Multiscale neighborhood-wise decision fusion for redundancy detection in image pairs

Charles Kervrann, Jérôme Boulanger, Thierry Pecot, Patrick Pérez, Jean Salamero

MULTISCALE NEIGHBORHOOD-WISE DECISION FUSION FOR REDUNDANCY DETECTION IN IMAGE PAIRS

CHARLES KERVRANN *, JÉRÔME BOULANGER †, THIERRY PÉCOT ‡, PATRICK PÉREZ §, AND JEAN SALAMERO ¶

Abstract. To develop better image change detection algorithms, new models able to capture spatio-temporal regularities and geometries present in an image pair are needed. In this paper, we propose a multiscale formulation for modeling semi-local inter-image interactions and detecting local or regional changes in an image pair. By introducing dissimilarity measures to compare patches and binary local decisions, we design collaborative decision rules that use the total number of detections obtained from the neighboring pixels, for different patch sizes. We study the statistical properties of the non-parametric detection approach that guarantees small probabilities of false alarms. Experimental results on several applications demonstrate that the detection algorithm (with no optical flow computation) performs well at detecting occlusions and meaningful changes for a variety of illumination conditions and signal-to-noise ratios. The number of control parameters of the algorithm is small and the adjustment is intuitive in most cases.

Key words. image analysis, image patches, non-parametric estimation, multiscale modeling, change detection, image motion, probability of false alarm

AMS subject classifications. 62H35, 68U10, 94A13, 60G35, 97K80

1. Introduction.

Occlusion and image change detection is a challenging problem for the accurate computation of correspondences in image sequence analysis and stereo vision. Theoretically, the pixels at the occlusion location should not be assigned any flow vector since there is no correspondence available in the other image. In this paper, we define the occluded regions as sets of pixels where the differences between two images are meaningful. These major changes are caused by appearance or disappearance of objects at considered location, and will be considered as not significant if the changes are due to camera motion/jitter or illumination changes in the scene.

1.1. Previous works.

There has been a substantial amount of work to handle changes in an image pair [3, 98, 2, 85, 66, 79, 1, 49]. For a recent survey, see [82]. Actually, change detection is of significant interest in an increasing number of applications, such as video-surveillance (e.g., in airports, museums, shops, etc), medical diagnosis [18, 80, 45, 87, 89], cell biology imaging [77, 19] and remote sensing [22, 54]. The challenge lies in distinguishing between meaningful changes related to unusual scene events and changes corresponding to camera motion, camera noise or atmospheric/lighting conditions. This can be generally achieved by using adaptive thresholds applied to image differencing or to image-background differences. The problem to be addressed further is to integrate the spatial-contextual information from the pixels to cope with camera jitter or animated texture in the background. In [22, 54, 76, 75, 91, 30, 95, 10, 11], several authors proposed to capture the spatial correlation among nearby pixels using Markov random fields (MRF). Energy minimization is usually performed by graph-cut algorithms [30, 95] or level set methods

*INRIA Rennes-Bretagne Atlantique, INRA-MIA Jouy-en-Josas, France (charles.kervrann@inria.fr)

†UMR 144 CNRS-Institut Curie Paris, France (jerome.boulanger@curie.fr)

‡INRIA Rennes-Bretagne Atlantique, INRA-MIA Jouy-en-Josas, UMR 144 CNRS-Institut Curie Paris, France

§INRIA Rennes-Bretagne Atlantique, Technicolor Corporate Research Rennes Laboratory, France (Patrick.Perez@technicolor.com)

¶UMR 144 CNRS-PICT IBiSA-Institut Curie Paris, France (salamero@curie.fr)

[75]. Nevertheless, MRF-based methods need the adjustment of weighting parameters to balance the prior energy terms and the fidelity terms. These parameters are usually adjusted for different image pairs and different signal-to-noise ratios, which may be considered as a limitation. In addition, the level of confidence of detected areas cannot be estimated since one focuses mainly on the global minima of the energy. Meanwhile, deterministic approaches have been also investigated to detect major changes in an image pair. For instance, in [8, 73] the authors proposed to compute the "intersection" of two images from the image level lines to detect meaningful changes. The main advantage is that the topographic map is contrast invariant, which may be attractive to compare two images depicting the same scene but illuminated differently. Nevertheless, as recently explained in [102], all illumination changes cannot be captured by this morphological image representation.

In the area of video analysis, a sequence with no moving object is traditionally used to learn the statistics of the static background [104, 41, 92, 72, 51, 10, 11]. Each pixel of the current image is then tested against the learned probability distribution function (PDF). Stauffer and Grimson [92] were probably the first authors who proposed a mixture of Gaussians to approximate the PDF of the background at each pixel. In motion analysis, detection of occluded areas is also known to be critical, especially when displacements are large [5, 93]. Therefore, several methods have attempted to simultaneously detect motion discontinuities and to compute optical flow [70, 64], or to detect the violation of motion consistency assumption [47, 49, 105]. Nevertheless, the optical flow estimated in occluded regions often appears over-smoothed and inaccurate in most applications. Curiously, not so much work has been done to handle the occlusion problem in the motion estimation area [14], whereas this problem has been widely studied in the context of stereo algorithms [35, 50, 60, 61, 88]. The most recent algorithms based on graph-cuts [101] or loopy belief propagation [103, 94] include a visibility label in the energy formulation to compute dense disparity [101]. Recently, Xiao *et al.* proposed to integrate occlusion penalties into the graph-cut framework by using a set of three-state pixel graphs with very impressive results [106]. A probabilistic framework for occlusion detection based on generative models was also studied by Fransens *et al.* in [42]. Nevertheless, the quality of occlusion detection based on optical flow techniques is not always satisfactory for processing real challenging and noisy sequences.

1.2. Our approach and related work. Our idea for better handling both occlusions and other sources of changes originates from the observation that two successive images are redundant ; the occlusions and change regions correspond to areas in one image which cannot be found in the second image. Our formulation is inspired by the Efros and Leung's exemplar-based approach for texture synthesis [39] and the detectors of repeated scene elements and self-similarities captured by patches, as respectively introduced in [63] and [90]. The redundancy property captured by image patches was previously exploited for image segmentation [44, 58], image denoising [17, 24, 56, 7, 40, 86], image inpainting [31], defect detection in images [108] and image representation [71]. The approach we propose is able to deal with situations as challenging as those presented in Fig. 1.1. First we assume that, to each patch in the first image, corresponds a small set of similar patches in the other image but not necessary an unique one. This was already suggested for image sequence denoising in [20, 23]. To detect the occlusions or changes occurring in two images, we propose further to collect a set of binary decisions obtained from pixels in a local neighborhood. This amounts to counting the number of neighboring patches in the second image which

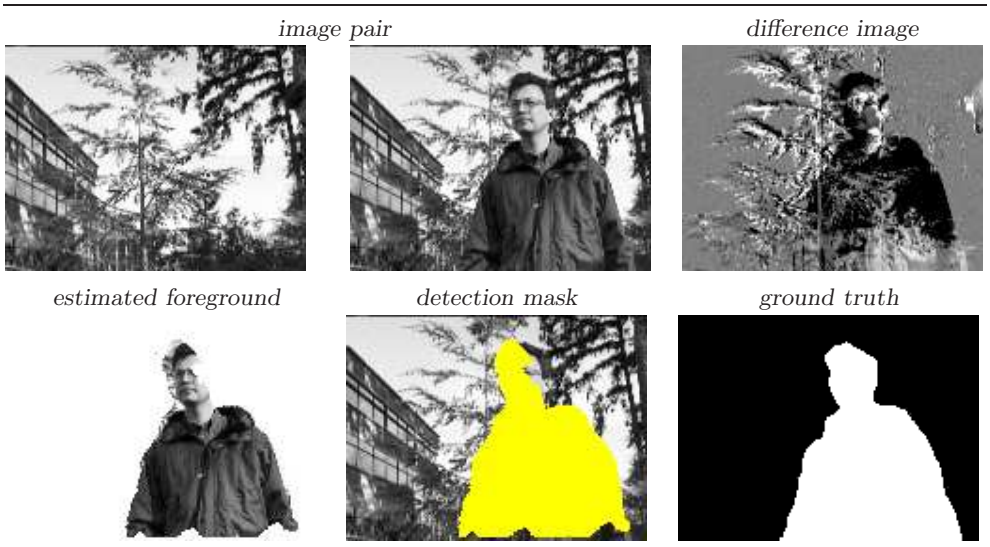


FIG. 1.1. *Change detection in an image pair (see [99, 46]). Our algorithm correctly highlights the person who appeared in the second image, while ignoring temporal changes due to complex motion of tree branches in the background.*

are similar to the current patch in the first image. At the current pixel, we make a neighborhood-wise decision by aggregation of the local decisions ("1" for change and "0" for no change), according to a decision-fusion principle [74, 55]. Finally, no additional regularization process is required to improve the binary detection maps since we benefit from the regularization property induced by the patches' overlap.

Since we compute probabilities of false alarms, our probabilistic approach is related to the *a contrario* modeling already investigated for change detection in [66, 100, 97, 87, 37, 83]. In the same spirit, Sabater *et al.* proposed recently in [88] a sophisticated *a contrario* block matching method to guarantee that on average not more than one wrong block match occurs in the image. Our method is also related to methods based on neighborhood agreement and votes as proposed in [66, 100, 97, 49]. Also, in [1] the authors performed statistical tests (under Gaussian hypothesis) from pixels within sliding windows (see also [2, 49]) as we also suggest ; in [4], the authors proposed a tracking algorithm (*"Frag-Track"*) which combines multiple votes and histogram comparisons in spatial neighborhoods [4] ; in [15, 16], the authors presented a generative and Bayesian method to detect unusual situations in an image sequence. Few examples on image pairs (visual inspection and defect detection) are reported in [16] but no objective comparison with existing change detection methods are given.

1.3. Main contributions. We propose an unsupervised change detection method based on binary decision analysis in local neighborhoods. The method is relatively straightforward since it amounts to counting similar patches in local neighborhoods and to comparing this number to adaptive thresholds. Unlike exemplar-based approaches, we examine the decisions made for different patch sizes and we provide a probabilistic multiscale framework to make final decisions with formal rules. Other contributions include:

1. Proposing a probabilistic framework based on detection theory to calculate the spatially-varying levels of confidence (i.e. probability of false alarm) for

each pixel. Generally, benchmarks and labeled images are necessary to evaluate the global performance of detectors. Our modeling framework can be used for evaluating the level of confidence of occlusions for any image pair.
2. Developing a method able to produce regularized detection maps with no explicit spatial regularization, no precise optical flow computation, no strong prior image model and no labeling/training stage as recommended with discriminative random field (DRF) [62, 67, 96].
3. Addressing the issue of robustness in image correspondence with respect to appearance variation due to illumination changes in the scene and low signal-to-noise ratios.
4. Estimating the detection thresholds with some originality.

Our method needs no statistical assumption on image pairs and noise and is able to robustly handle small object shifts without compensating camera motion. Unlike many subtraction methods which need a series of recorded training images, our algorithm is able to produce satisfying results using an image pair only. It is worth noting that MRF-based methods are traditionally used for change detection but the level of confidence of occluded areas cannot be estimated since one focuses on the global minima of the energy. In this paper, we address also the issue of parameter setting and the determination of spatially-varying detection thresholds. In the second part of the paper, we propose a MRF-CRF (Conditional Random Field) modeling with explicit spatial priors for regularization to compare experimentally the two approaches.

The remainder of the paper is organized as follows: in Section 2, we describe basically the approach based on patch comparisons and we present a family of dissimilarity measures, robust to gradual or severe variations in the appearance. In Section 3, we describe the probabilistic approach for image pair analysis and present collaborative decision rules in neighborhoods. Section 4 describes the multiscale framework to fuse binary decisions at different spatial scales. In Section 5, we present the algorithm and we propose a strategy to set the control parameters. In Section 6, we examine the properties of the detector. Finally, in Section 7 we present experimental results on several examples and we study an alternative approach based on global energy minimization in the spirit of most approaches in computer vision.

2. Image redundancy and change detection. Unlike usual methods (kernel-based [41, 72, 91] or mixture of Gaussian-based [92, 107, 51]) which assume a time series of images as input, we consider the scenario where we have in our possession two images, without prior knowledge of the scene as considered in [65]. Background subtraction based on temporal information cannot be performed as described in [41, 92, 30, 95, 32] since we are using two images only.

2.1. Principle. In order to describe our detection method, let us first introduce some useful notations. Consider a gray-scale image pair $u = (u(\mathbf{x}))_{\mathbf{x}\in\Omega}$ and $v = (v(\mathbf{x}))_{\mathbf{x}\in\Omega}$ defined over a bounded domain $\Omega \subset \mathbb{R}^2$. Our study examines the situations where a change occurs in the image pair (u, v). In order to test robustly the similarity between u and v, we focus on image patches as semi-local image features able to capture local geometries and contextual information.

Our idea is to guess a patch at location $\mathbf{x}$ in u from patches at locations $\mathbf{y}$ taken in the (fixed size) semi-local neighborhood $B(\mathbf{x}) \subset \Omega$ where $N = |B(\mathbf{x})|$ denotes the number of elements observed at locations $\mathbf{y}$ in the second image v. As a starting point, if the distance between the reference patch at location $\mathbf{x}$ and the patches at locations $\mathbf{y} \in B(\mathbf{x})$ taken respectively in u and v is large enough, we can conclude

that a meaningful change has occurred provided that the global illumination of the scene is the same in the two images. In [23, 39, 20], it has been confirmed that the L_2 distance is a good candidate to express the amount of dissimilarity between image patches. Also, theoretically the range of the search space can be as large as the whole image for scene change detection since the occluded objects are not present in one of the two images. In the area of image sequence analysis, the search space is smaller but assumed to be larger than the expected maximum motion amplitude of moving objects in the scene. Nevertheless, we shall see that examining small neighborhoods enables to detect occlusions reliably for most studied situations.

In what follows, we propose to make local binary decisions about the presence/absence of patches in the second image v that are similar to the reference patch in the first image u. To infer the decision thresholds, we studied the usual Gaussian image model where $u = u_0 + \epsilon$ and $v = v_0 + \eta$ are assumed to be noisy versions of "true" images u_0 and v_0 and the "errors" ϵ and η are i.i.d (independent identically distributed) Gaussian zero-mean random variables. This parametric modeling approach is probably too restrictive to analyze any real image pair, but this starting point is usually considered in several change detection applications (e.g. see [85, 2, 82]). First, we conducted experiments and examined the probability distribution function (PDF) of dissimilarity measures (L_2 distances) between patches taken in the two input images. Even by considering homogeneous lighting conditions, the same patch sizes and the same neighborhood sizes, the PDFs are affected by the area of missing objects and accordingly, are not stable from an image pair to another. Furthermore, we considered the approximation of these PDFs by mixtures of Gaussians. Mixture of Gaussians is a powerful technique to estimate the modes and the tails of the PDFs. Unfortunately it was hard to estimate a valuable threshold to detect meaningful changes on all tested pairs. Learning the variabilities from the tested image pair (i.e. without training set) seems not to be desirable for meaningful change detection. Finally, these ideas were not investigated further and we decided rather to exploit a single reference image to estimate a specific detection threshold for each individual pixel.

2.2. A family of patch dissimilarity measures. The proposed detection method is based on the patch dissimilarity measure (0 when patches are maximally similar), usually chosen as the weighted distance between patches:

$$\phi_{uv}(\mathbf{x}, \mathbf{y}) = \sum_{\mathbf{t} \in \mathbb{R}^2} W_n(\mathbf{t})\, g(u(\mathbf{x}+\mathbf{t}), v(\mathbf{y}+\mathbf{t})), \tag{2.1}$$

where $g : \mathbb{R} \to \mathbb{R}^+$ is a measurable function fixed in advance by the user (measuring a kind of distance between two image pixels). The dissimilarity measure is non-negative and the smaller the value of the distance is, the more similar patches are. The function $W_n(\cdot)$ is used to assign spatial weights to the patch elements and n refers to the size of the square or circular patches. An usual L_2 distance between square $\sqrt{n} \times \sqrt{n}$ patches is obtained if we choose $W_n(\mathbf{t}) = \mathrm{rect}_n(\mathbf{t}) \triangleq \mathbb{1}[\|\mathbf{t}\|_\infty \leq \frac{\sqrt{n}}{2}]$ where $\mathbb{1}[\cdot]$ is the indicator function. If $W_n(\mathbf{t}) = G_n(\mathbf{t}) \triangleq e^{-\|\mathbf{t}\|_2^2/n}$, the central pixels in the patch contribute more to the distance than the pixels located at the periphery.

Obviously many dissimilarity measures are not invariant to a number of transformations which may arise in most applications, especially variations in brightness. Such cases are very common and the detection method should be robust to moving cast shadows, gradual or sudden intensity variations, specularities, changing lighting directions. This problem can be alleviated by explicitly removing shadows, specularities or undesirable effects in a pre-processing stage. For instance, the image pairs can

be normalized by exploiting the general framework based on intrinsic images and illumination eigenspaces (e.g. see [69]). Nevertheless, a time series of images is generally required and the strategy we adopt here consists in considering other dissimilarity measures (see [2, 21]) defined as:

$$\phi_{uv}(\mathbf{x},\mathbf{y}) = \sum_{\mathbf{t}\in\mathbb{R}^2} W_n(\mathbf{t})\,((u(\mathbf{x}+\mathbf{t})-u_\rho(\mathbf{x}))-(v(\mathbf{y}+\mathbf{t})-v_\rho(\mathbf{y})))^2, \tag{2.2}$$

$$\phi_{uv}(\mathbf{x},\mathbf{y}) = \sum_{\mathbf{t}\in\mathbb{R}^2} W_n(\mathbf{t})\,\left(u(\mathbf{x}+\mathbf{t})-\frac{u_\rho(\mathbf{x})}{v_\rho(\mathbf{y})}v(\mathbf{y}+\mathbf{t})\right)^2, \tag{2.3}$$

$$\phi_{uv}(\mathbf{x},\mathbf{y}) = 1-\frac{\sum_{\mathbf{t}\in\mathbb{R}^2} W_n(\mathbf{t})\,u(\mathbf{x}+\mathbf{t})v(\mathbf{y}+\mathbf{t})}{\sqrt{\sum_{\mathbf{t}\in\mathbb{R}^2} W_n(\mathbf{t})\,u^2(\mathbf{x}+\mathbf{t})}\times\sqrt{\sum_{\mathbf{t}\in\mathbb{R}^2} W_n(\mathbf{t})\,v^2(\mathbf{y}+\mathbf{t})}}, \tag{2.4}$$

where $u_\rho = G_\rho \star u$ and $v_\rho = G_\rho \star v$ are images convolved with a Gaussian kernel G_ρ with standard deviation ρ. The dissimilarity measure (2.2) is introduced to eliminate unwanted changes corresponding to global or local additive contrasts between the two input images. Under the Lambertian assumption, the relation between observed intensity, illumination and reflectance is multiplicative [2, 69]. Accordingly, invariance to multiplicative brightness changes can be obtained by considering the dissimilarity measure (2.3). A symmetric version defined as $(\frac{u(x+t)}{u_\rho(x)}-\frac{v(y+t)}{v_\rho(y)})^2$ can be considered also and the experimental results shown in Section 7.1 are very close to those obtained with (2.3). Note that ρ influences the computation in matching, and setting ρ to a very large value means that brightness variation between the two images is global and the same for every pixels. The correlation dissimilarity measure (2.4) only considers the similarity between the angles formed by vectorized image patches and discards the scaling on the magnitude. These dissimilarity measures may have the advantage of being not sensitive to illumination variation and outperforms the dissimilarity measure (2.5) in that case. The following dissimilarity measures are not invariant to global illumination variations but may be considered also in some applications:

$$\phi_{uv}(\mathbf{x},\mathbf{y}) = \sum_{\mathbf{t}\in\mathbb{R}^2} W_n(\mathbf{t})\,(u(\mathbf{x}+\mathbf{t})-v(\mathbf{y}+\mathbf{t}))^2, \tag{2.5}$$

$$\phi_{uv}(\mathbf{x},\mathbf{y}) = \sum_{\mathbf{t}\in\mathbb{R}^2} W_n(\mathbf{t})\,(u(\mathbf{x}+\mathbf{t})-v(\mathbf{y}+\mathbf{t}))^p, \tag{2.6}$$

$$\phi_{uv}(\mathbf{x},\mathbf{y}) = \sum_{\mathbf{t}\in\mathbb{R}^2} W_n(\mathbf{t})\,\left(\sum_{p=0}^{m}(u^{(p)}(\mathbf{x}+\mathbf{t})-v^{(p)}(\mathbf{y}+\mathbf{t}))^2\right), \tag{2.7}$$

$$\phi_{uv}(\mathbf{x},\mathbf{y}) = \sum_{\mathbf{t}\in\mathbb{R}^2} W_n(\mathbf{t})\,\left[\frac{(u(\mathbf{x}+\mathbf{t})-v(\mathbf{y}+\mathbf{t}))^2}{4\sigma^2}-\log\left(\mathrm{erf}\left(\frac{u(\mathbf{x}+\mathbf{t})+v(\mathbf{y}+\mathbf{t})}{2\sigma}\right)\right.\right.$$
$$\left.\left.+\,\mathrm{erf}\left(\frac{2\max(\sup_{\mathbf{x}\in\mathbb{R}^2}u(\mathbf{x}),\sup_{\mathbf{x}\in\mathbb{R}^2}v(\mathbf{x}))-u(\mathbf{x}+\mathbf{t})-v(\mathbf{y}+\mathbf{t})}{2\sigma}\right)\right)\right], \tag{2.8}$$

where (2.7) is based on the m first derivatives of u and v (Sobolev norm) and (2.8) is related to the Maximum Likelihood dissimilarity measure [68] in the case of intensity-independent Gaussian noise with variance σ^2 where $\mathrm{erf}(\cdot)$ denotes the Gauss error function.

In the next sections, ϕ_{uv} will denote any of these dissimilarity functions. Additional dissimilarity measures and functions $g(\cdot)$ based on the normalized cross-correlation [78] and more general probabilistic dissimilarity measures [84, 68] can be also investigated. In our experiments, we will focus on (2.3), (2.4) and (2.5) for

demonstration. Also it is worth noting that the input images can be preliminarily modified by local histogram equalization that preserves image level lines [28, 102], midway image equalization [34] or dynamic histogram warping [29].

3. Neighborhood-wise decision. In this section, we describe the proposed probabilistic framework for change detection. We have sketched a procedure based on patch comparisons, which tells whether a local change at the current pixel has occurred or not. In what follows, we make a collective decision from binary decisions obtained at neighboring points.

3.1. Fusion of local binary decisions. Consider a reference patch at location $\mathbf{x}$ to be compared to neighboring patches at locations $\mathbf{y} \in B(\mathbf{x})$ in the second image v. Let $\phi_{uv}(\mathbf{x}, \mathbf{y})$ denote the dissimilarity measure at pixel $\mathbf{y}$ defined for instance as the L_2 distance between two patches (as explained in Section 2.2):

$$\phi_{uv}(\mathbf{x}, \mathbf{y}) \;=\; \sum_{\mathbf{t} \in \mathbb{R}^2} W_n(\mathbf{t}) \, (u(\mathbf{x} + \mathbf{t}) - v(\mathbf{y} + \mathbf{t}))^2. \tag{3.1}$$

Based on $\phi_{uv}(\mathbf{x}, \mathbf{y})$, we make a binary decision regarding the presence of a similar patch in the semi-local search area. The decision depends on whether $\phi_{uv}(\mathbf{x}, \mathbf{y})$ exceeds a threshold $\tau(\mathbf{x})$.

Further, the collaborative neighborhood-wise decision for change detection is obtained through a fusion rule from individual binary decisions in the neighborhood $B(\mathbf{x})$. An intuitive choice is to count the total number of positive decisions defined as:

$$S_N(\mathbf{x}) \;=\; \sum_{\mathbf{y} \in B(\mathbf{x})} \mathbb{1}[\phi_{uv}(\mathbf{x}, \mathbf{y}) \geq \tau(\mathbf{x})], \tag{3.2}$$

where $N = |B(\mathbf{x})|$ and $\tau(\mathbf{x})$ is a spatially-varying threshold. If the number of decisions exceeds the overall decision threshold $T \in \{1, \cdots, N\}$ assumed to be constant for the whole image, we declare that a change occurs at pixel $\mathbf{x}$. The neighborhood-wise decision $D(\mathbf{x}) \in \{0, 1\}$ at pixel $\mathbf{x}$ is then defined as:

$$D(\mathbf{x}) \;\stackrel{\triangle}{=}\; \mathbb{1}[S_N(\mathbf{x}) \geq T]. \tag{3.3}$$

Cooperation among neighboring points tends potentially to enhance the ability to detect meaningful changes, namely, whether a change occurs or not within the search area $B(\mathbf{x})$. This natural decision procedure that fuses local binary decisions has been already proposed in [49, 52] for change detection and in [74, 55] for analyzing distributed sensors in a wireless network.

3.2. Controlling the number of false alarms. In our framework, we make no assumption about the image formation since there may be different sources of occlusions present in the image which cannot be modeled *a priori*. We require only $\phi_{uv}(\mathbf{x}, \mathbf{y}) > 0$. The null hypothesis H_0 of no occlusion present in the image is defined as follows:

DEFINITION 3.1. *Under the null hypothesis H_0, for $\mathbf{y} \in B(\mathbf{x})$, the events $\{\phi_{uv}(\mathbf{x}, \mathbf{y}) \geq \tau(\mathbf{x})\}$ are independent and occur with probability $\mathbf{P}(\phi_{uv}(\mathbf{x}, \mathbf{y}) \geq \tau(\mathbf{x}) | H_0)$.*

If we assume that the events $\{\phi_{uv}(\mathbf{x}, \mathbf{y}) \geq \tau(\mathbf{x})\}$ are independent, the probability that there exists a collection of at least T pixels $\mathbf{y} \in B(\mathbf{x})$ which declare $\phi_{uv}(\mathbf{x}, \mathbf{y}) \geq$

$\tau(\mathbf{x})$ is given by:

(3.4)

$$P_{fa}(\mathbf{x},T) \;=\; \sum_{k=T}^{N} \sum_{\mathcal{X}^k \in \mathcal{X}_N^k} \left(\prod_{\mathbf{y}\in\mathcal{X}^k} \mathbf{P}(\phi_{uv}(\mathbf{x},\mathbf{y}) \geq \tau(\mathbf{x})|H_0) \right) \left(\prod_{\mathbf{y}\in\bar{\mathcal{X}}^k} (1 - \mathbf{P}(\phi_{uv}(\mathbf{x},\mathbf{y}) \geq \tau(\mathbf{x})|H_0)) \right),$$

where $\mathcal{X}^k$ is a set of k pixels ($\bar{\mathcal{X}}^k$ is the complementary set of $\mathcal{X}^k$) and $\mathcal{X}_N^k$ is the set of all such sets. Note that the computation burden of (3.4) is known to become daunting as the number of pixels increases in $B(\mathbf{x})$. In the simple case (also considered in the framework) where the probabilities $\mathbf{P}(\phi_{uv}(\mathbf{x},\mathbf{y}) \geq \tau(\mathbf{x})|H_0)$ are identical (i.e. the events $\{\phi_{uv}(\mathbf{x},\mathbf{y}) \geq \tau(\mathbf{x})\}$ are i.i.d) for all the pixels $\mathbf{y} \in B(\mathbf{x})$, i.e.

$$\mathbf{P}(\phi_{uv}(\mathbf{x},\mathbf{y}) \geq \tau(\mathbf{x})|H_0) \;\stackrel{\triangle}{=}\; p_{fa}(\mathbf{x},\tau(\mathbf{x})), \tag{3.5}$$

the probability of false alarm $P_{fa}(\mathbf{x},T)$ defined in (3.4) reduces to the well-known right tail of the binomial distribution:

$$P_{fa}(\mathbf{x},T) \;\stackrel{\triangle}{=}\; \sum_{k=T}^{N} \left(\begin{array}{c} N \\ k \end{array} \right) (p_{fa}(\mathbf{x},\tau(\mathbf{x})))^k (1 - p_{fa}(\mathbf{x},\tau(\mathbf{x})))^{N-k}. \tag{3.6}$$

In practice, there is no way to derive an explicit form of $p_{fa}(\mathbf{x},\tau(\mathbf{x}))$ at location $\mathbf{x}$. The probabilities depend on the spatial position and the location of occluded pixels are not known in the image. To overcome this difficulty, we examined several upper-bounds of $P_{fa}(\mathbf{x},T)$. The Chebyshev's and Chernoff's upper-bounds actually depend on the knowledge on $p_{fa}(\mathbf{x},\tau(\mathbf{x}))$ which is not accessible for our purpose. Accordingly, to compute an upper-bound, we propose to use the following lemma:

LEMMA 3.2. *From the Markov's inequality, the probability of false alarm is bounded as*

$$P_{fa}(\mathbf{x},T) \;=\; \mathbf{P}\left(e^{S_N(\mathbf{x})} \geq e^T|H_0\right) \;\leq\; e^{-T}\mathbf{E}\left[e^{S_N(\mathbf{x})}|H_0\right].$$

The bound in Lemma 3.2 suggests that the probability of false alarms can be bounded from above at each location in the image, provided that the threshold $\tau(\mathbf{x})$ is known and $\mathbf{E}\left[e^{S_N(\mathbf{x})}|H_0\right]$ can be empirically computed as it is proposed in Section 4.2.

Actually, since it is unrealistic to choose $\tau(\mathbf{x})$ to achieve a desired local probability of false alarm, we investigate strategies for estimating these spatially-varying thresholds. In the next section, we address also the setting of parameter T, the size n of patches and the size of the search window $B(\mathbf{x})$ and we recommend a multiscale approach to reduce the number of false alarms.

4. Multiscale probabilistic decision framework. The multiscale approach, which uses the property of patch repetitions across scales, is adopted traditionally to analyze several spatial contexts [43, 71, 38, 81, 33].

4.1. Parameter setting and motivations. For a given patch size n and a neighborhood size N, at each location $\mathbf{x}$ in the image domain, the proposed adaptive decision mechanism is as follows:

1. for each pixel $\mathbf{y} \in B(\mathbf{x})$ in v, the dissimilarity measure $\phi_{uv}(\mathbf{x},\mathbf{y})$ is compared to the threshold $\tau(\mathbf{x})$;
2. given $S_N(\mathbf{x}) = \sum_{\mathbf{y}\in B(\mathbf{x})} \mathbb{1}[\phi_{uv}(\mathbf{x},\mathbf{y}) \geq \tau(\mathbf{x})]$, we make a neighborhood-wise decision $D(\mathbf{x}) = \mathbb{1}[S_N(\mathbf{x}) \geq T] \in \{0,1\}$ at location $\mathbf{x}$.

While this procedure is straightforward, the detection results depend very much on the threshold T and patch size n. First, to produce a very low number of false alarms, a natural idea is to set $T = N$, yielding to

$$\mathbf{P}\left(S_N(\mathbf{x}) = N|H_0\right) \;=\; P_{fa}(\mathbf{x}, N) \;=\; \left(p_{fa}(\mathbf{x}, \tau(\mathbf{x}))\right)^N \tag{4.1}$$

since the probabilities $p_{fa}(\mathbf{x}, \tau(\mathbf{x})) \equiv \mathbf{P}(\phi_{uv}(\mathbf{x}, \mathbf{y}) \geq \tau(\mathbf{x})|H_0)$ are assumed to be identical for all pixels $\mathbf{y} \in B(\mathbf{x})$. This means that a change occurs at pixel $\mathbf{x}$ if all the distances are higher than $\tau(\mathbf{x})$, i.e. all the pixels are in agreement in the neighborhood $B(\mathbf{x})$. By setting $T = N$ also justified in [94, 52], a pixel $\mathbf{x}$ with no match is then considered as being subject to a significant change, or occluded. To avoid the precise setting of the size n of patches, we propose to embed the previous procedure in a multiscale framework described below.

4.2. Multiscale and probabilistic modeling. Let $D_\ell(\mathbf{x})$ be the binary random variable whose value is 1 at location $\mathbf{x}$ when a change is detected for a given patch size $n_\ell = (2\ell+1)^2, 1 \leq \ell \leq L$, and 0 otherwise. In what follows, L is the number of patch sizes considered at each location and the size of $B(\mathbf{x})$ is constant for all pixels in the image whatever the considered patch size. We denote $\{D_1(\mathbf{x}), \cdots, D_L(\mathbf{x})\}$ the set of binary random variables at each location $\mathbf{x} \in \Omega$.

In our analysis, for very small search windows, a pixel is likely associated to a detected change for a large number of different and strongly correlated patch sizes at the same location in the image. Accordingly, the Bernouilli random variables $D_\ell(\mathbf{x})$ are *not identically distributed* and *not really independent* mainly because the patches with different sizes are nested. The decisions are correlated for two successive patch sizes since they have many pixels in common at a given location $\mathbf{x}$. Nevertheless, the sum $\sum_{\ell=1}^{L} D_\ell(\mathbf{x})$ of Bernouilli random variables is known to converge in distribution to a Poisson distribution, provided that the dependencies between the variables are not too large. Actually, there are several ways to prove a Poisson approximation result and a general formulation is as follows: let $D_1(\mathbf{x}), D_2(\mathbf{x}), \cdots, D_L(\mathbf{x})$ be such that $\mathbf{E}[D_\ell(\mathbf{x})|H_0] = P_{fa,\ell}(\mathbf{x}, T), 1 \leq \ell \leq L$ where $P_{fa,\ell}(\mathbf{x}, T)$ is the probability of false alarm (3.6) for a given patch size $n_\ell = (2\ell + 1)^2$. The probability of success is different for each Bernouilli variable $D_\ell(\mathbf{x})$ and we define $\lambda_T(\mathbf{x}) \;\triangleq\; \sum_{\ell=1}^{L} P_{fa,\ell}(\mathbf{x}, T)$.

Chen-Stein method for Poisson approximation. According to the Chen-Stein method, $\sum_{\ell=1}^{L} D_\ell(\mathbf{x})$ tends in distribution to a Poisson law with mean $\lambda_T(\mathbf{x})$ as $L \to \infty$. A friendly exposition of the Chen-Stein method for deriving Poisson approximations in terms of bounds on the total variation distance can be found in [6]. By convention, we denote the total variation distance between two discrete probability distributions P and Q on $\mathbb{N}$ by:

$$d_{TV}(P, Q) = \sup_{A \subset \mathbb{N}} |P(A) - Q(A)|, \tag{4.2}$$

and we use the theorem statement below to evaluate the distance between the distribution of the sum of dependent random indicator (Bernouilli) variables and the Poisson distribution.

THEOREM 4.1. **(Arratia 1990)** *[6],* **(Barbour 1992)** *[9] Let $I = \{1, 2, \cdots, |I|\}$ be a finite index set where $|I|$ denotes the cardinal of I and let $\{X_i : i \in I\}$ be non-independent Bernouilli variables with $p_i = \mathbf{P}(X_i = 1)$. For each variable X_i we choose*

a neighborhood set $\aleph_i \subset I$ of indices such that X_i is independent of variables X_j with $j \notin \aleph_i$. We also require $i \in \aleph_i$. Let $Y = \sum_{i\in I} X_i$ and let Z be a Poisson random variable with mean $\lambda = \mathbf{E}[Y] = \sum_{i\in I} \mathbf{E}[X_i] < \infty$. Then the total variation distance between the distribution of Y and the distribution of Z satisfies:

$$d_{TV}(\mathcal{L}(Y), \mathcal{P}(Z)) \;\leq\; \frac{1-e^{-\lambda}}{\lambda}(Q_1+Q_2) + \min\left(1, \sqrt{\frac{2}{\lambda e}}\right) Q_3,$$

where $\mathcal{L}(Y)$ denotes the distribution of Y, $\mathcal{P}(Z)$ denotes a Poisson distribution with mean λ and

$$\begin{aligned} Q_1 &\triangleq \sum_{i\in I} \sum_{j\in\aleph_i} \mathbf{E}[X_i]\mathbf{E}[X_j], \\ Q_2 &\triangleq \sum_{i\in I} \sum_{j\in\aleph_i, j\neq i} \mathbf{E}[X_i X_j], \\ Q_3 &\triangleq \sum_{i\in I} \mathbf{E}[|\mathbf{E}[X_i | X_j : j\notin \aleph_i] - \mathbf{E}[X_i]|]. \end{aligned}$$

The term Q_1 describes the size of the neighborhoods. The term Q_2 can be viewed as a second order interaction term and accounts for the strength of the dependence inside the set $\aleph_i$. The term Q_3 is related to the strength of dependence of X_i with random variables outside its neighborhood; in particular $Q_3 = 0$ if X_i is independent of $\{X_j : j \notin \aleph_i\}$.

Convergence to the Poisson distribution happens when all Q_1, Q_2 and Q_3 tend to zero as $|I| \to \infty$. The main task is to compute these terms and determine the conditions when they become negligibly small.

The evaluation of the Q_1 and Q_2 terms involve the computation of the first two moments of Y only. Moreover, if we choose $\aleph_i$ as the set $\aleph_i = \{i-w, i-w+1, \cdots, i, \cdots, i+w-1, i+w\}$, with $i \in \aleph_i$ and $|\aleph_i| = 2w+1$, it follows that $Q_3 = 0$ since the Bernouilli variable X_i is independent of X_j for $j \notin \aleph_i$. To calculate Q_1 and Q_2, we assume that $\mathbf{P}(X_i = 1) \approx p, \forall i \in I$ and $|\aleph_i|$ is constant as i varies over I. It follows that $\lambda = \sum_{i=1}^{|I|} \mathbf{E}[X_i] = \sum_{i=1}^{|I|} \mathbf{P}(X_i = 1) \approx |I|p < \infty$ and

$$Q_1 \;=\; \sum_{i=1}^{|I|} \sum_{j\in\aleph_i} \mathbf{P}(X_i=1)\mathbf{P}(X_j=1) \;=\; \sum_{i=1}^{|I|} \sum_{j\in\aleph_i} p^2 \;=\; (2w+1)|I|p^2. \tag{4.3}$$

To compute the bound Q_2, we assume that $\mathbf{E}[X_i X_j]$ is independent of i since all the X_is are identically distributed and have the same pairwise dependencies:

$$\begin{aligned} Q_2 &= \sum_{i=1}^{|I|} \sum_{j\in\aleph_i, j\neq i} \mathbf{P}(X_i = X_j = 1), \\ &= \sum_{i=1}^{|I|} p \sum_{j\in\aleph_i, j\neq i} \mathbf{P}(X_j = 1 | X_i = 1) \;\leq\; 2w|I|p. \end{aligned} \tag{4.4}$$

Finally, the Q_1 and Q_2 terms approach 0 if we assume $w|I|p \to 0$ as $|I| \to \infty$. For that, we impose p to converge to 0 faster than $w|I|$ converges to infinity as $|I| \to \infty$.

Under these conditions, the total variation distance

$$d_{TV}(\mathcal{L}(Y),\mathcal{P}(Z)) \;\leq\; \frac{1-e^{-\lambda}}{\lambda}\left((2w+1)|I|p^2+2w|I|p\right).$$

tends to 0 and Y approaches a Poisson random variable with mean λ.

Computation of probabilities of false alarm. In our multiscale modeling, we assume local dependencies between two adjacent scales. Therefore, we capture all dependencies in the set $\aleph_\ell=\{\ell-1,\ell,\ell+1\}$ (i.e. $w=1$) with $\ell=\{1,\cdots,L\}$. Under the aforementioned conditions $Lp\to 0$ as $L\to\infty$ with $\mathbf{P}(D_\ell(\mathbf{x})=1)=p$ and $\mathbf{P}(D_\ell(\mathbf{x})=0)=1-p,\forall\ell\in\{1,\cdots,L\}$, the probability of false alarm denoted as $P_{FA}(\mathbf{x},L)$ at location $\mathbf{x}$ is given by the Poisson tail with parameter $\lambda_T(\mathbf{x})=\sum_{\ell=1}^{L}P_{fa,\ell}(\mathbf{x},T)$:

$$P_{FA}(\mathbf{x},L) \;\triangleq\; \mathbf{P}\left(\sum_{\ell=1}^{L}D_\ell(\mathbf{x})>k_D(\mathbf{x})|H_0\right) \;=\; 1-\sum_{k=0}^{k_D(\mathbf{x})}\left(\lambda_T(\mathbf{x})\right)^k\frac{e^{-\lambda_T(\mathbf{x})}}{k!}, \tag{4.5}$$

where $(k_D(\mathbf{x})+1)$ is the actual number of changes (positive decisions) detected for the different patch sizes at location $\mathbf{x}$: $0\leq k_D(\mathbf{x})\leq L-1$.

Now, we focus on the computation of the probability of false alarm $P_{FA}(\mathbf{x},L)$ for application on real images. From Lemma 3.2, the probability $P_{fa,\ell}(\mathbf{x},N)$ can be bounded at each pixel as (if $T=N$)

$$P_{fa,\ell}(\mathbf{x},N) \;\leq\; e^{-N}\mathbf{E}\left[e^{S_{N,\ell}(\mathbf{x})}|H_0\right], \tag{4.6}$$

where $S_{N,\ell}(\mathbf{x})$ is the total number of positive decisions obtained from the pixels in the search window for a given patch size $n_\ell, 1\leq\ell\leq L$. In our change detection scenario, we need also to make a pointwise decision from spatial contexts given the entire image. This can be actually achieved by considering very large spatial neighborhoods in order to both capture long range information and to detect unusual events when comparing two images. In the special case when $\lambda_N(\mathbf{x}) \;\equiv\; \lambda_N=\sum_{\ell=1}^{L}P_{fa,\ell}(N)$ is constant as adopted in our experiments, we consider the upper-bound (3.2) to get:

$$\left\{\begin{array}{rcl} P_{FA}(\mathbf{x},L) & = & 1-\displaystyle\sum_{k=0}^{k_D(\mathbf{x})}\left(\lambda_N\right)^k\frac{e^{-\lambda_N}}{k!}, \\ \lambda_N & = & \displaystyle\sum_{\ell=1}^{L}e^{-N}\mathbf{E}\left[e^{S_{N,\ell}(\mathbf{x})}|H_0\right] \;\approx\; \sum_{\ell=1}^{L}\frac{e^{-N}}{|\Omega_{0,\ell}|}\sum_{\mathbf{y}\in\Omega_{0,\ell}}e^{S_{N,\ell}(\mathbf{y})}, \end{array}\right. \tag{4.7}$$

where $\Omega_{0,\ell}=\{\mathbf{x}\in\Omega : S_{N,\ell}(\mathbf{x})<N\}$. Finally, we consider that a change occurs at pixel if

$$P_{FA}(\mathbf{x},L) \;\leq\; \alpha(\mathbf{x}), \tag{4.8}$$

for a desired level of significance $\alpha(\mathbf{x})$. The final goal is to make a decision by setting a probability of false alarm $\alpha(\mathbf{x})$ with $0<\alpha(\mathbf{x})<1$. Actually, we are not in position to assume that the detected pixels are independent and it would be more suitable to compare the number of detected pixels to the expectation of this number as recommended in [36, 66, 100, 87, 37]. This may be possible if we examine the number of detection tests performed and by applying the Bonferroni strategy for multiple tests.

This amounts to setting $\alpha(\mathbf{x}) = \varepsilon/|\Omega|$ where $|\Omega|$ is the number of tested pixels and ε is a user-defined expected number of false alarms over the entire image. We can refer to the *a contrario* framework yielding the same control as Bonferroni while allowing one to set $\varepsilon \geq 1$ if needed. Setting $\varepsilon = 1$ as we do in most experiments (see also [66, 100, 27, 37]), means that about 1 pixel on average is falsely detected but the remaining detections are "meaningful". It is a sound choice as the number of false alarms generally has an exponential behavior with respect to event properties so the dependence on ε is rather a log-dependence [36].

Let $H_\varepsilon(u,v) : \Omega \to \{0,1\}$ be the final multiscale change detection map defined as:

$$H_\varepsilon(u,v)(\mathbf{x}) = \begin{cases} 1 & \text{if } P_{FA}(\mathbf{x}, L) \leq \varepsilon/|\Omega|, \\ 0 & \text{otherwise.} \end{cases} \tag{4.9}$$

PROPOSITION 4.2. *The expected number of false alarms in Ω is lower than ε.*
Proof.

$$\begin{aligned} \mathbf{E}\left[\sum_{\mathbf{x}\in\Omega} H_\varepsilon(u,v)(\mathbf{x})|H_0\right] &= \sum_{\mathbf{x}\in\Omega} \mathbf{E}[H_\varepsilon(u,v)(\mathbf{x})|H_0], \\ &= \sum_{\mathbf{x}\in\Omega} \mathbf{P}(H_\varepsilon(u,v)(\mathbf{x}) = 1|H_0), \\ &= \sum_{\mathbf{x}\in\Omega} \mathbf{P}\left(\sum_{\ell=1}^{L} D_\ell(\mathbf{x}) > k_D(\mathbf{x})|H_0\right), \\ &= \sum_{\mathbf{x}\in\Omega} P_{FA}(\mathbf{x}, L) \;\leq\; \sum_{\mathbf{x}\in\Omega} \frac{\varepsilon}{|\Omega|} \;\leq\; \varepsilon. \end{aligned}$$

□

In practice, we can avoid the setting of the value ε provided that a meaningful change is expected to occur. This means we can derive a minimal value ε_{min} to detect at least one occluded pixel in the image:

PROPOSITION 4.3. *No change detection occurs in the image pair if*

$$\varepsilon \;\geq\; \varepsilon_{min} \;\triangleq\; |\Omega|\left(1 - \sum_{k=0}^{L-1} (\lambda_N)^k \frac{e^{-\lambda_N}}{k!}\right).$$

Proof. By definition, a meaningful change is detected if

$$\frac{\varepsilon}{|\Omega|} \;\geq\; 1 - \sum_{k=0}^{k_D(\mathbf{x})} (\lambda_N)^k \frac{e^{-\lambda_N}}{k!} \;\geq\; 1 - \sum_{k=0}^{L-1} (\lambda_N)^k \frac{e^{-\lambda_N}}{k!}.$$

and we obtain the required assertion. □

Finally, we choose to consider a uniform threshold to detect meaningful changes in the entire image as:

$$\alpha(\mathbf{x}) \;=\; \alpha \;\triangleq\; \frac{\varepsilon}{|\Omega|}. \tag{4.10}$$

As mentioned earlier, the Bonferroni correction is known to be too conservative for a detection problem under multiple i.i.d tests. The correction adjusts the threshold

for each individual test in order to satisfy a lower false alarm rate value $\varepsilon/|\Omega|$ but reduces the detection capability considerably. Although the chance of false alarm at each location $\mathbf{x}$ is only α, the chance of at least one falsely alarmed pixel is much higher since the neighboring pixels are spatially correlated. Using the False Discovery Rate (FDR) approach [12], we could expect to improve the detection results while maintaining a given false discovery rate. FDR is defined as the expected ratio of the number of observations falsely classified into alternate hypotheses to the total number of observations classified into the alternate hypotheses. When the total of observations come from the null hypothesis as assumed in the paper, FDR is proved to be equivalent to the family-wise error rate (FWER). In all other cases, FWER is bounded below by FDR. In the context of our change detection problem, considering only the probability of false alarm yielded satisfactory detection results as shown in Section 6 and we did not investigate further FDR for the time being.

5. Algorithm and implementation. In this section, we describe the multiscale change detection procedure.

5.1. Estimation of spatially-varying thresholds. To apply the algorithm summarized in Fig. 5.2, we need to address the estimation problem of $\tau(\mathbf{x})$ already introduced in Section 3.1.

To compute the spatially-varying thresholds $\tau(\mathbf{x})$, we propose here to adopt a non-parametric approach to capture the variability sources related to spatial contexts. *Our idea is to estimate adaptive detection thresholds for each individual pixel and from a unique reference image.*

To derive the adaptive thresholds for change detection, we postulate that all positive decisions correspond to a distance $\phi_{uv}(\mathbf{x},\mathbf{y})$ higher than the highest distance at pixel $\mathbf{x}$ computed from the *reference image* u: $\tau(\mathbf{x}) \triangleq \tau_u(\mathbf{x})$. Note that considering only one training image for change detection has been already suggested in [51] but a mixture of Gaussians was necessary to derive a unique decision threshold for the entire image. We define

$$\begin{aligned} \tau_u(\mathbf{x}) &= \max\left(\sup_{\mathbf{y}\in b(\mathbf{x})} \phi_{uu}(\mathbf{x},\mathbf{y}),\ \bar{\tau}\right), \\ \bar{\tau} &= \frac{1}{|\Omega|}\sum_{\mathbf{x}\in\Omega} \inf_{\mathbf{y}\in b(\mathbf{x})} \phi_{uu}(\mathbf{x},\mathbf{y}), \end{aligned} \tag{5.1}$$

where $b(\mathbf{x})$ is a small neighborhood around $\mathbf{x}$ (e.g. 3×3 square windows or balls of $\mathbb{R}^2$ of radius $r_b = 1$ pixel, i.e. set of pixels $\mathbf{y} \in b(\mathbf{x})$ such that $\|\mathbf{y}-\mathbf{x}\|_2^2 \leq r_b^2$). By introducing a minimal value $\bar{\tau}$ in (5.1) defined as the average of the lowest distance computed over the image domain Ω, we improve the robustness to low signal-to-noise ratios. In this learning procedure, the nearby patches taken in $b(\mathbf{x})$ around $\mathbf{x}$ are assumed to correspond to perturbed configurations of the central patch located at pixel $\mathbf{x}$. By examining only the local neighborhoods in the reference image u, we derive minimal adaptive thresholds (5.1), yielding to robust change detection results as we shall see in our experiments (Section 7).

Intuitively, the background is assumed to be nearly static with residual motions due to camera instability, non-constant camera exposure, other sources of measurement noise or other irrelevant background dynamics, all described by stochastic processes. In order to cope with such random local changes, the threshold $\tau_u(\mathbf{x})$ is defined as the higher distance in the neighborhood since motion cannot be predicted.

Formally, let $\mathbf{s}(\mathbf{x}) \in \mathbb{R}^2$ be a local random shift (or displacement) vector such that $\mathbf{E}[\|\mathbf{s}(\mathbf{x})\|_2] = 0$ and $\mathbf{E}[\|\mathbf{s}(\mathbf{x})\|_2^2] = \sigma_{\mathbf{s}}^2$. The "jittering" noise can be described by a stochastic process: a patch at random position $\mathbf{y} \in b(\mathbf{x})$ defined as $\mathbf{y} = \mathbf{x} + \mathbf{s}(\mathbf{x})$ is expected to be moved to the position $\mathbf{x}$ in the second image v if no meaningful change occurs [48]. The size of the local neighborhood $b(\mathbf{x})$ is related to the amplitude of camera jitter and/or background dynamics. Also, we assume that the patches at location $\mathbf{x}$ and location $\mathbf{y}$ for some $\mathbf{y} \in b(\mathbf{x})$ are similar if we use a single image u to derive a minimal detection threshold at location $\mathbf{x}$. Finally, the probability that the amplitude $\|\mathbf{s}(\mathbf{x})\|_2$ of the background motion is higher than the radius r_b of local neighborhoods is upper-bounded and decreases as $\tau_u(\mathbf{x})$ increases if the functions u and v have a minimal regularity (K-Lipschitz):

PROPOSITION 5.1. *Assume that $b(\mathbf{x})$ is a ball in $\Omega \subset \mathbb{R}^2$ with radius $r_b = \sqrt{|b(\mathbf{x})|/\pi}$ and $u \in \mathbb{R}^\Omega$ is K-Lipschitz, namely $\exists K \geq 0$ such that $|u(\mathbf{x}) - u(\mathbf{y})| \leq K\|\mathbf{x} - \mathbf{y}\|_2$. Then,*

$$\mathbf{P}\left(\|\mathbf{s}(\mathbf{x})\|_2 \geq \sqrt{\frac{|b(\mathbf{x})|}{\pi}}\right) \;\leq\; \frac{nK\sigma_{\mathbf{s}}^2}{\tau_u(\mathbf{x})} \;\leq\; \frac{nK\sigma_{\mathbf{s}}^2}{\bar{\tau}}.$$

Proof. Assume $\exists \kappa > 0$ such that $r_b = \kappa\sigma_{\mathbf{s}}$. From the Chebyshev's inequality, we have

$$\mathbf{P}\left(\|\mathbf{s}(\mathbf{x})\|_2 \geq \kappa\sigma_{\mathbf{s}}\right) \;\leq\; \frac{1}{\kappa^2} \;=\; \frac{\sigma_{\mathbf{s}}^2}{r_b^2}. \tag{5.2}$$

Consider the usual Euclidean distance to evaluate the dissimilarity between image patches and assume $u \in \mathbb{R}^\Omega$ is K-Lipschitz, that is $\exists K \geq 0$ such that $|u(\mathbf{x}) - u(\mathbf{y})| \leq K\|\mathbf{x} - \mathbf{y}\|_2$. From (5.1), it follows

$$\begin{aligned} \tau_u(\mathbf{x}) \;&=\; \max\left(\sup_{\mathbf{y}\in b(\mathbf{x})} \sum_{\mathbf{t}\in\mathbb{R}^2} \mathrm{rect}_n(\mathbf{t})\,(u(\mathbf{x}+\mathbf{t}) - u(\mathbf{y}+\mathbf{t}))^2, \bar{\tau}\right), \\ &\leq \max\left(\sup_{\mathbf{y}\in b(\mathbf{x})} K^2 n\|\mathbf{x}-\mathbf{y}\|_2^2, \bar{\tau}\right), \\ &\leq K^2 n r_b^2. \end{aligned} \tag{5.3}$$

From (5.2) and (5.3), we get

$$\mathbf{P}\left(\|\mathbf{s}(\mathbf{x})\|_2 \geq r_b\right) \;\leq\; \frac{\sigma_{\mathbf{s}}^2}{r_b^2} \;\leq\; \frac{nK^2\sigma_{\mathbf{s}}^2}{\tau_u(\mathbf{x})} \;\leq\; \frac{nK^2\sigma_{\mathbf{s}}^2}{\bar{\tau}}.$$

□

The size of the local neighborhood $b(\mathbf{x})$ corresponds to the expected "jittering" amplitude or, in image sequence analysis, to the movement amplitude of animated texture in the background. Note that if $r_b \to 0$ then $\tau_u(\mathbf{x}) \to 0$ and a change is likely detected at location $\mathbf{x}$ in the image. Finally, the thresholds are computed directly from image data and the computation of the noise variance is not required since it is assumed to be a constant value in the vote procedure.

In other respects, the method we have described will not produce the same detection results if we compare u to v and vice-versa, mainly because the thresholds are estimated either from u or v. It may be desirable in some applications to get the same

answer in both cases. This can be achieved by combining the definitions of thresholds and dissimilarity measures from u and v as follows:

$$\begin{array}{lcl} \tau(\mathbf{x}) & = & \min\left(\tau_u(\mathbf{x}), \tau_v(\mathbf{x})\right), \\ \phi_{uv}(\mathbf{x},\mathbf{y}) & \stackrel{\triangle}{=} & \min\left(\phi_{uv}(\mathbf{x},\mathbf{y}), \phi_{vu}(\mathbf{x},\mathbf{y})\right). \end{array} \tag{5.4}$$

We present in Section 7.1 (see Fig. 7.9) an experiment with synthetic images to demonstrate the interest of this modeling.

5.2. Implementation. Our final algorithm is presented in Fig. 5.2 and can serve in different applications. Note that the size N of the search window depends on the motion amplitude (due to camera "jittering" or residual motion in the background) we do not want to detect. Since, we focus on the change and occlusion detection problem, we set $N = 3 \times 3$ (5×5 at most) (or circular windows with radius 1 pixel) in the applications with still cameras. This means we are testing dissimilarities in very local neighborhoods. Accordingly, we can choose $|b(\mathbf{x})| \equiv |B(\mathbf{x})|$, also fixed whatever the considered patch size is, which limits the number of algorithm parameters. Increasing the search window size should produce the same results since the missing or occluded patches are not visible in the second image for any location. Nevertheless, undesirable but similar patterns can be found if there is repeated structure or texture in the background. We can use Proposition 4.3 to choose ε since we guarantee that no meaningful detection occurs in the image pairs if ε is higher than ε_{min}. Finally, we evaluate the performance of the proposed algorithm in section 7.

Typically, the computing time are of 71.79 seconds for a 360×288 image shown in Figs. 7.1-7.2 ($L = 11, N = 9$) with a C++ implementation (with no code optimization) on an Intel Core 17 64 bits, 2,66 Ghz. The timings for several scales are given in Fig. 5.2. The complexity increases significantly with the number L of scales since the patch size is defined as $N_\ell = (2\ell + 1)^2$ at a given scale $1 \leq \ell \leq L$.

6. Invariances. In this section, several invariance properties of our detector are given. We consider the Euclidean dissimilarity measure (2.5).

Shift and linear invariance: Let $s(u) = c_0 + c_1 u$ be a linear contrast change. It follows that, $\forall c_0, c_1 \in \mathbb{R}$ and $\forall (u, v) \in \mathbb{R}^\Omega \times \mathbb{R}^\Omega$, we have[1]

$$H_\varepsilon(u, v) = H_\varepsilon(s(u), s(v)).$$

Translation invariance: Assume that Ω is a torus and let $\mathbf{s} \in \mathbb{R}^2$. If $\mathrm{T}_\mathbf{s}$ is a translation operator defined as $\mathrm{T}_\mathbf{s} \circ u(\mathbf{x}) = u(\mathbf{x} - \mathbf{s})$ for all $u \in \mathbb{R}^\Omega$, the change detector satisfies:

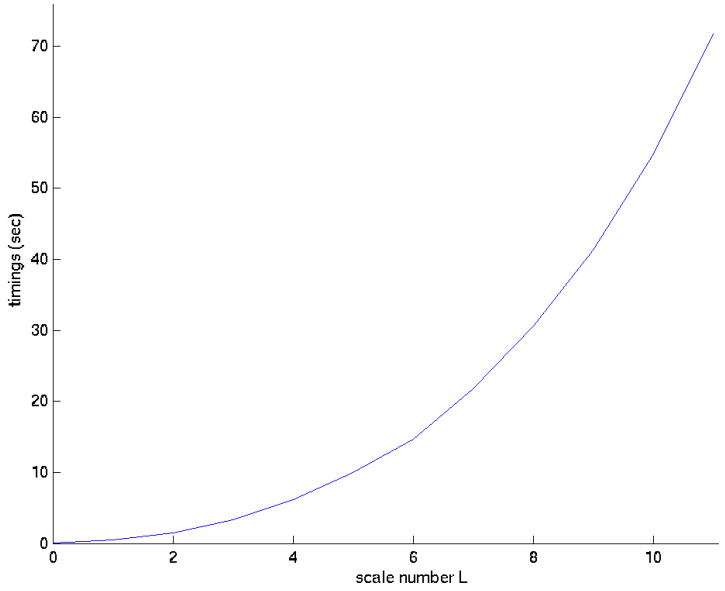
$$H_\varepsilon(T_\mathbf{s} \circ u, T_\mathbf{s} \circ v) = \mathrm{T}_\mathbf{s} \circ H_\varepsilon(u, v).$$

$\pi/2$ *rotation invariance:* If $\mathrm{R}_{\pi/2}$ is a $\pi/2$ rotation sending Ω onto itself, then, for all $(u, v) \in \mathbb{R}^\Omega \times \mathbb{R}^\Omega$, we have:

$$H_\varepsilon(\mathrm{R}_{\pi/2} \circ u, \mathrm{R}_{\pi/2} \circ v) = \mathrm{R}_{\pi/2} \circ H_\varepsilon(u, v).$$

[1] *Sketch of proof:* Assume with loss of generality $\bar{\tau} = 0$. Hence, we have

$$S_N(\mathbf{x}) = \sum_{\mathbf{y} \in B(\mathbf{x})} \mathbb{1}\left[\phi_{uv}(\mathbf{x},\mathbf{y}) \geq \sup_{\mathbf{y} \in b(\mathbf{x})} \phi_{uu}(\mathbf{x},\mathbf{y})\right] = \sum_{\mathbf{y} \in B(\mathbf{x})} \mathbb{1}\left[\phi_{s(u)s(v)}(\mathbf{x},\mathbf{y}) \geq \sup_{\mathbf{y} \in b(\mathbf{x})} \phi_{s(u)s(u)}(\mathbf{x},\mathbf{y})\right].$$



scale number L	1	2	3	4	5	6	7	8	9	10	11
timings (sec)	0.45	1.44	3.29	6.22	9.95	14.70	21.77	30.65	41.28	54.84	71.79

FIG. 5.1. *Timings wrt scale number for computing detection maps (image pair shown in Figs. 7.1-7.2)*

If we consider circular search windows (not used in practice) and $W_n(\cdot) = G_n(\cdot)$, the detector is invariant to any rotation of angle $\nu \in [0, 2\pi]$:

$$H_\varepsilon(\mathrm{R}_\nu \circ u, \mathrm{R}_\nu \circ v) = \mathrm{R}_\nu \circ H_\varepsilon(u, v).$$

Symmetry invariance: It may be also desirable to provide the same change detection results when u is compared to v and vice-versa. The algorithm given in Fig. 5.2 is then modified using (5.4) as explained in Section 5.2 to get this invariance.

Invariance to illumination changes: By considering the set of distances given in Section 3.1, invariance to illumination changes can be achieved.

Because of the complex nature of the algorithm, the theoretical property of *scale invariance* is extremely difficult to obtain. This could be investigated in future work since it is an important issue in computer vision and desirable for shape recognition. Nevertheless, our method considers a multiscale framework to avoid the precise setting of n, which is known to be related to the scale of noise and texture in the image. In video-surveillance applications, the objects may also have different sizes. Therefore, the multiscale framework is suitable to analyze the robustness of the decision rules at various scales while slightly increasing the computation time.

7. Experimental results. To evaluate our multiscale method, we conducted experiments on a variety of image pairs and applications including video surveillance and blotch detection in old digitized movies. In this section, we present results on

Asymmetric change detection algorithm

Let L be the number of patch scales, ϕ_{uv} the dissimilarity measure used for patch comparison and ρ the smoothing parameter for brightness invariance (see (2.2) and (2.3)).

- For $\ell = 1 \cdots L$
  1. For each pixel $\mathbf{x} \in \Omega$ and 3×3 (or 5×5) neighborhoods $b(\mathbf{x})$, compute
  $$\tau_{u,\ell}(\mathbf{x}) \;=\; \max\left(\sup_{\mathbf{y}\in b(\mathbf{x})} \phi_{uu,\ell}(\mathbf{x},\mathbf{y}),\; \bar{\tau}_\ell\right),$$
  where $\tau_{u,\ell}(\mathbf{x})$, $\bar{\tau}_\ell$ and $\phi_{uu,\ell}(\mathbf{x},\mathbf{y})$ are respectively the thresholds $\tau_u(\mathbf{x})$ and $\bar{\tau}$ and the dissimilarity measure $\phi_{uu}(\mathbf{x},\mathbf{y})$ computed for a given patch size $n_\ell = (2\ell+1)^2$.
  2. For each pixel $\mathbf{x} \in \Omega$ and 3×3 (or 5×5) search window $B(\mathbf{x})$, compute
  $$S_{N,\ell}(\mathbf{x}) \;=\; \sum_{\mathbf{y}\in B(\mathbf{x})} \mathbb{1}[\phi_{uu,\ell}(\mathbf{x},\mathbf{y}) \geq \tau_\ell(\mathbf{x})].$$
  3. Compute $P_{fa,\ell}(N) \;=\; \dfrac{1}{|\Omega|}\displaystyle\sum_{\mathbf{y}\in\Omega} e^{S_{N,\ell}(\mathbf{y})-N}$ and decide that a change occurs at pixel $\mathbf{x}$, for a given patch size l, if $S_{N,l}(\mathbf{x}) = N$.
- Compute $\lambda_N \;=\; \displaystyle\sum_{\ell=1}^{L} P_{fa,\ell}(N)$.
- The final decision at pixel $\mathbf{x} \in \Omega$ is defined as:
$$H_\varepsilon(u,v)(\mathbf{x}) \;=\; \mathbb{1}[P_{FA}(\mathbf{x},L) \leq \alpha],$$
with $\alpha = \varepsilon/|\Omega|$ and $P_{FA}(\mathbf{x},L) \;=\; 1 - \displaystyle\sum_{k=0}^{k_D(\mathbf{x})} (\lambda_N)^k \,\frac{e^{-\lambda_N}}{k!}$ where $0 \leq k_D(\mathbf{x}) \leq L-1$ is the actual number of changes detected at location $\mathbf{x}$ for the set of L patch sizes.

FIG. 5.2. *Asymmetric change detection algorithm between images u and v.*

different image pairs with illumination and motion variations. We compare also our results to those produced by algorithms of the state-of-the-art for several image pairs. In all our experiments, the parameters are those given in Fig. 5.2. We used the Euclidean dissimilarity measure (2.5) in most experiments. In the case of illumination changes, we considered the dissimilarity measure (2.3) and (2.4). Moreover, the size of neighborhoods $b(\mathbf{x})$ and the size of search windows $B(\mathbf{x})$ are the same (i.e. $B(\mathbf{x}) \equiv b(\mathbf{x})$). In all experiments, we have chosen $W_n(\mathbf{t}) = G_n(\mathbf{t}) \triangleq e^{-\|\mathbf{t}\|_2^2/n}$ to delineate better the foreground object borders. For display purposes only, the change detection masks are superimposed (in yellow) on the first original image. The probabilities of false alarms are also shown for several tested image pairs.

7.1. Evaluation of robustness on real images. In the two first examples, our multiscale method is applied to outdoor and traffic scenes respectively. In Fig. 1.1, we considered $|b(\mathbf{x})| = |B(\mathbf{x})| = 5 \times 5$ ($N = 25$) because of moderate but meaningful motion in the animated background. The size of Gaussian patches ranges from $n = 3 \times 3$ to $n = 81 \times 81$ since the area of the missing object is relatively large ($L = 40$) and we found $\hat{\lambda}_N = 16.78$. On this example, the quality of the detection mask can be

FIG. 7.1. *Example of an image pair.*

visually assessed by comparison with the ground truth [99]. The method is robust to undesirable motions in the background corresponding to trees shaken by the wind. In Figs. 7.1-7.2, 3×3 neighborhoods and 3×3 search windows were used ($N = 9$). Since the background represents a large part of the image, we set $L = 11$ and we found $\hat{\lambda}_N = 0.88$. We also examined the detection maps and the counting values $S_N(\mathbf{x})$ for different and arbitrary patch sizes. Low count numbers are labeled with cold colors and high count numbers are labeled with hot colors. In these two typical examples, the masks are quite regularized with no hole because of the sliding window process and patch overlapping. Figure 7.2 shows that the object boundaries are better delineated if we consider a set of patches with different sizes instead of a single arbitrary patch size. In Figs. 7.3-7.4, other examples of change detection are shown for video-surveillance scenarios for which the illumination conditions are relatively stable.

In Fig. 7.5, the image pair is composed of two consecutive frames of an old movie. We used a set of small Gaussian patches ($L = 3$ pixels) to detect all the blotches known to suddenly appearing in the image at random locations [13, 97, 25]. These artifacts correspond typically to bright or dark small regions caused by dirt (on the positive or on the negative). The loss of gelatin covering the film caused by mishandling or aging of the film is the second well-known degradation process. Generally, the spatial coherence of blotches means that the neighboring pixels are corrupted as well but the temporal coherence is not preserved. In this experiment, we considered small neighborhoods (3×3 square windows) and we found $\hat{\lambda}_N = 0.050$ for the tested image pair. We analyzed the images in both directions and the algorithm given in Fig. 5.2 has been then modified accordingly using (5.4). These detected regions can be repaired further by inpainting methods (e.g. [39]) if successfully detected. Too many alarms is not desirable in this application since inpainting methods may fail.

Robustness to low signal-to-noise ratios. To demonstrate the robustness to low signal-to-noise ratios, we evaluated the algorithm by applying different levels of noise (with added artificial white Gaussian noise) on two image pairs. Figure 7.6 shows that the method (with the same parameters) produces similar detection masks even if the signal-to-noise ratio is very low ($\sigma = 30$). No undesirable connected component is detected.

On real noisy images, the difficulty is that surveillance cameras normally capture small non-rigid figures, such as walking persons or moving cars, on low contrast and low resolution formats. Typically, Fig. 7.7 shows a snowy traffic scene with low contrast. We applied the usual algorithm to detect the car in the back-

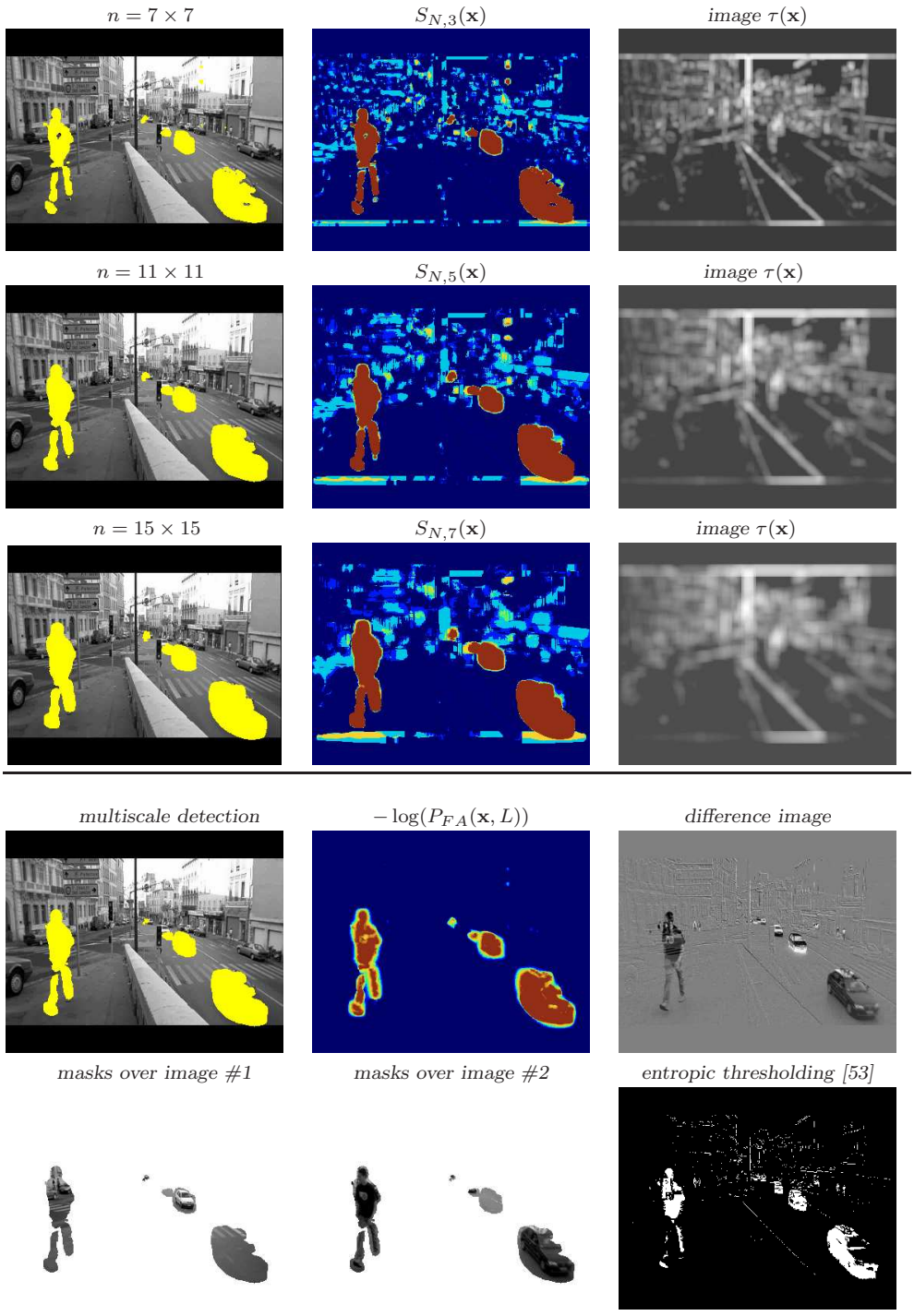


FIG. 7.2. *Change detection in the image pair shown in Fig.7.1. The images $S_{N,\ell}(\mathbf{x})$ correspond to the number of positive decisions in 3×3 search windows $B(\mathbf{x})$ for several patch sizes and scales: $n_3 = 7 \times 7$, $n_5 = 11 \times 11$ and $n_7 = 15 \times 15$ respectively. The level sets corresponding to $T = N$ are superimposed in yellow on the original image for each patch size (left column). The last row shows the detection results we obtained by using a multiscale representation when the range of patches varies from $n = 3 \times 3$ to $n = 23 \times 23$ pixels ($L = 11$). For comparison, the histogram of the absolute difference image was thresholded with the Kapur's entropic method [53].*

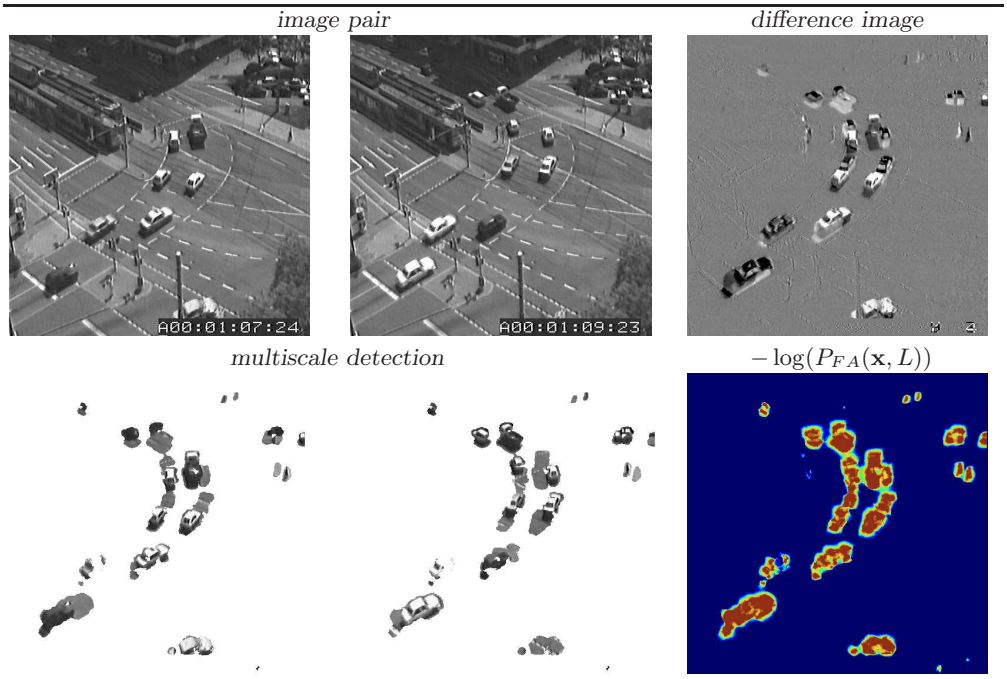


FIG. 7.3. *Occlusion detection in a traffic scene (images at time $t = 1$ and $t = 50$) using 5×5 search windows ($L = 12$) (see* `http://i21www.ira.uka.de/image_sequences/`*).*

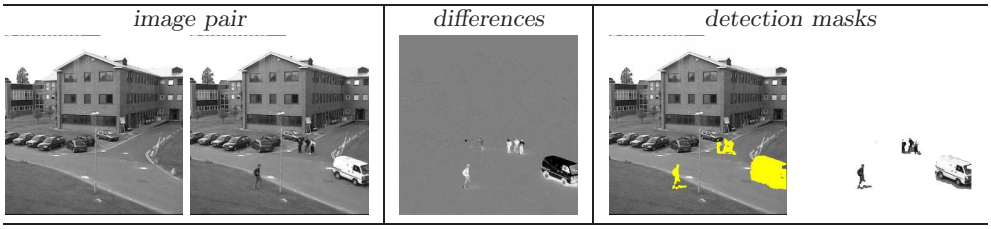


FIG. 7.4. *Change detection results by using the L_2 dissimilarity measure (2.5): $L = 3, \hat{\lambda}_N = 0.13$ (PETS 2006 dataset).*

ground. Unfortunately, a large number of undesirable regions are also detected due to non-stationarities in the animated background for several search windows. To overcome this problem, we modified the algorithm by simply considering a unique threshold defined as: $\tau = \sup_{\mathbf{x} \in \Omega} \tau(\mathbf{x})$. Because the statistics of temporal and spatial noises are not the same in the image as assumed in Section 3.2, we proposed successfully this simple modification. In Fig. 7.7, the expected object is now correctly identified by our multiscale method and no additional region and hole is extracted ($N = 9, L = 15, \hat{\lambda}_N = 0.44$).

Robustness to illumination variation. On the example shown in Fig. 7.8, we compared two photographs of the same scene at two different dates in the year (summer and spring). The structural changes are better estimated using the dissimilarity measure (2.4) (correlation coefficient) which is parameter free, since illumination

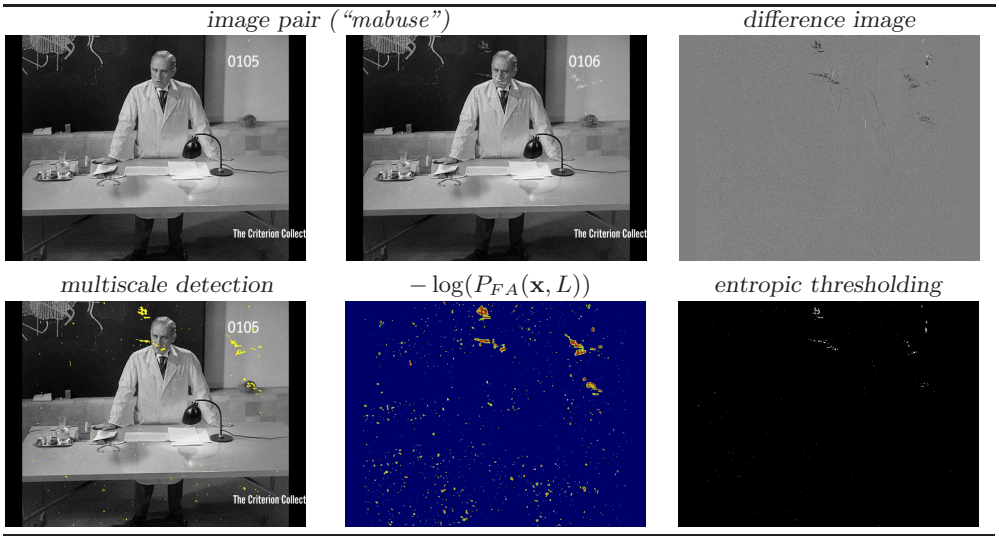


FIG. 7.5. *Blotch detection in an image pair from an old movie ($L = 3$). The histogram of the absolute difference image was thresholded with the Kapur's entropic method [53].*

changes are irrelevant for scene understanding. The expected changes correspond to moving objects (e.g. cars) and still objects with appearance changes according to seasons (e.g. trees).

Sharp edge detection and accuracy. In this section, we address the problem of occlusions in the case of high contrasts and sharp edges. In the neighborhood of an edge, we may consider two situations: i/ the edge pixel is present on the first image and $\tau(\mathbf{x})$ is large; ii/ the edge pixel is present only in the second image and $\tau(\mathbf{x})$, estimated from the first image, is then close to 0 (e.g. homogeneous region). A typical toy example is shown in Fig. 7.9 where we consider two synthetic images with several objects in common. In this experiment, the images are corrupted with artificial Gaussian noise ($\sigma = 5$). If the spatially-varying thresholds $\tau(\mathbf{x})$ are computed from u which contains more objects, the algorithm cannot recover all the occluded pixels in the second image, especially pixels belonging to bright thin structures and straight lines. By increasing the number of scales, the locations of edges and T-junctions are not accurate and the detection results are poor. On the contrary, the detection are more satisfying if the spatially-varying thresholds are computed from v, as expected. The values of $\tau_v(\mathbf{x})$ are indeed lower in the neighborhood of occlusion edges. Meaningful occluded pixels are detected and robustness to noise is improved if we consider more patch sizes. Moreover the accuracy of object boundary locations is not altered too much if L is increased.

Finally, since it is a difficult task to choose to detect the occluded pixels from a unique reference image u or v, we recommend to adopt the (symmetric) modeling (5.2): $\tau(\mathbf{x}) = \min(\tau_u(\mathbf{x}), \tau_v(\mathbf{x}))$. As shown in Fig. 7.9 (right), the results are very similar to those obtained with v as a reference image. The error pixels (wrt "ground truth" (top right)) shown in white in Fig. 7.9 (bottom right) are due to noise and staircase edges in the original images which cannot be easily recovered with small

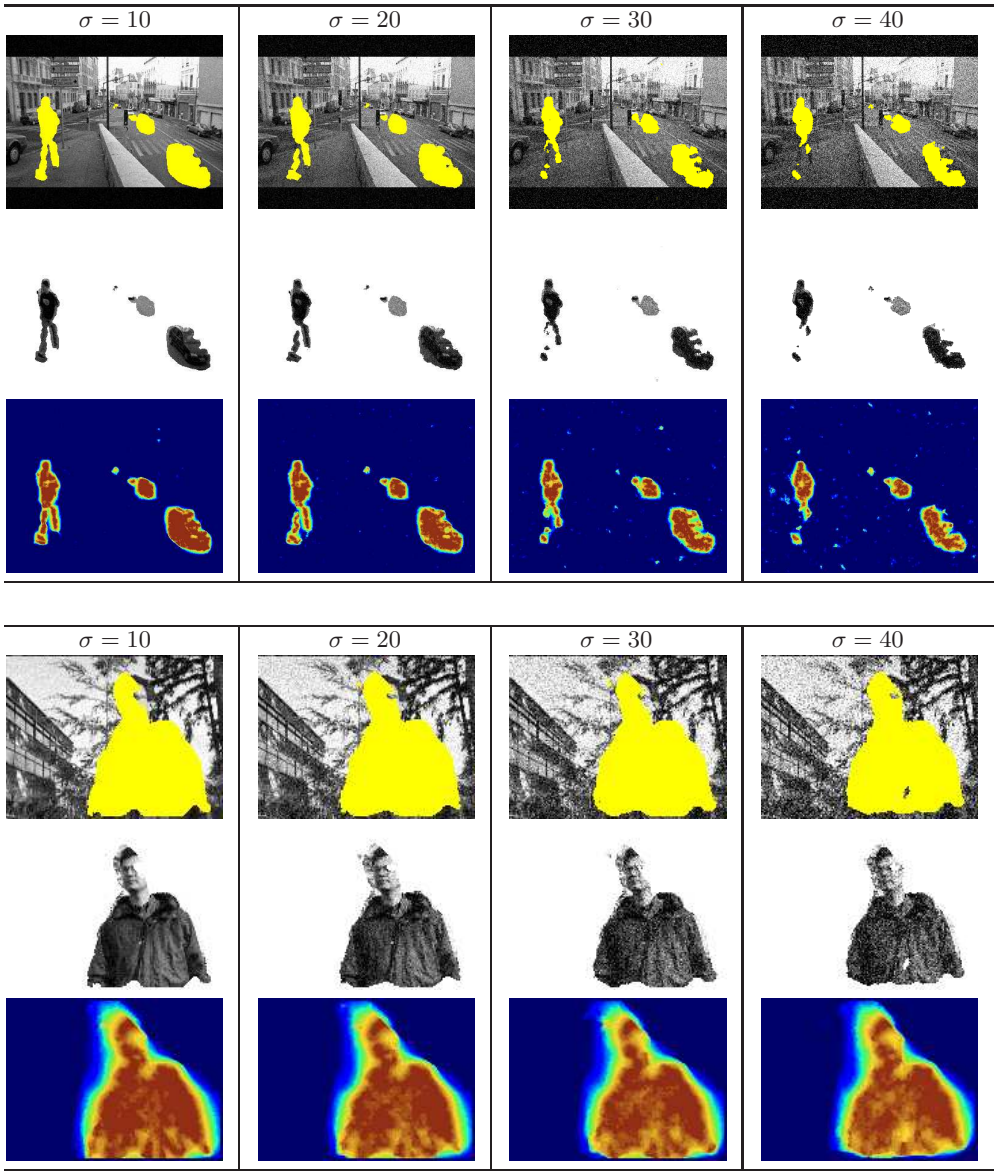


FIG. 7.6. *Robustness to white Gaussian noise on two image pairs. The two first rows show the detection results and the third row shows the probabilities of false alarm* $(-\log(P_{FA}(\mathbf{x}, L)))$.

square patches and neighborhoods.

7.2. Comparisons with state-of-the-art methods. We compare our approach to commonly-used and more recent background subtraction methods [92, 41, 107, 32, 78, 26] applied to images taken from video-surveillance image sequences. All these change detection methods exploit several frames of the sequence to determine the foreground objects unlike our multiscale detection method.

For several image pairs, we used ground-truths to evaluate the performance of

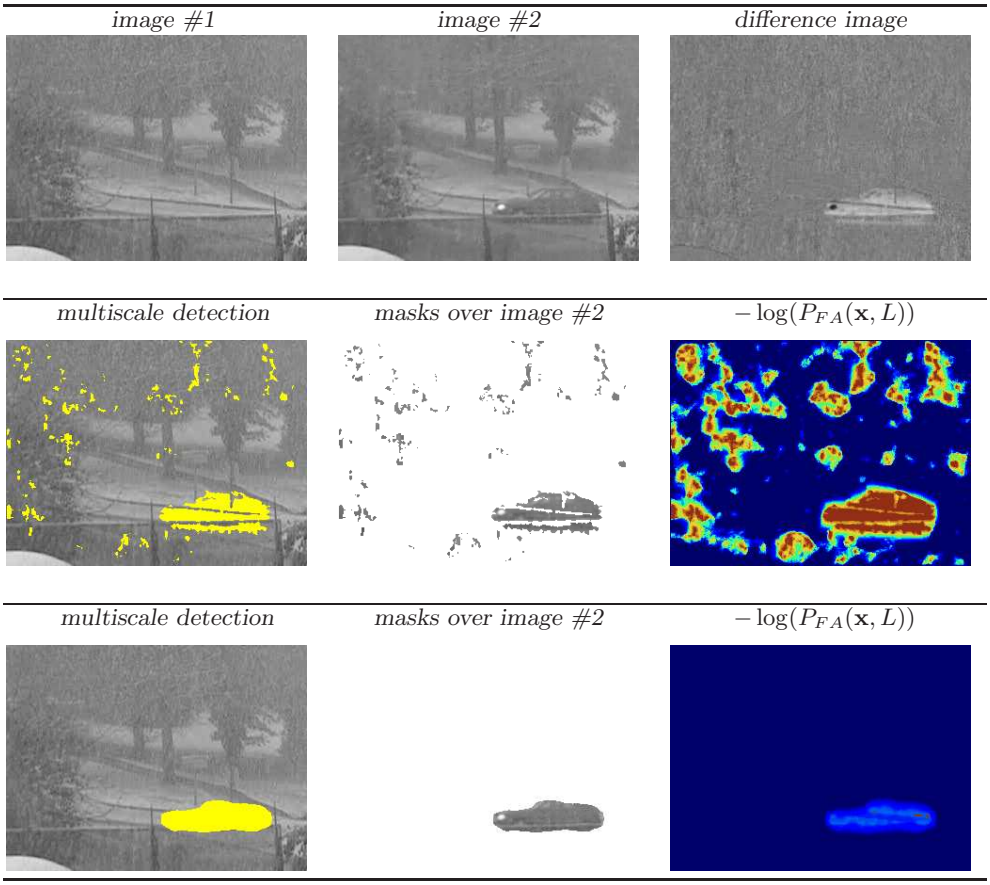


FIG. 7.7. *Car detection in a snowy traffic scene. The second row corresponds to detection performed as previously ; the third row shows the detection results by considering a unique threshold (see text).*

the algorithm. We considered the usual "RECALL" and "PRECISION" rates defined in the range $[0, 1]$ as:

$$\text{RECALL} = \frac{\text{number of pixels correctly labeled as foreground}}{\text{number of foreground pixels in the ground-truth}},$$

$$\text{PRECISION} = \frac{\text{number of pixels correctly labeled as foreground}}{\text{number of foreground pixels detected by the algorithm}}.$$

Typically, the algorithm must achieve an inherent trade-off between RECALL and PRECISION. It is worth noting that our results are compared to "RECALL" and "PRECISION" rates obtained with methods that exploit an image sequence as input.

The following experiments demonstrate first the robustness of the method to illumination changes. We tested several image pairs with the dissimilarity measures (2.3) and (2.4) to handle illumination conditions, which are not necessarily the same in the two input images. Figures 7.10 and 7.11 show two image pairs extracted from the *"light switch"* and *bootstrapping"* benchmarks described in [99]. We compared our results obtained for different standard deviations ρ involved in the dissimilarity measure (2.3) to ground truths. Our method generally produces satisfying results using

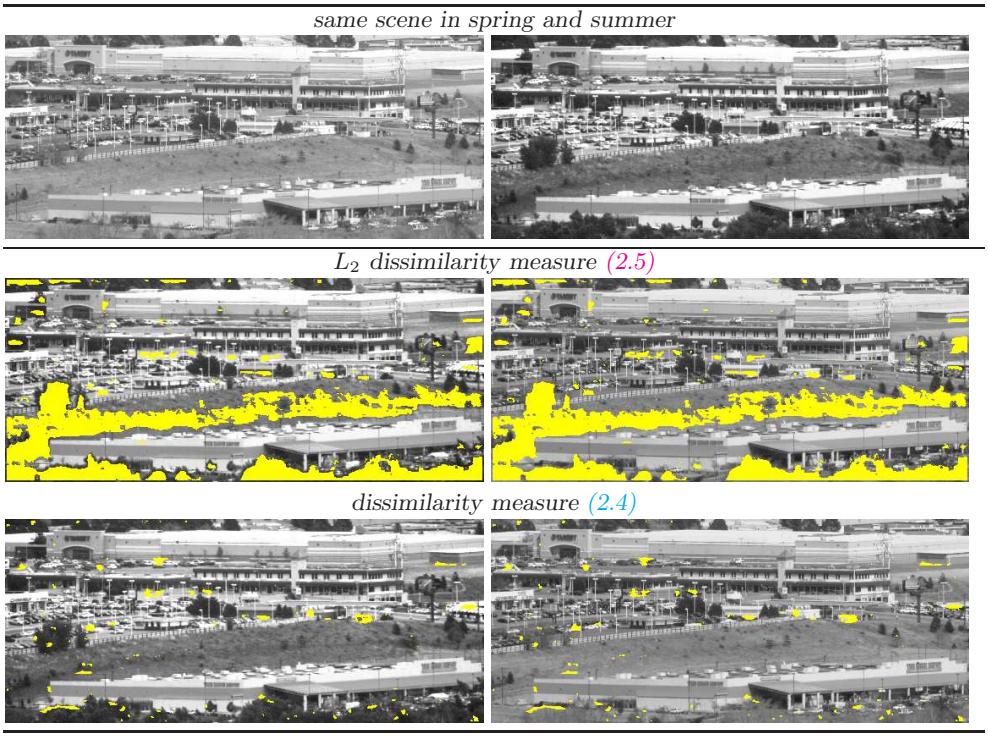


FIG. 7.8. *Change detection with several dissimilarity measures in an outdoor scene at two different dates in the year (spring and summer) (5×5 search windows and $L = 7$).*

a single reference image when compared to the state-of-the-art methods exploiting series of temporal images [99] or color information [78]. By setting $\rho = 1$, the method is relatively robust to sudden illumination (see Fig. 7.10, $N = 9, L = 10, \hat{\lambda}_N = 3.04$) and shadow effects (see Fig. 7.11, ($N = 9, L = 15, \hat{\lambda}_N = 4.69$)) but the expected number of false alarms needs to be increased ($\varepsilon = 100$ and $\varepsilon = 10$ respectively) to detect the desired regions. However, the method suffers from missed detection problem in constant areas, as it was expected. This is the case of Fig. 7.10 where a portion of the person's body is detected as unchanged. In most cases, the dissimilarity measure (2.4) (correlation coefficient) which is parameter free, enabled to produce similar results and can be attractive in many applications.

On the tested images (with ground truths), the performance of baseline techniques are known to be limited as reported in [99, 46]. The potential of our method shown in Fig. 7.12 is satisfying when compared to the state-of-the-art background subtraction methods which can handle illumination changes and moving objects [99, 41, 92, 72, 107, 32, 46, 78]. Our method did the same job using two input images and sometimes outperformed several methods [92, 41, 2, 46, 78] which require a long image sequence as input. Nevertheless, background subtraction cannot be used to detect changes between two images only (background model must be learned beforehand). Note that motion in the background (see Fig. 1.1) makes detecting changes by baseline methods very challenging (see [99]).

In Fig. 7.13 (third row), the masks of moving objects are estimated by considering respectively 50 frames and a mixture of three Gaussians [92] (third row, middle), and

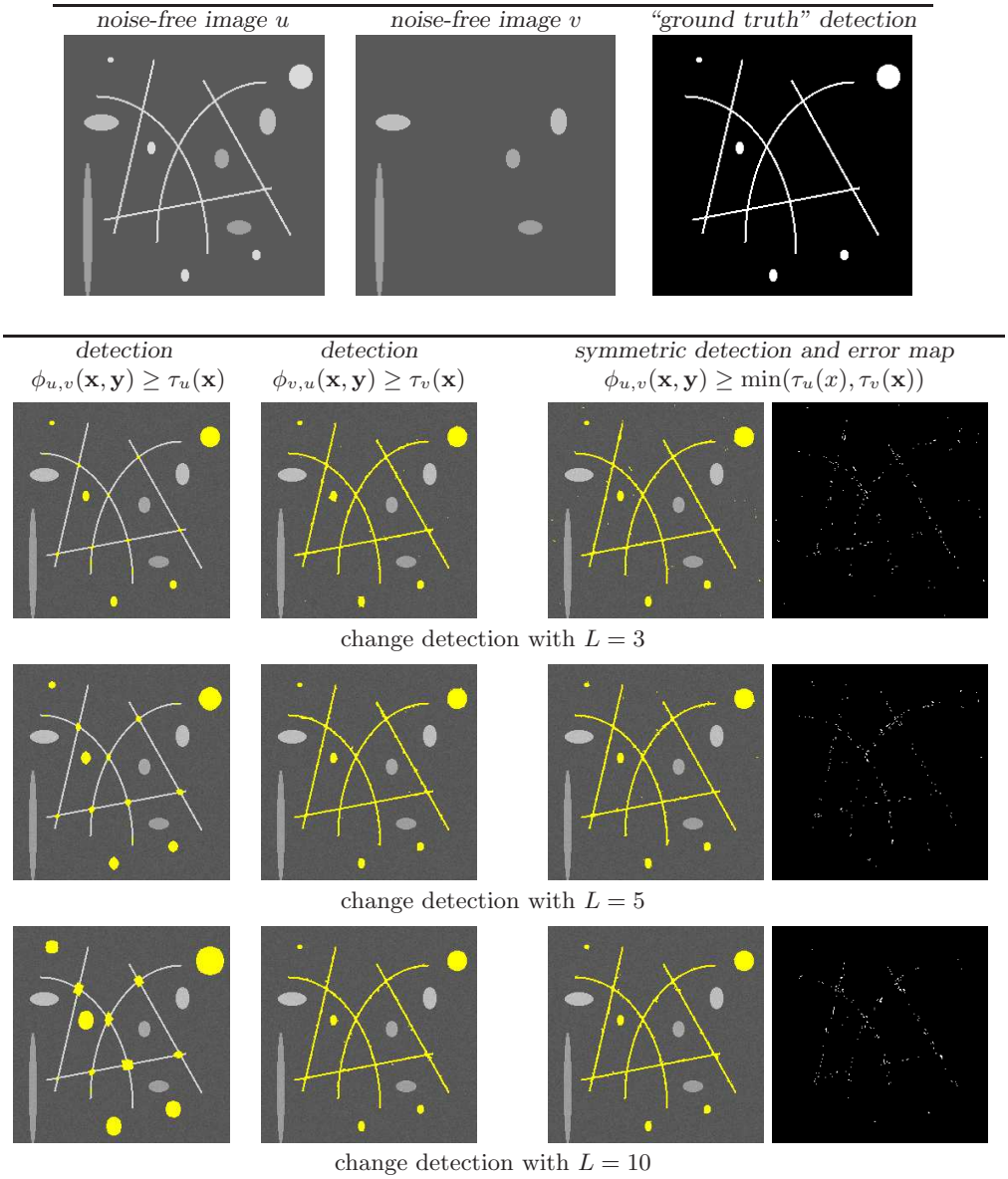


FIG. 7.9. *Change detection in a synthetic image corrupted with white Gaussian noise ($\sigma = 5$): Top: noise-free image pair and "ground truth" detection. Bottom: multiscale change detection ($L = 3, 5, 10$). The spatially-varying thresholds $\tau(\mathbf{x})$ are computed either from u (first column) or v (second column) or from (u, v) (third column) and the error pixels (symmetric detection (5.2)) are shown in white (third column).*

a long term model ($th = 10^{-6}$ (third row, left), [41]). We present also the results obtained by a recent mean shift-based clustering method [26] which exploits motion and photometry consistency. The images of the sequence are normally acquired with a fixed camera and the tested background subtraction methods provide satisfactorily results in most cases. The pedestrian in front of the water, the bike and the cars are isolated but holes are created in the masks in Fig. 7.13 because of the uniform appear-

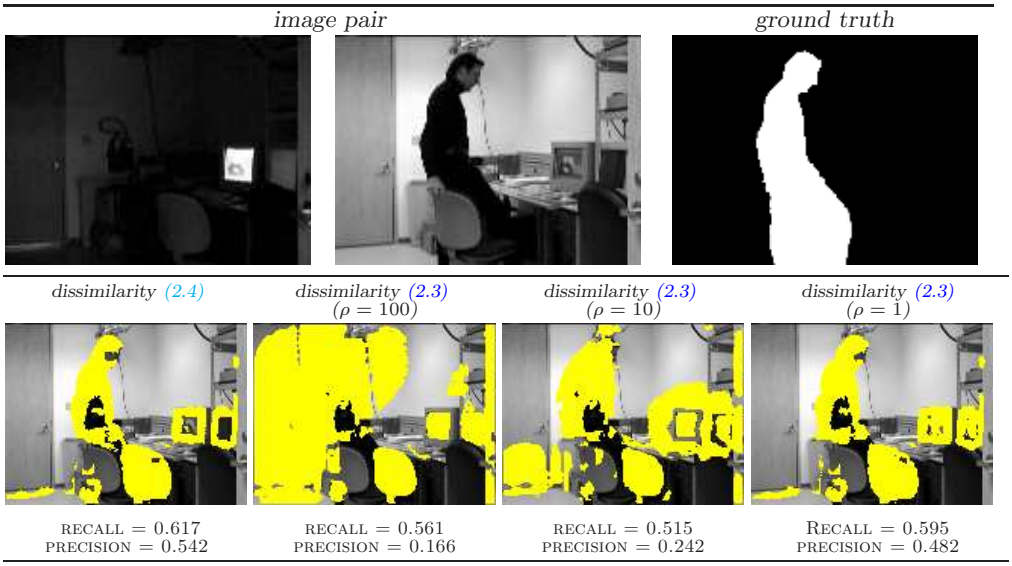


FIG. 7.10. *Robustness to sudden illumination changes with two different dissimilarity measures and different values of* ρ *(*$\varepsilon = 100$*) (see [99, 46, 78] for comparison).*

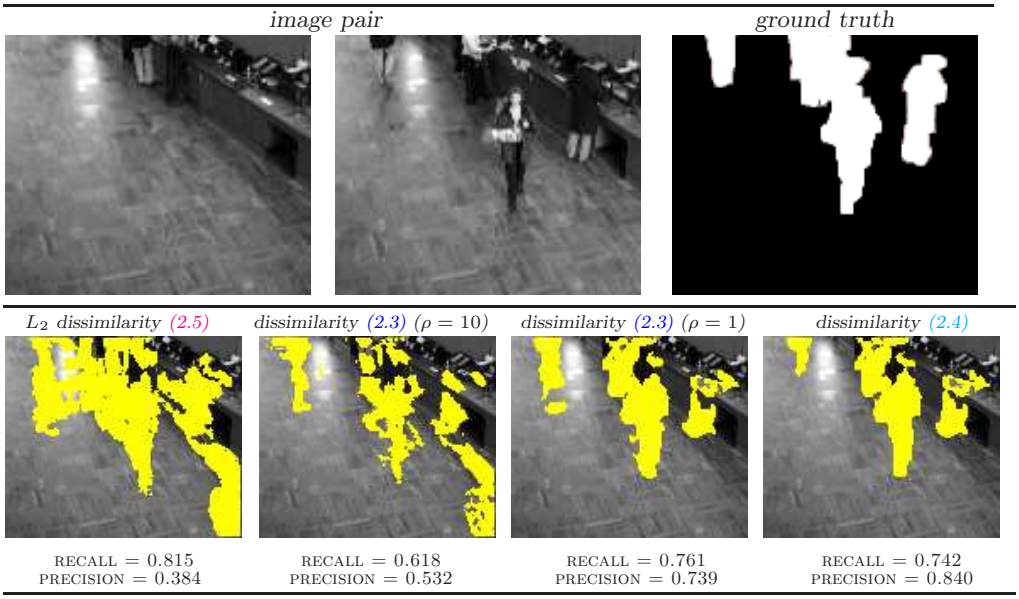


FIG. 7.11. *Robustness to shadows and specularities with the dissimilarity measures (2.3) (*$\varepsilon = 10$*) and (2.5) and (2.4) (*$\varepsilon = 1$*) (see [99, 46] for comparison).*

ance of the coat. The third method [26] produced regularized masks but no object and no car is respectively detected on the water and on the bridge in the background. We compared this image ($t = 84$) to an arbitrarily chosen reference image (Fig. 7.13, first row, left) showing the background and with no moving object. The difference image shows that the tested image (with moving objects) is slightly blurred when compared

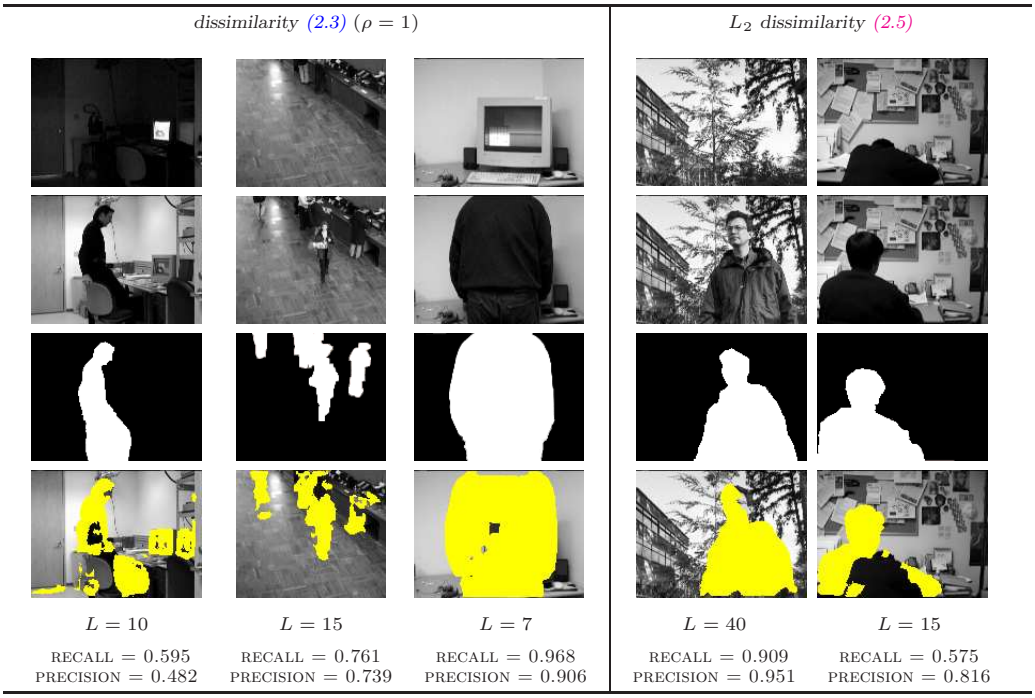


FIG. 7.12. *Change detection masks computed by the proposed method for the image pairs tested in [99, 46] (see also [78]). Two first rows: image pairs ; third row: ground truths ; fourth row: our detection results with the dissimilarity measures (2.3) and (2.5).*

to the reference image showing only the background. Additionally we observe residual movement due to vibrations of the camera. Nevertheless, the proposed multiscale modeling and the sliding window process enable to produce regularized masks and to detect the cars on the bridge and moving meaningful objects on the water ($L = 15$, $\hat{\lambda}_N = 3.72$). Another comparison with other competitive change detection methods on image pair taken from [41, 32] is shown in Fig. 7.14.

In Fig. 7.16, we compare our results to those obtained by Zivkovic & van der Heijden [107] and Pilet *et al.* [78] on three images taken from an image sequence from the PETS 2006 dataset. Unlike these two methods, our method exploits only the reference image shown on top row of Fig. 7.16. The persons are reliably detected and the results are very similar to those obtained in [78]. Finally, we compared our method to the three layer MRF method [10, 11] applied to an image pair for which the ground truth is available. In Fig. 7.15, our algorithm missed mainly a non-contrasted vehicle and detected a pair of crosses as expected.

The experiments we presented demonstrate that our multiscale method works well for tested image pairs. Intuitively, the number of scales is related to the area of missing objects or occlusions in the image pairs. The next experiments will verify the potential of the method for other applications given an image pair.

7.3. Occlusion and discontinuity detection. Additional experiments have been carried by running the change detection algorithm on more challenging test images. Results in Figs. 7.17-7.18 focus on the problem of detecting space-time occlusions in consecutive frames as in [64, 106, 52]. On the classic *"flower garden"* sequence,

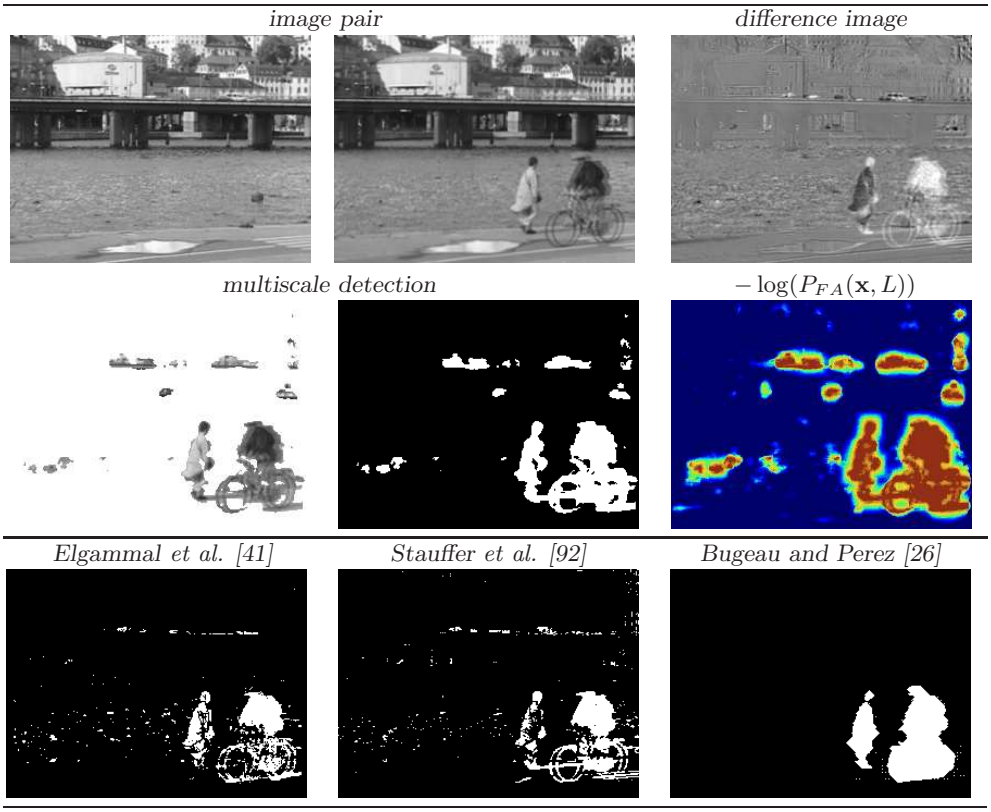


FIG. 7.13. *Comparison of background subtraction methods with our approach that utilizes a single reference image showing the background with no moving object.*

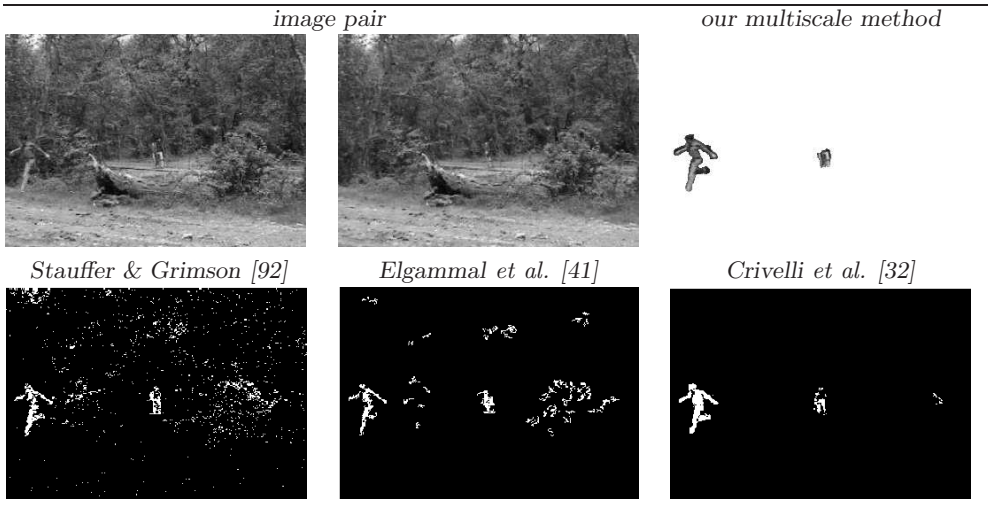


FIG. 7.14. *Comparison with three background subtraction methods using several frames from the analyzed image sequence [41, 32] and our multiscale change detection method that uses only a single reference image shown on top left (L_2 dissimilarity measure: $L = 5, \hat{\lambda}_N = 0.1$).*

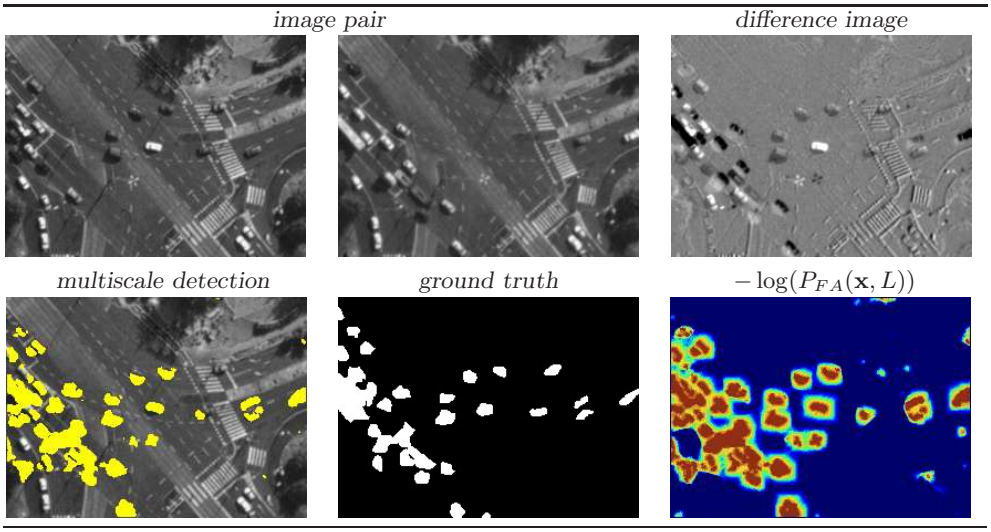


FIG. 7.15. *Occlusion detection in a traffic scene using* 3×3 *search windows* ($L = 12$, RECALL $= 0.759$, PRECISION $= 0.616$) (`http://web.eee.sztaki.hu/bcsaba/aerialObjectMotionBenchmark.htm`) *(see [10, 11]).*

where the large camera pan generates occlusions around the tree in the foreground, our detector extracted most of these occlusions without using color information or motion information. In this scenario, if many patches similar to the first one can be found (at a distance parametrized by the size $\sqrt{N} \times \sqrt{N}$ of $B(\mathbf{x})$), no change is detected. As shown in examples in Figs. 7.17-7.18, detected locations can be interpreted as meaningful changes in the scene corresponding to: 1) appearance or disappearance of scene parts ; 2) occlusions ; 3) motions of amplitude larger than $\sqrt{N/2}$ pixels. If $B(\mathbf{x})$ is large enough, the detector is potentially invariant to a wide range of movements, including those caused by camera displacement, to the extent of only detecting the two first types of events. In this example, the input images are highly redundant. Then, it makes sense to examine the situation with 3×3 search windows $B(\mathbf{x})$. A limited number of patch sizes ($L = 3$) is more suitable since occluded areas are small. Similar textured patches are actually found in the second image because of texture redundancy. If $B(\mathbf{x})$ is larger, several patches along discontinuities may be found in the second image and the set of occluded pixels would be smaller. This phenomena is related to the so-called "aperture problem" which is well known in motion estimation. Another critical issue we did not address in this paper is the capture of the camera motion amplitude. Larger search windows must be considered in that case. An alternative is to compensate the camera motion in a pre-processing step. The remaining detected regions correspond to occluded regions.

7.4. Comparison with global cost functionals. In order to demonstrate the benefits of the change detection method, we compare the results with more conventional techniques. As baseline algorithms and for the sake comparison, we chose to minimize two cost functionals, with a min-cut/max-flow algorithm, in the spirit of most approaches for change detection. Each studied energy is the combination of two terms, each of them corresponding to a precise property which must be satisfied by the optimization solution.

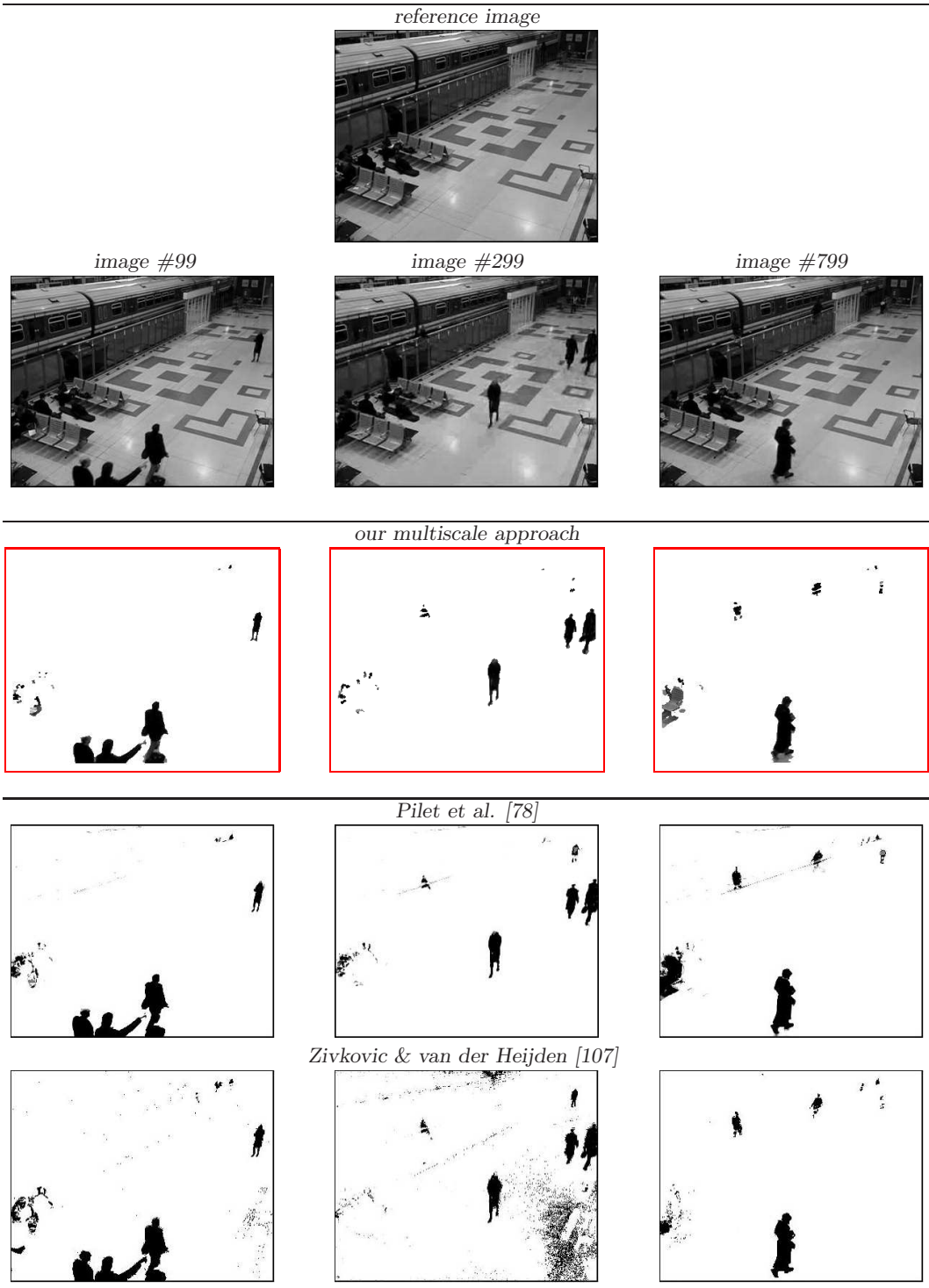


FIG. 7.16. *Detection on three images from the PETS 2006 dataset. The methods [107, 78] exploit the image sequence unlike our method which uses only a single reference image (top row).*

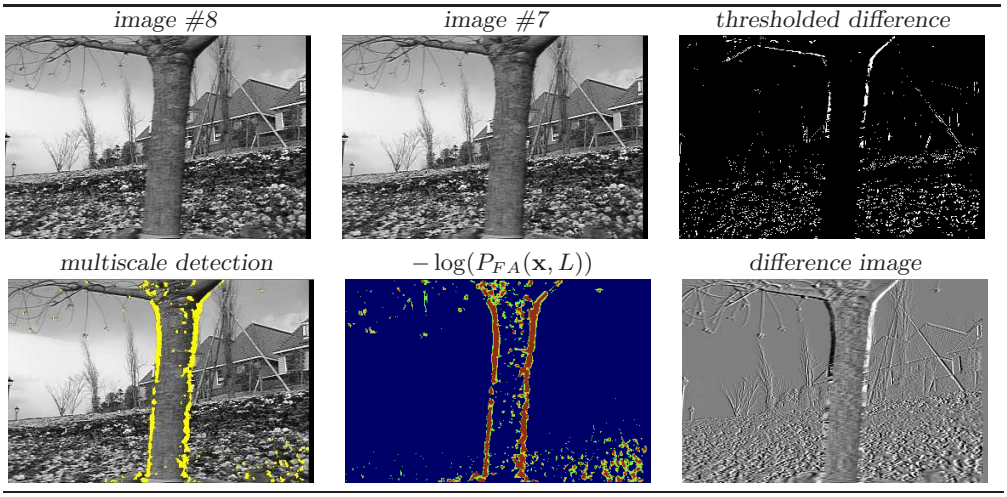


FIG. 7.17. *Detection of spatio-temporal discontinuities in "flower garden" (3×3 neighborhoods, 3×3 search windows, $L=3$). The discontinuity areas (superimposed on the original image #7) correspond to the highest level set of the image $-\log(P_{FA}(\mathbf{x}, L))$.*

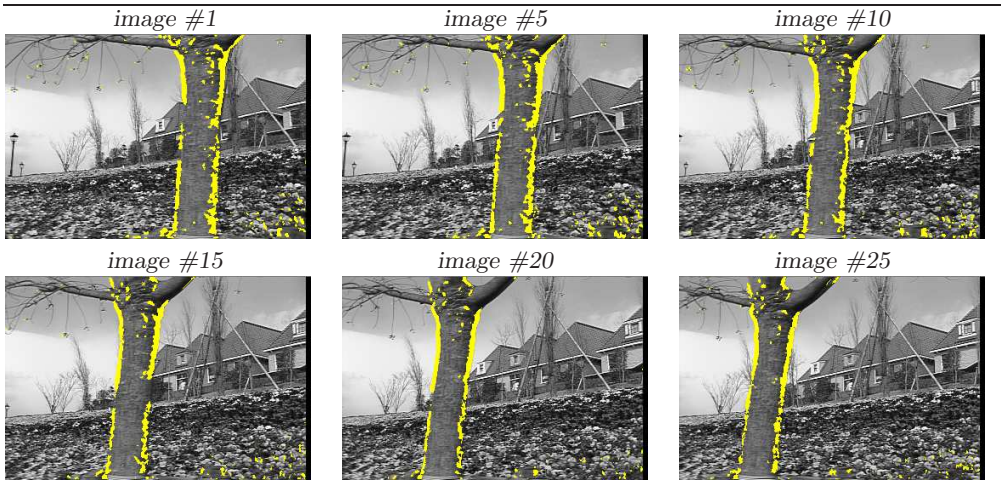


FIG. 7.18. *Space-time discontinuities in "flower garden".*

Let Θ the set of configurations $\theta = (\theta(\mathbf{x}))_{\mathbf{x}\in\Omega}$ with $\theta(\mathbf{x}) \in \{0, 1\}$ for all $\mathbf{x} \in \Omega$. We consider the following patch-based cost functional [57]:

$$J_{uv}(\theta) = \sum_{\mathbf{x}\in\Omega} \sum_{\ell=1}^{L} a_\ell \left| \frac{1}{N} S_{N,\ell}(\mathbf{x}) - \theta(\mathbf{x}) \right| + \sum_{<\mathbf{x},\mathbf{y}>} \psi(\theta(\mathbf{x}), \theta(\mathbf{y})). \tag{7.1}$$

The first terms in (7.1) tends to label a pixel as subject to change if the number of positive decisions (when u is compared to v) is high for a large number of patch sizes. We set $a_\ell = \frac{a'_\ell}{\sum_{\ell=1}^{L} a'_\ell}$ with either $a'_\ell = 1$ to consider equally the maps $\{S_{N,\ell}\}$ or $a'_\ell = (2\ell + 1)^{-1}$ to give more weights to maps $\{S_{N,\ell}\}$ obtained with small patch

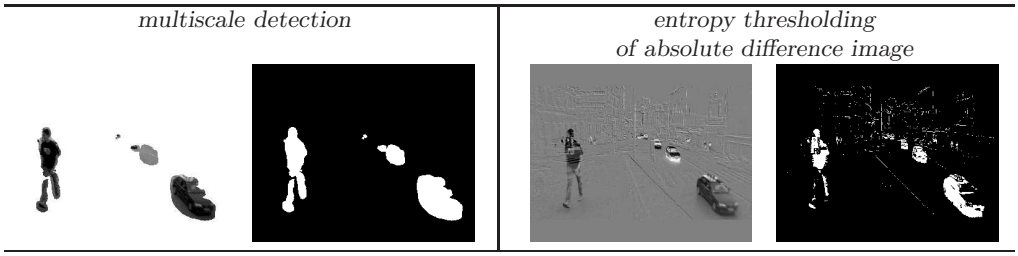


FIG. 7.19. *Our multiscale detector ($L = 11$) applied to the image pair shown in Fig. 7.1.*

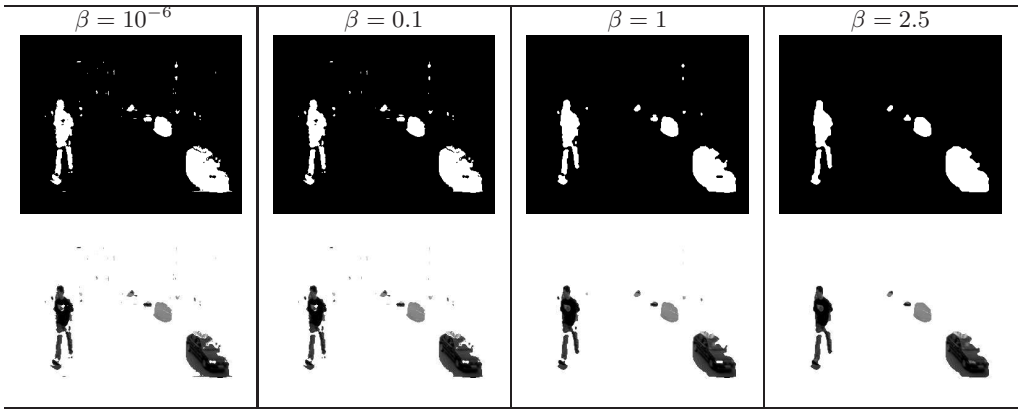


FIG. 7.20. *Detection results for different values of β which controls the patch-based energy functional $J_{uv}(\theta)$ with $a'_\ell = (2\ell+1)^{-1}, \ell = 1, \cdots, L$ (image pair shown in Fig. 7.1).*

sizes as it may be recommended in specific applications. In our experiments, we evaluated these two arbitrary weights but other rules to combine the maps $\{S_{N,\ell}\}$ can be found. The second term in (7.1) takes the form of an Ising MRF (Markov Random Field) model. It is defined on pairs $< \mathbf{x}, \mathbf{y} >$ of neighboring pixels (with respect to 4-neighborhood system) as

$$\psi(\theta(\mathbf{x}), \theta(\mathbf{y})) = \begin{cases} \beta & \text{if } \theta(\mathbf{x}) = \theta(\mathbf{y}), \\ 0 & \text{otherwise,} \end{cases} \tag{7.2}$$

where $\beta > 0$ is the balance parameter manually adjusted. This parameter has significant effects on the qualitative properties of the minimizer. Imposing more regularity by increasing β tends to simultaneously decrease the perimeter and the number of connected components.

On the image pair shown in Fig. 7.1, we present the detection results obtained respectively by automatic thresholding of the absolute image difference (Fig. 7.19 (right)). In Fig. 7.20, regular masks are estimated by minimizing (7.1) but small undesirable objects are extracted for all the tested β values. Moreover, rough blobs are generally isolated as it is confirmed on additional examples and applications (see Fig. 7.22). To overcome these difficulties, more sophisticated cost functionals and Conditional Random Fields (CRF) models including edge terms could be designed as proposed in [59, 95, 52, 96]. Finally, the object boundaries are better delineated using our multiscale framework (Fig. 7.21).

It is worth noting that a compromise is not easily achieved to produce results

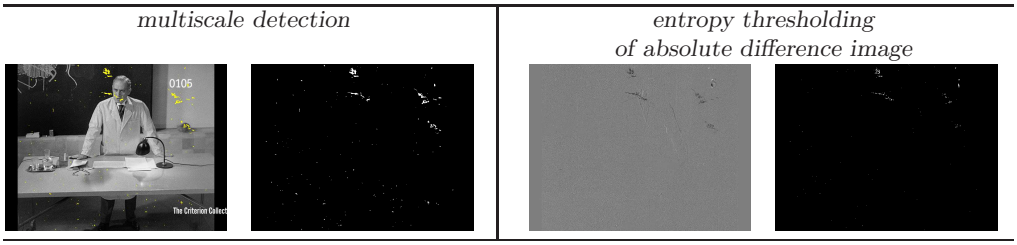


FIG. 7.21. *Our multiscale detector ($L = 3$) applied to the image pair shown in Fig. 7.5.*

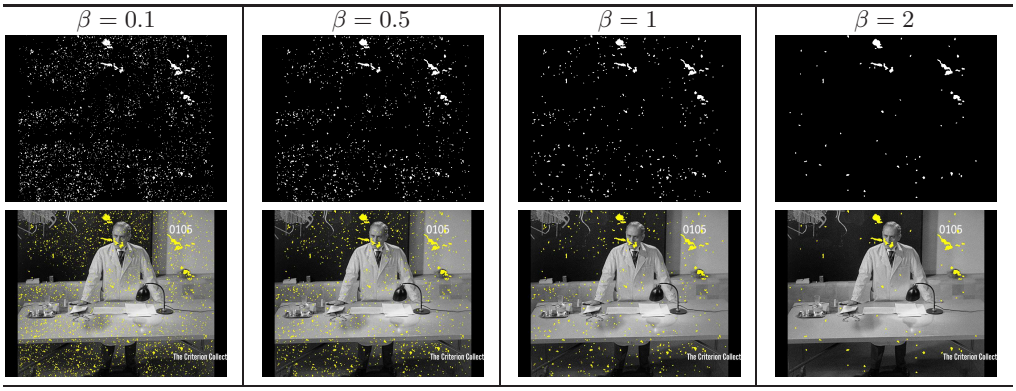


FIG. 7.22. *Detection results for different values of β which controls the patch-based energy functional $J_{uv}(\theta)$ with $a'_\ell = (2\ell + 1)^{-1}, \ell = 1, \cdots, L$ (image pair shown in Fig. 7.5).*

visually similar to those obtained using our testing method which globally analyzes the occurrence of dependent counts for different patch sizes (see Fig. 7.21). It turns out that the cost functional (7.1) is not equivalent to our probabilistic detection approach. Intuitively, β is a parameter which can be thought as a parameter controlling the number of false alarms. Unlike the energy-based counterpart, a reliability measure expressed formally in terms of false alarm rates is provided by using our approach.

8. Conclusion. In this paper, we have described a non-parametric multiscale change detector and we have presented a theoretical study of its statistical properties. The method robustly detects areas in images where the redundancy property captured by image patches does not hold. It is particular robust to many types of variations such as local appearance changes, residual motions and scale variations. In this approach, local and independent decisions for nearby patches are collected and a decision is made for a change detection if the number of dissimilar patches exceeds a given threshold. This procedure is performed for different patch sizes and a multiscale fusion decision rule is used for final change detection. This approach is capable of extracting clean occlusion/change masks.

An important feature of the approach is that image motion does not have to be computed explicitly. Applied to the specific problem of foreground detection in an image sequence with static camera, our method, using only two input images, performs as well as and sometimes better than methods [92, 41, 2, 78] which require a long image sequence as input. We demonstrated on real and complex image pairs the

ability of this unified approach to detect appearance/disappearance of objects, motion occlusions, and blotches in old movies. We explored the estimation of the patch size and of the search window size, and addressed the robustness to global illumination changes. To our knowledge, no previous method addressed all these variations at the same time and fuse multiple decisions as we did.

In all experiments, gray level intensity values are used for matching although color images could be considered. We did not address the stereo problem yet since the displacements are traditionally large. In a forthcoming work, we plan to investigate a conditional random field modeling from the false alarm probability maps provided by the change detection method described in the paper.

Acknowledgments. The authors thank the anonymous reviewers for helping to improve the presentation of the work.

REFERENCES

[1] T. Aach and A. Condurache, *Transformation of adaptive thresholds by significance invariance for change detection*, in Proceedings of the IEEE Workshop on Statistical Signal Processing, Bordeaux, 2005, pp. 637–642.

[2] T. Aach, L. Dumbgen, R. Mester, and D. Toth, *Bayesian illumination-invariant motion detection*, in Proceedings of the IEEE International Conference on Image Processing, vol. 3, Thessaloniki, Greece, 2001, pp. 640–643.

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

[4] A. Adam, E. Rivelin, and I. Shimshoni, *Robust fragments-based tracking using the integral histogram*, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, New-York, NY, 2006, pp. 798–805.

[5] L. Alvarez, R. Deriche, T. Papadopoulo, and J. Sanchez, *Symmetrical dense optical flow estimation with occlusions detection*, in Proceedings of the European Conference on Computer Vision, vol. 1, Copenhagen, Denmark, 2002, pp. 721–735.

[6] R. Arratia, L. Goldstein, and L. Gordon, *Poisson approximation and the Chen-Stein method*, Statistical Science, 5 (1990), pp. 403–434.

[7] S.P. Awate and R.T. Whitaker, *Unsupervised, information-theoretic, adaptive image filtering for image restoration*, IEEE Trans. Pattern Analysis Machine Intelligence, 28 (2006), pp. 364–376.

[8] C. Ballester, E. Cubero-Castan, M. Gonzalez, and J.M. Morel, *Contrast invariant image intersection*, in Advanced Mathematical Methods in Measurement and Instrumentation, 2000, pp. 41–55.

[9] A.D. Barbour, L. Holst, and S. Janson, *Poisson approximation*, Oxford - University Press, 1992.

[10] C. Benedek, T. Szirányi, Z. Kato, and J. Zerubia, *A multi-layer MRF model for object-motion detection in unregistered airborne image-pairs*, in Proceedings of the IEEE International Conference on Image Processing, vol. 6, San Antonio, TX, 2007, pp. 141–144.

[11] ———, *Detection of object motion regions in aerial image pairs with a multi-layer Markovian model*, IEEE Trans. Image Processing, 18 (2009), pp. 2303–2315.

[12] Y. Benjamini and Y. Hochberg, *Controlling the false discovery rate: a practical and powerful approach to multiple testing*, J. Royal Statistical Society B, 57 (1995), pp. 289–300.

[13] J. Biemond, P.M.B. van Roosmalen, and R.L. Lagendijk, *Improved blotch detection by postprocessing*, in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 6, Washington, DC, 1999, pp. 3101–3104.

[14] M.J. Black and D.J. Fleet, *Probabilistic detection and tracking of motion boundaries*, Int. J. Computer Vision, 38 (2000), pp. 231–245.

[15] O. Boiman and M. Irani, *Detecting irregularities in images and in video*, in Proceedings of the IEEE International Conference on Computer Vision, vol. 1, Beijing, China, 2005, pp. 462–469.

[16] ———, *Detecting irregularities in images and in video*, Int. J. Computer Vision, 74 (2007), pp. 17–31.

[17] J.S. De Bonet, *Noise reduction through detection of signal redundancy*, 1997. Rethinking Artificial Intelligence, MIT AI Lab.

[18] M. Bosc, F. Heitz, J.-P. Armspach, D. Namer, I.and Gounot, and L. Rumbach, *Automatic change detection in multimodal serial MRI: application to multiple sclerosis lesion evolution*, NeuroImage, 20 (2003), pp. 643–656.

[19] J. Boulanger, A. Gidon, J. Salamero, J.-B. Sibarita, and C. Kervrann, *Repetitive and transient event detection in fluorescence video-microscopy*, in Focus on Microscopy, Osaka-Awaji, Japan, 2008.

[20] J. Boulanger, C. Kervrann, and P. Bouthemy, *Space-time adaptation for patch based image sequence restoration*, IEEE Trans. Pattern Analysis Machine Intelligence, 29 (2007), pp. 1096–1102.

[21] T. Brox, O. Kleinschmidt, and D. Cremers, *Efficient nonlocal means for denoising of textural patterns*, IEEE Trans. Image Processing, 17 (2008), pp. 1083–1092.

[22] L. Bruzzone and D.F. Prieto, *Automatic analysis of the difference image for unsupervised change detection*, IEEE Trans. Geoscience and Remote Sensing, 38 (2000), pp. 1171–1182.

[23] A. Buades, B. Coll, and J.M. Morel, *Nonlocal image and movie denoising*, Int. J. Computer Vision, 76 (2008), pp. 123–139.

[24] A. Buades, B. Coll, and J. M. Morel, *A review of image denoising algorithms, with a new one*, SIAM J. Multiscale Modeling and Simulation, 4 (2005), pp. 490–530.

[25] T. Buades, J. Delon, Y. