i, j)$  of the  $k$ -th image by the mask of moving objects,

$L = \{a, b\}$  – set of labels which can be assigned to a particular pixel by the mask of the moving object,

Superscript  $T$  denotes the transposition of a vector.

## 1 Introduction

The field of labels assigned to the pixels specifies the mask of moving objects and represents a particular realization of a hidden MRF. The energy of this field includes the energy of the hidden layer itself as well as the energy representing the influence of the observation on the hidden layer. In the MRF model considered, there are three sources of observation: brightness difference  $o_k$  at a given pixel  $(i, j)$ , the value obtained from the mask of temporal changes  $\bar{o}_k$  for the same pixel, and brightness  $f_k$  at pixel  $(i, j)$  in the  $k$ -th image. The MRF model with two observation sources is described in [2] and [4], and the authors confirm that their model is not appropriate for motion detection of objects with uniform brightness.

The new model is especially advantageous for such objects. The labels assigned to individual pixels in the mask of moving objects are found by minimization of the total energy  $W_k$  of the field of labels of a pair of images, that is of the  $k$ -th and  $(k + 1)$ -th image. This energy can be expressed as

$$W_k = W_{s_1} + W_{s_2} + W_t + W_c, \quad (1)$$

where:

$W_{s_1}$  – spatial energy of the field of labels for  $k$ -th image,

$W_{s_2}$  – spatial energy of the field of labels for  $(k + 1)$ -th image,

$W_t$  – temporal energy for the field of labels of  $k$ -th and  $(k + 1)$ -th image conditioned on the observation  $\bar{o}_k$ ,

$W_c$  – consistency energy of the field of labels of  $k$ -th and  $(k + 1)$ -th image with observation  $[o_k, f_k]^T$ .

The field of labels assigned to the  $k$ -th image is considered final, whereas the field of labels assigned to the  $(k + 1)$ -th image will be changed subsequently when the next pair of images  $(k + 1)$ -th and  $(k + 2)$ -th is analyzed.

Finding the minimum of the energy  $W_k$  is described in [1] and [3]. In the current paper the problem of building the appropriate model is considered.

## 2 A hidden MRF with Three Observation Sources

Instead of dealing directly with total energy it is much easier to consider the local energy of the field ([1] and [3]). Without getting into details, it can be said that by minimization of the individual local energies one can presumably achieve the minimum of the total energy. The local energy at a pixel  $(i, j)$  is

$$U_{ij} = U_{ijs_1} + U_{ijs_2} + U_{ijt} + U_{ijc}, \quad (2)$$

where:

$U_{ijs_1}$  – local spatial energy of pixel  $(i, j)$  of  $k$ -th image,

$U_{ijs_2}$  – local spatial energy of pixel  $(i, j)$  of  $(k + 1)$ -th image,

$U_{ijt}$  – local temporal energy of the realization  $[e_k, e_{k+1}]^T$ , at pixel  $(i, j)$  of  $k$ -th and  $(k + 1)$ -th image,

$U_{ijc}$  – consistency energy of the realization  $[e_k, e_{k+1}]^T$  with observation  $[o_k, f_k]^T$  at a pixel  $(i, j)$ .

When calculating the energy according to Eq. (2) one has to know the vector  $\Theta$  of the parameters of the model. This vector can be written as

$$\Theta^T = [\Theta_s^T, \Theta_t^T, \Theta_c^T], \quad (3)$$

where  $\Theta_s$  – part of the parameter vector responsible for the spatial energy,  $\Theta_t$  – part responsible for temporal energy, and  $\Theta_c$  – part responsible for the consistency energy.

For purposes of calculating the spatial energy, a simple model of the field is assumed, in which only 2-pixel cliques are included. These cliques can have

horizontal, vertical, diagonal to the left, or diagonal to the right position. The potential  $V_{cs}$  of any 2-pixel clique is assumed to have the form

$$V_{cs} = \begin{cases} \beta_s & \text{if the labels of both pixels are different,} \\ -\beta_s & \text{if the labels of both pixels are the same,} \end{cases} \quad (4)$$

where  $cs$  denotes a spatial 2-pixel clique, and  $\beta_s$  is a positive constant.

