HAL
archives-ouvertes.fr

PATCH-BASED MARKOVMODELS FOR CHANGE DETECTION IN IMAGE SEQUENCE ANALYSIS

Thierry Pécot, Charles Kervrann

▶ **To cite this version:**

Thierry Pécot, Charles Kervrann. PATCH-BASED MARKOVMODELS FOR CHANGE DETECTION IN IMAGE SEQUENCE ANALYSIS. LNLA - International Workshop on Local and Non-Local Approximation in Image Processing - 2008, Aug 2008, Lausanne, Switzerland. 2008. <hal-00919698>

HAL Id: hal-00919698
https://hal.inria.fr/hal-00919698

Submitted on 17 Dec 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

PATCH-BASED MARKOV MODELS FOR CHANGE DETECTION IN IMAGE SEQUENCE ANALYSIS

Thierry Pécot and Charles Kervrann

INRIA Rennes - Bretagne Atlantique, Campus Universitaire de Beaulieu, F-35042 Rennes, FRANCE
INRA, UR341 Mathématiques et informatique appliquées, F-78352 Jouy-en-Josas, FRANCE
thpecot@irisa.fr, ckervran@irisa.fr

ABSTRACT

Change detection between two images is challenging and needed in a wide variety of imaging applications. Several approaches have been yet developed, especially methods based on difference image. In this paper, we propose an original patch-based Markov modeling framework to detect spatial irregularities in the difference image with low false alarm rates. Experimental results show that the proposed approach performs well for change detection, especially for images with low signal-to-noise ratios.

1. INTRODUCTION

Change detection represents an important tool in video analysis and is used in a large number of applications [1, 2, 3]. It consists in identifying the missing regions in one image corresponding to appearance or disappearance of objects, motion of objects or shape changes of objects.

Several methods have been developed for change detection. Traditionally, image difference followed by thresholding is the simplest and the most popular approach for change detection. When considering this approach, the most important step is to define the detection threshold. It can be chosen empirically as almost in specific applications [2]. Nevertheless, the commonly-used automatic thresholding methods can be classified into gray-level distribution based [4] and spatial properties based [5]. A review of image difference followed by thresholding based methods is proposed in [5, 6, 7]. More sophisticated approaches have been also developed; i) predictive models [8] exploit the relationships between nearby pixels both in space and time (when an image sequence is available); ii) methods [9] are based on the fact that the decision rule is casted into a statistical hypothesis test. A recent review of change detection algorithms can be found in [7].

In our approach, we want to exploit also the difference image. However, the thresholded difference image detect mainly the object outlines. Indeed, if the object does not move enough, pixels representing the object inside in the first image can still represent the object inside in the second image. Moreover, the difference image is also known to be sensitive to low signal-to-noise ratios, and undesirable artifacts are then detected. Therefore, we propose an original patch-based Markov modeling that detects irregularities and regularizes the difference image. This method is able to robustly detect moving objects between two images. The remainder of the paper is organized as follows: in section 2, we describe the patch-based Markov model; in section 3, we propose a statistical method to automatically threshold the "potential" map. In section 4, we demonstrate the performance of the method on real image sequences.

2. PATCH-BASED MARKOV MODELS FOR IMAGE REPRESENTATION

The difference image shows high values in regions where objects move between the two images. However, the difference image is not robust to low signal-to-noise ratios and shows low values in the inside moving objects. New statistical models are then required to improve motion detection for image sequence analysis.

In our approach, we apply a Markov random fields (MRF) framework [10] to analyse the difference image. In contrast to the usual pixel-wise MRF methods, a recent line of work consists in modeling non-local interactions from image patches; in [11, 12, 13], the redundancy property and patch-based representation can be exploited to detect unusual spatial patterns seen in the scene. In our study, we propose an original patch-based Gibbs/MRF modeling to regularize and detect irregularity in the difference image. In the presence of noise, patch-wise Markov models produce potential maps which are much more regular than those obtained with difference image.

More formally, consider a gray-scale image $v_t = (v(x)_t)_{x\in\Omega}$ defined over a bounded domain $\Omega \subset \mathcal{R}^2$ at time t. The difference image $u = (u(x))_{x\in\Omega}$ is defined as:

$$u(x) = v(x)_{t+1} - v(x)_t. \tag{1}$$

Then, we consider the following difference image:

$$u(x) = u_0(x) + \varepsilon(x), \tag{2}$$

where u_0 is the true difference image and the errors $\varepsilon(x)$ are assumed to be iid (independent identically distributed) Gaussian zero-mean random variables with unknown variance σ_u^2.

In order to robustly detect redundancy in u, we focus on difference image patches as non-local image features able to capture spatial regularities. Our idea is to

perform pairwise comparisons of n-dimensional patches $\underline{u}(x)$ within a fixed-size semi-local neighborhood. For the sake of simplicity, a vectorized image patch $\underline{u}(x)$ is defined as the $\sqrt{n} \times \sqrt{n}$ square neighborhood of point x and the pixels are ordered lexicographically. In [11], it has been confirmed that the l_2 distance $\|\underline{u}(x) - \underline{u}(y)\|^2$ is able to express the amount of similarity between image patches $\underline{u}(x)$ and $\underline{u}(y)$. Intuitively, if the distance is large enough, we can conclude that the patches centered at pixels x and y belong to different spatial contexts. Besides, in homogeneous regions, the noise being assumed to be Gaussian, the score

$$z(x,y) \triangleq \frac{\|\underline{u}(x) - \underline{u}(y)\|^2}{2\sigma_u^2} \tag{3}$$

follows a central chi-squared distribution with n degrees of freedom, i.e. $z(x,y) \sim \chi_n^2$. Hence, the probability distribution function (pdf) can be also expressed as

$$p(z(x,y)) \propto \exp\left[\left(\frac{n}{2} - 1\right)\log(z(x,y)) - \frac{z(x,y)}{2}\right]. \tag{4}$$

The key idea here is to combine the Markov random fields (MRF) framework with patch-based representation. Instead of defining the underlying potential function by hand or training [12], we exploit patch-based score statistics given in (4). Let $G = (V, E)$ be a graph where V denotes the nodes and E the edges connecting the nodes. Moreover, a neighborhood system connecting all the nodes in the square window (larger than patches) and centered at pixel x is defined. The Hammersley-Clifford theorem establishes that the pdf of the proposed graphical model is a Gibbs distribution of the form:

$$p(u) \propto \exp - \sum_{<x,y>} \phi(x,y),$$

where $< x, y >$ denotes the set of cliques in the neighborhood and $\phi(x,y)$ is the homogeneous local interaction potential function. We arbitrarily choose $\phi(x,y) = \log(p(z(x,y)))$ and write the pdf of u as

$$p(u) \propto \exp - \sum_{<x,y>} \log(p(z(x,y)), \tag{5}$$

$$\propto \exp - \left[\sum_{<x,y>} (n-2)\log\left(\|\underline{u}(x) - \underline{u}(y)\|^2\right) - \frac{\|\underline{u}(x) - \underline{u}(y)\|^2}{4\sigma_u^2}\right]. \tag{6}$$

In what follows, all the pairwise comparisons between neighboring patches wrt vertical and horizontal directions will be considered. This Gibbs model is parametrized by only one parameter, i.e. the noise variance σ_u^2, and will be directly estimated by Maximum Likelihood (ML) from data. As explained in [12], the normalization term involved in the Gibbs model (6) is intractable but is not required for our purpose, as we shall see later. The proposed Gibbs distribution also includes spatial correlation in the modeling since neighboring patches overlap in the neighborhood.

For illustration, the difference image (Fig. 1c)) shows high values in areas corresponding to the jumping woman, more precisely in the woman outlines. In Fig. 1d-e), the logarithm of the local conditional probability:

$$\Phi(x) = -\frac{1}{\mathcal{Z}(x)} \log p(\underline{u}(x) | \underline{u}(y), y \in \Delta(x)) \tag{7}$$

is computed at pixel x from local interactions specified for both the pixel-wise and patch-wise MRFs, with $\Delta(x)$ the neighborhood centered at pixel x. The normalization term for this potential is defined as:

$$\mathcal{Z}(x) = \sum_{\underline{u}(x) \in \mathcal{P}} \exp - \left[\sum_{y \in \Delta(x)} (n-2)\log\left(\|\underline{u}(x) - \underline{u}(y)\|^2\right) - \frac{\|\underline{u}(x) - \underline{u}(y)\|^2}{4\sigma_u^2}\right], \tag{8}$$

with $\mathcal{P}$ the set of all possible patches. If we consider patches 3×3 for an image coded in 8 bits, $|\mathcal{P}| = 256^9$, and $\mathcal{Z}(x)$ cannot be computed. But, this term tends to 0 when the patches are different. Hence, it makes sense to only consider $\mathcal{P}_r$ the set of patches in the whole image. Then $|\mathcal{P}_r| = n \times m$, with n and m the image dimensions (without taking into account the border conditions). It turns out that the normalization term is nearly constant in all the image and we adopt the following form:

$$\Phi(x) = \sum_{y \in \Delta(x)} (n-2)\log\left(\|\underline{u}(x) - \underline{u}(y)\|^2\right) - \frac{\|\underline{u}(x) - \underline{u}(y)\|^2}{4\sigma_u^2}, \tag{9}$$

In our experiments, the patch-wise MRF enables to regularize the difference image and the related potential map is noiseless when compared to the potential map obtained with the pixel-wise MRF (see Fig. 1).

3. IRREGULARITY DETECTION AND FALSE ALARM RATE

Given the noise variance σ_u^2, the patch-wise Gibbs model can be then used to assess the regularity of an input difference image. Based on semi-local interactions (9), a potential map $\Phi = (\Phi(x))_{x \in \Omega}$ is computed as explained above. As expected, in the presence of irregularities the potential is high, which means that the proposed model cannot wholly capture all the spatial image features. To detect irregularities, we then propose to automatically threshold the potential map. Since the difference image is relatively homogeneous, only a few areas correspond to high values. We want to test the hypothesis for each pixel x that the potential $\Phi(x)$ is meaningful (hypothesis H_0) or not (hypothesis H_1). A very convenient way to define this notion consists in assuming that the potential $\Phi(x)$ are independent and distributed according to the following mixture

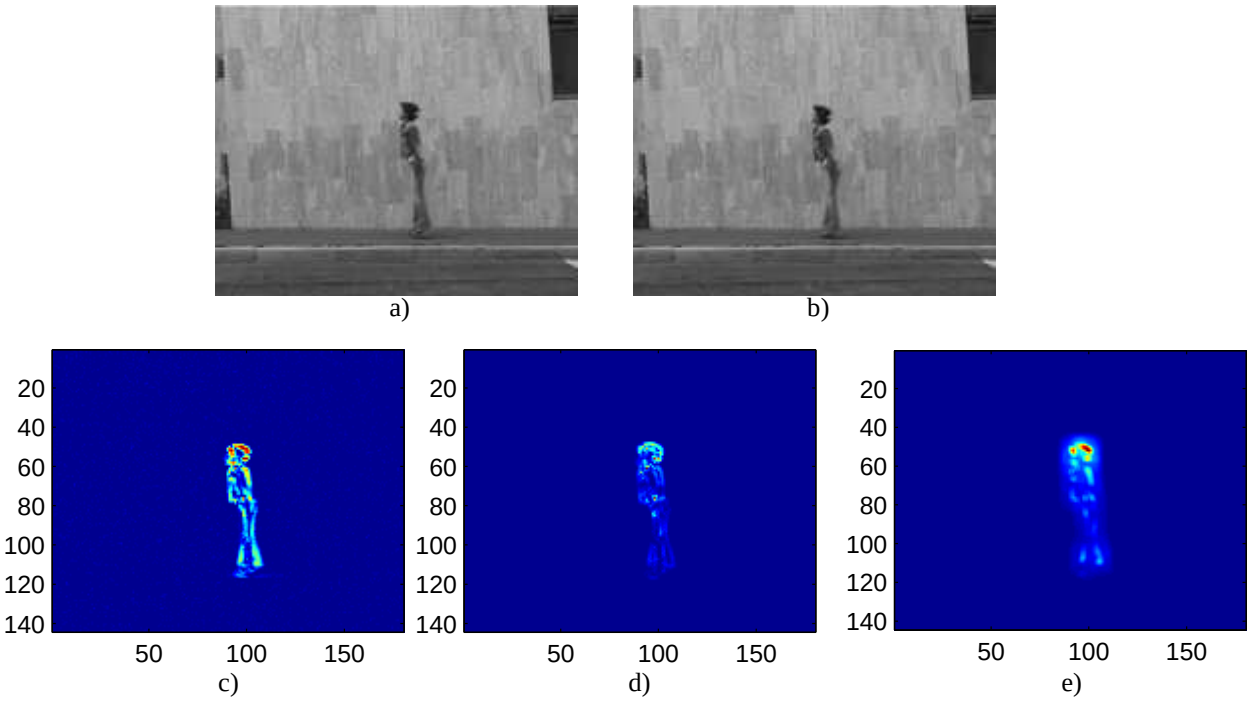

Figure 1. a-b) Two consecutive images taken from an image sequence; c) difference image; d) potential map using pixel-wise MRF modeling; e) potential map using patch-wise MRF modeling (blue regions correspond to low potential values and red regions to high potential values).

distribution:

$$f(\Phi(x)) = \delta(\Phi(x) < \epsilon) f_0(\Phi(x)) + (1 - \delta(\Phi(x) < \epsilon)) f_1(\Phi(x)), \quad (10)$$

where $\delta(\cdot)$ denotes the Kronecker, f_0 denotes the pdf of potential under H_0 and f_1 is the pdf of potentials under H_1. In the mixture model, f_0 is a Dirac function centered at "0" ($\epsilon = 0_+$ subject to $\epsilon > 0$) and the tail f_1 of the pdf is approximated by a Pareto distribution of the form:

$$f_1(\Phi(x)) = \frac{k\epsilon^k}{(\Phi(x))^{k+1}}, \ \forall\, \Phi(x) \geq \epsilon, \quad (11)$$

with parameters k and $\epsilon > 0$. A Maximum Likelihood estimate for k can be easily derived, i.e.:

$$\hat{k} = N \left(\sum_{x \in \Omega:\, \Phi(x) \geq \epsilon} \log(\Phi(x)) - \log(\epsilon) \right)^{-1},$$

where $N = \#\{x \in \Omega : \Phi(x) \geq \epsilon\}$. The Pareto distribution is recommended to describe tails of pdfs and performs well as shown in Fig. 2. Accordingly, the probability that $\Phi(x)$ is larger than a threshold $\tau \geq \epsilon$ is:

$$\mathcal{P}\{\Phi(x) \geq \tau\} = (\tau/\epsilon)^{-k}. \quad (12)$$

Therefore, for a given false alarm probability $\mathcal{P}_{FA} := \mathcal{P}\{\Phi(x) \geq \tau\}$ selected by the user (and assumed to be constant in the whole image), one can compute the corresponding threshold τ as:

$$\tau = \exp\left(\log(\epsilon) - \frac{\log(\mathcal{P}_{FA})}{k} \right). \quad (13)$$

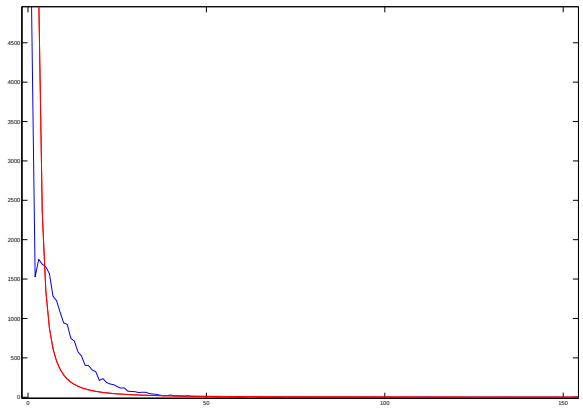

Figure 2. Approximation of the observed pdf (blue) of Φ by a mixture of two distributions (red) for the image sequence shown in Fig. 1.

In practice, we compute first the ML estimate $\hat{k}$ and, for a given false alarm probability $\mathcal{P}_{FA}$, we derive the threshold τ which are different for each processed image in the sequence.

4. EXPERIMENTAL RESULTS

We propose four experiments to illustrate the potential of our method to detect changes between two images. In these experiments, we compare the change detection maps obtained with different patch sizes and neighborhood sizes.

In the first experiment (Fig. 3), a woman is running against a non-uniform background. This woman is far

Figure 3. Two consecutive images (180×144) taken from an image sequence, and the corresponding thresholded difference image.

Figure 4. Two consecutive images (160×120) taken from an image sequence, and the corresponding thresholded difference image.

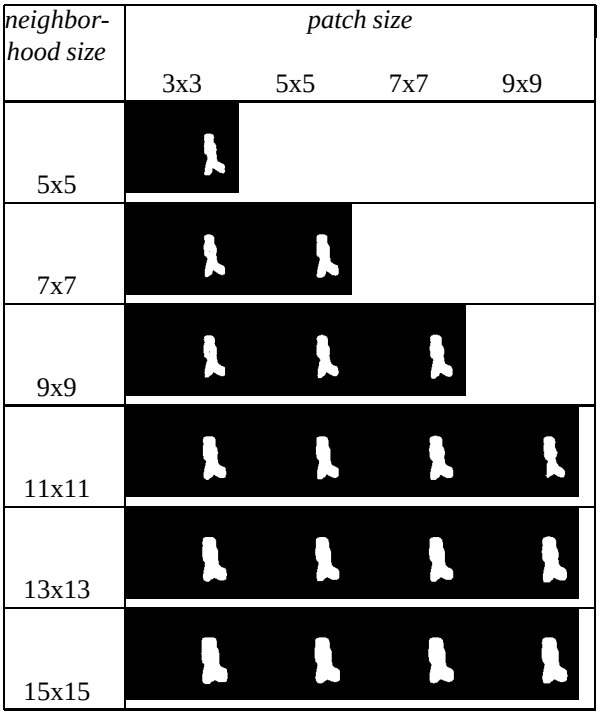

Figure 5. Change detection maps resulting from the difference image of Fig. 3 and obtained for different patch sizes and different neighborhood sizes.

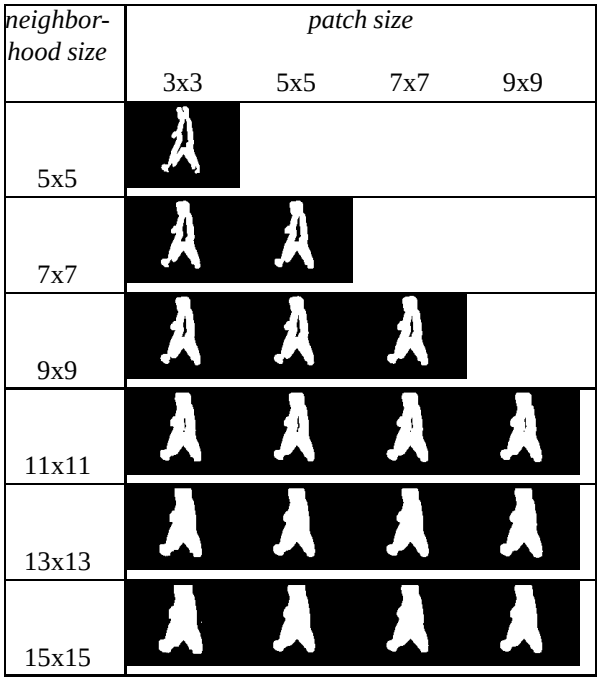

Figure 6. Change detection maps resulting from the difference image of Fig. 4 and obtained for different patch sizes and different neighborhood sizes.

Figure 7. Two consecutive images (180×144) perturbed by a Gaussian noise (standard deviation equal to 10), and the corresponding thresholded difference image.

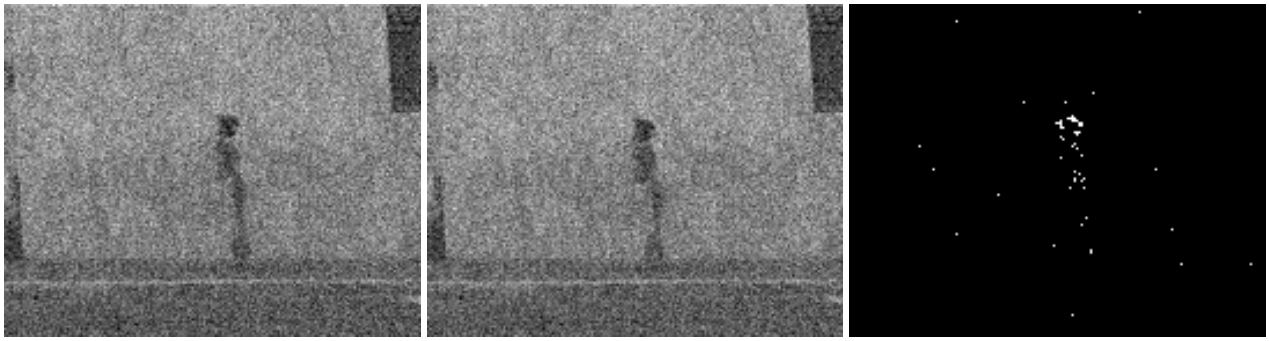

Figure 8. Two consecutive images (180×144) perturbed by a Gaussian noise (standard deviation equal to 30), and the corresponding thresholded difference image.

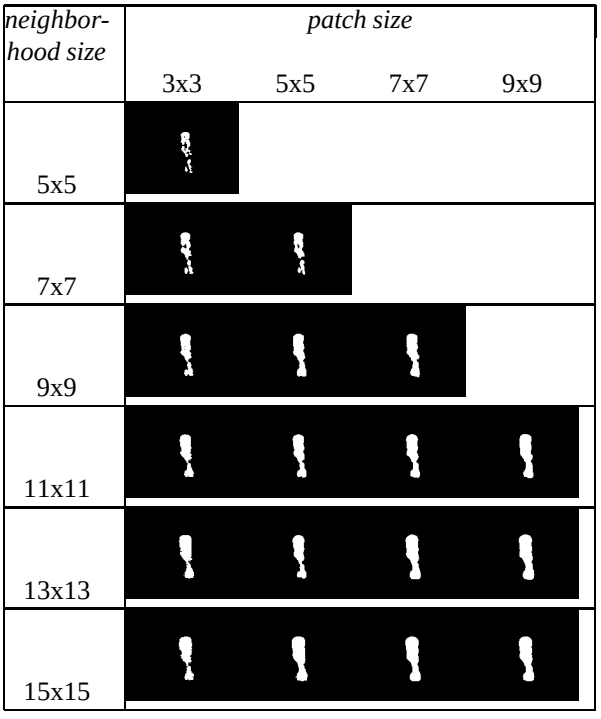

Figure 9. Change detection maps resulting from the difference image of Fig. 7 and obtained for different patch sizes and different neighborhood sizes.

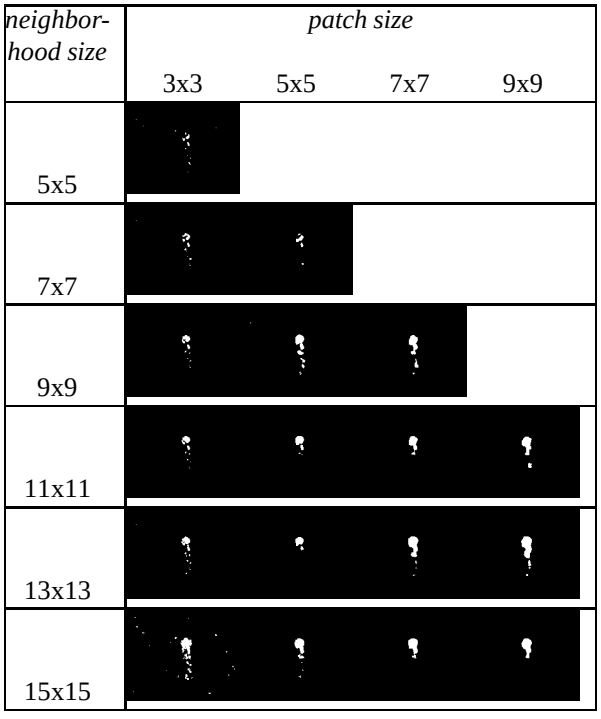

Figure 10. Change detection maps resulting from the difference image of Fig. 8 and obtained for different patch sizes and different neighborhood sizes.

from the video camera and appears as a small object in the image. As expected, the thresholded difference image detects only the woman outlines. The change detection maps obtained with our method are shown in Fig. 5. As the moving object is small, small patch and neighborhood sizes are appropriate to detect the whole woman while larger patch and neighborhood sizes tend to overestimate the size of the person and to blur the outlines.

In the second experiment (Fig. 4), a man is walking against a uniform background. He is close to the video camera and occupies a large space in the image. Once again, the thresholded difference image detects only the outlines, and the shoe outlines are not fully detected because the man does not move enough between the two images. The change detection maps resulting from our method are shown in Fig. 6. As the moving object is large, the patch and/or the neighborhood sizes have to be larger than in the first experiment to detect the whole object. Hence, the patch and neighborhood sizes have to be chosen following the size and field depth of the objects of interest in the images.

In the third experiment (Fig. 7), a woman is jumping against a non-uniform background far from the video camera, and the images are noisy (Gaussian noise with a standard deviation equal to 10). Consequently, the thresholded difference image detects a large number of artifacts that do not correspond to moving objects. On the contrary, the change detection maps (Fig. 9) only detect the moving object. But, even if the woman is far from the video camera and consequently little in the image, the patch and neighborhood sizes have to be large enough because of the noise.

The same successive images than in the previous experiment are used for the fourth experiment (Fig. 8) but noisier (Gaussian noise with a standard deviation equal to 30). In that case (Fig. 10), it is impossible to detect the moving regions with the thresholded difference image. Indeed, in the previous experiment, a large number of artifacts are detected but the moving object too. In this experiment, it is impossible to have an idea about the moving object using the thresholded difference image whereas the change detection maps manage to detect it. The whole object is not detected but its localization in the image is detected. In that case, the patch size must be bigger than 3×3 if we do not want to detect artifacts.

5. CONCLUSION

In this paper, we have proposed a general probabilistic and patch-based framework for irregularity detection in the difference image in order to detect the moving regions in an image sequence. These detection maps are more regular than the difference image for detecting the whole moving objects and are much more robust to low signal-to-noise ratios. Moreover, no Gibbs model based on patch interactions was ever developed to our knowledge. In practice, this method only requires the setting of the false alarm probability.

6. REFERENCES

[1] R. Collins, A. Lipton, and Kanade T., "Introduction to the special section on video surveillance," in *IEEE Trans. Pattern Anal. Mch. Intell.*, 2000, vol. 22, pp. 745–746.

[2] L.M. Tavares de Carvalho, J. Clevers, S.M. De Jong, and A.K. Skidmore, "Forestry database updating based on remote sensing change detection," in *XII Brazilian Remote Sensing Symposium*, april 2005, pp. 465–472.

[3] M. Bosc, F. Heitz, J.P. Armspach, I. Namer, D. Gounot, and L. Rumbach, "Automatic change detection in multimodal serial mri: Application to multiple sclerosis lesion evolution," *Neuroimage*, vol. 20, pp. 643–656, 2003.

[4] N. Otsu, "A threshold selection method from gray-level histogram," in *IEEE Trans. System, Man and Cybernetics*, 1979, vol. 19, pp. 62–66.

[5] P.L. Rosin, "Thresholding for change detection," *Computer vision and image understanding*, vol. 86, pp. 79–95, 2002.

[6] P.L. Rosin and T. Ellis, "Evaluation of global image thresholding for change detection," *Pattern Recognition Letters*, vol. 24, pp. 2345–2356, 2003.

[7] R.J. Radke, A. Srinivas, O. Al-Kofahi, and B. Roysam, "Image change detection algorithms: a systematic survey," *IEEE Trans. Image Processing*, vol. 14, no. 3, pp. 294–307, Mar. 2005.

[8] Y.Z. Hsu, H.-H. Nagel, and G. Reckers, "New likelihood test methods for change detection in image sequences," in *Computer Vision, Graphics, and Image Processing*, 1984, vol. 26, pp. 73–106.

[9] T. Aach and A. Kaup, "Statistical model-based change detection in moving video," *Signal processing*, vol. 31, pp. 165–180, 1993.

[10] J. Besag, "Spatial interaction and the statistical analysis of lattice systems," *Journal Royal Statistical Society*, vol. 36, pp. 192–236, 1974.

[11] A. Buades, B. Coll, and J.M. Morel, "A review of image denoising algorithms, with a new one," *Multiscale Modeling and Simulation: A SIAM Interdisciplinary Journal*, vol. 4, no. 2, pp. 490–530, 2005.

[12] Stefan Roth and Michael J. Black, "Fields of experts: A framework for learning image priors.," in *Proc. of IEEE CVPR'2005*, San Diego, USA, June 2005, vol. 2, pp. 860–867.

[13] S.P. Awate and R.T. Whitaker, "Unsupervised, information-theoretic, adaptive image filtering for image restoration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 3, pp. 364–376, 2006.