Each spatial energy  $U_{ijs_1}, U_{ijs_2}$  is then equal to the sum of potentials of eight cliques, each of which includes the pixel  $(i, j)$ , that is

$$U_{ijs} = \sum_{cs \in C'} V_{cs}, \quad (5)$$

where  $cs$  denotes a spatial 2-pixel clique, and the summation is over the set  $C'$  of all cliques including the pixel  $(i, j)$ . It is possible to assign a different weight to each of the clique orientations. In the simplest case all the weights are the same, and the vector  $\Theta_s$  is

$$\Theta_s = [\beta_s, \beta_s, \beta_s, \beta_s]^T, \quad (6)$$

where  $\beta_s$  is a positive constant.

It is convenient to define the consistency energy  $U_{ijc}$  in terms of the auxiliary vector function  $\Psi$

$$\Psi[e_k(i, j), e_{k+1}(i, j)] = \begin{cases} [m_{01}, m_{02}, \sigma_{01}^2, \sigma_{02}^2, \rho_0]^T & \text{if } [e_k(i, j), e_{k+1}(i, j)] = (b, b) \\ [m_{11}, m_{12}, \sigma_{11}^2, \sigma_{12}^2, \rho_1]^T & \text{if } [e_k(i, j), e_{k+1}(i, j)] = (a, b) \\ [m_{21}, m_{22}, \sigma_{21}^2, \sigma_{22}^2, \rho_2]^T & \text{if } [e_k(i, j), e_{k+1}(i, j)] = (b, a) \\ [m_{31}, m_{32}, \sigma_{31}^2, \sigma_{32}^2, \rho_3]^T & \text{if } [e_k(i, j), e_{k+1}(i, j)] = (a, a) \end{cases} \quad (7)$$

where:

$a$  – label indicating that the pixel belongs to the moving object,

$b$  – label indicating that the pixel belongs to the background,

$(a, a), \dots, (b, b)$  – pair of labels assigned to a pixel in two consecutive images,

$m_{01}$  – mean value of the background-to-background brightness difference,

$m_{02}$  – mean value of the brightness of the background,

$\sigma_{01}^2$  – variance of the background-to-background brightness difference,

$\sigma_{02}^2$  – variance of the brightness of the background,

$\rho_0$  – correlation coefficient between the brightness of the background-to-background difference and the brightness of the object.

The remaining parameters  $m_{11}, \dots, \rho_3$  in Eq. (7) denote, respectively, the mean values, the variances, and the correlation coefficients for three other situations: object-to-background, background-to-object, and object-to-object, which can occur for two consecutive images.

It is assumed that the consistency of the observation  $[o_k, f_k]^T$  at a pixel  $(i, j)$  with the realization of the field of labels  $[e_k, e_{k+1}]^T$  at the same pixel is characterized by the 2-D Gaussian distribution with parameters  $m_{l1}, m_{l2}, \sigma_{l1}^2, \sigma_{l2}^2, \rho_l$ ,

where  $l = 0, \dots, 3$ , depending on which realization  $(a, a), \dots, (b, b)$  is considered. The consistency energy at a pixel  $(i, j)$  is

$$U_{ijc} = \frac{1}{2(1 - \rho_l^2)} \left[ \left( \frac{o_k - m_{l1}}{\sigma_{l1}} \right)^2 - \rho_l \left( \frac{o_k - m_{l1}}{\sigma_{l1}} \right) \left( \frac{f_k - m_{l2}}{\sigma_{l2}} \right) + \left( \frac{f_k - m_{l2}}{\sigma_{l2}} \right)^2 \right] \quad (8)$$

and the vector  $\Theta_c$  is

$$\Theta_c = [m_{01}, m_{02}, \sigma_{01}^2, \sigma_{02}^2, \rho_0, \dots, m_{31}, m_{32}, \sigma_{31}^2, \sigma_{32}^2, \rho_3]^T \quad (9)$$

The temporal energy  $U_{ijt}(e_k, e_{k+1}, \bar{o}_k)$  at a pixel  $(i, j)$  depends on the value at the same pixel of the mask of temporal changes occurring between  $k$ -th and  $(k+1)$ -th image. In the case of motion of objects with constant brightness there is a high probability of the situation when the value of the mask of temporal changes indicates *no changes* but the pixel should belong to the moving object. The brightness observation  $f_k$  allows one to properly describe such a situation. The energy of a temporal clique is found according to the following table

	$\bar{o}_k = 0$	$\bar{o}_k = 1$
$(b, b)$	$-\beta_t$	$\beta_t$
$(a, b)$	$\beta_t$	$-\beta_t$
$(b, a)$	$\beta_t$	$-\beta_t$
$(a, a)$	$-\beta_t$	$-\beta_t$

where  $\bar{o}_k = 1$  indicates that changes at a pixel  $(i, j)$  between  $k$ -th and  $(k+1)$ -th images have been detected,  $\bar{o}_k = 0$  – changes have not been detected, and  $\beta_t$  is a positive constant. In accordance with the above table the vector  $\Theta_t$  is

$$\Theta_t = [-\beta_t, \beta_t, \beta_t, -\beta_t, \beta_t, -\beta_t, -\beta_t, -\beta_t]^T. \quad (10)$$

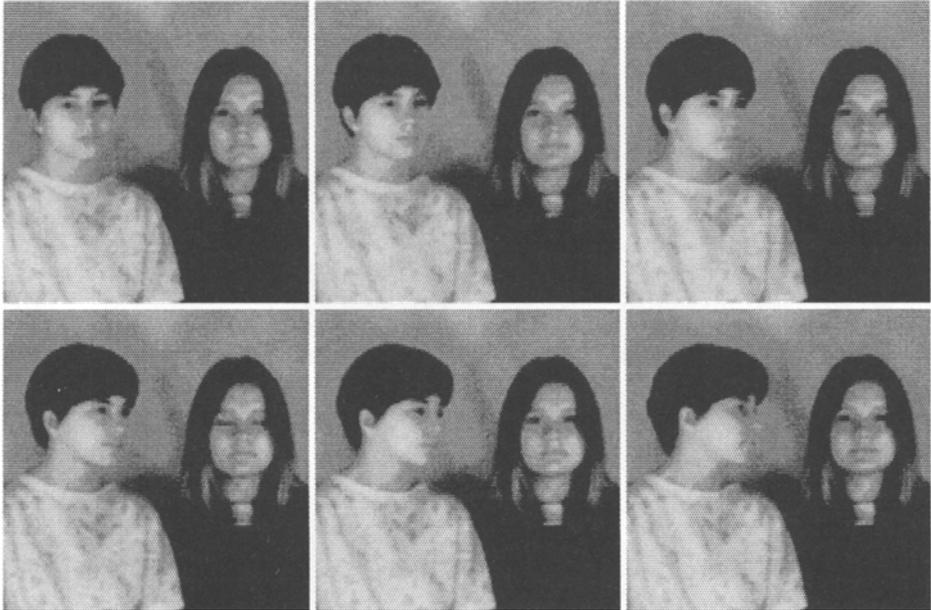
The assumed MRF model has 32 parameters, some of which are equal. Typically one assumes  $\beta_s = 10$  and  $\beta_t = 100$ . However, there are 20 unknown parameters for consistency energy. The authors estimated these parameters via the "teaching sequence." A sample of a sequence was taken, and one obtained the mask of the moving object by manually indicating pixels of the mask. Then, having both the original sequence and the masks of moving objects one estimated the mean values, variances, and correlation coefficients for  $o_k$  and  $f_k$  by means of ML estimation.

### 3 Example of the Use of the MRF with Three Observation Sources

A sequence of images is shown in Fig. 1. The masks of temporal changes for this sequence were obtained by approximating the brightness in the  $3 \times 3$  window by a linear function and then carrying out the ML test for determining whether the remaining variability of the brightness indicated a change. The mask of temporal changes typically consists of a number of disconnected areas. The hair was not included in this mask except for parts of the contour of the head. In Figs. 2 and 3 contours of the masks of moving objects are shown.

## 4 Conclusion

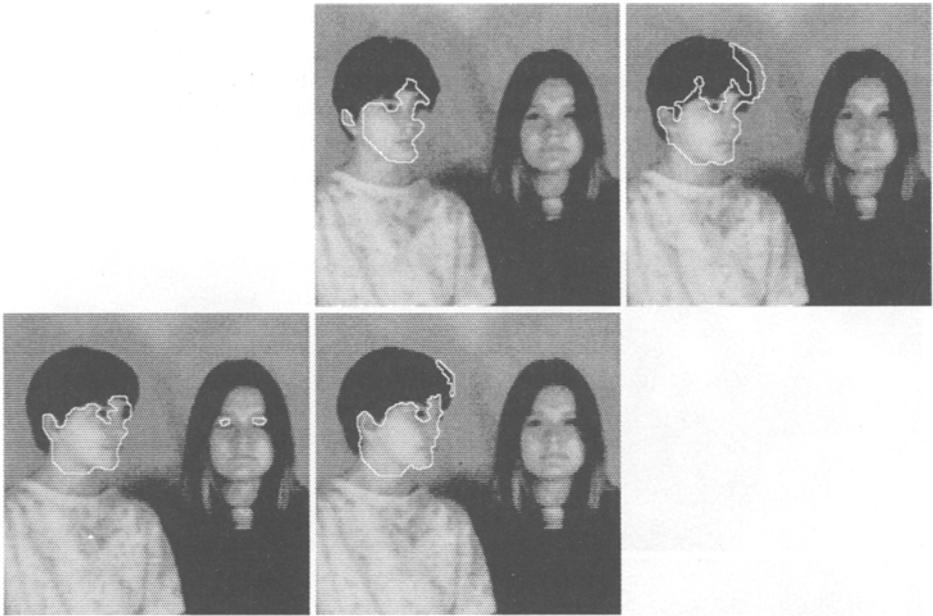
The results obtained confirm that the MRF model using three observations offers a sensitive method of motion detection. Experiments showed that the model with two observations gives very low quality masks for sequences such as in Fig. 1. The model with three observations is particularly recommended for moving objects with uniform brightness.



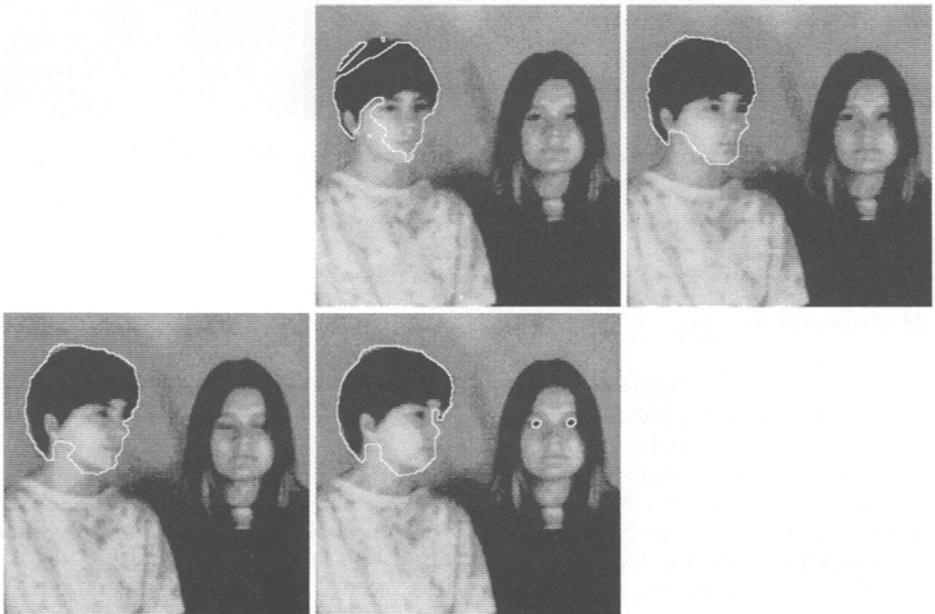
**Fig. 1.** A sequence showing the rotational motion of the girl on the left around her vertical axis.

## References

1. Besag, J.: On the statistical analysis of dirty pictures. *Journ. Royal Statist. Soc.* **48** series B (1986) 259–302
2. Bouthemy, P., Lalande, P.: Detection and tracking of moving objects based on a statistical regularization method in space and time. *Proc. European Conf. Comp. Vision, Antibes* (1990) 307–311
3. Geman, S., Geman, D.: Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *Pattern Anal. Mach. Int.* **6** (1984) 721–741
4. Lalande, P., Bouthemy, P.: A statistical approach to the detection and tracking of moving objects in an image sequence. *Proc. 5th ESPC EUSIPCO 90, Barcelona* (1990)



**Fig. 2.** Contours of the masks of moving objects superimposed on the images of Fig. 1. The model parameters were estimated with the hair included in the background.



**Fig. 3.** Contours of the masks of moving objects superimposed on the images of Fig. 1. The model parameters were estimated with the hair included in the moving object.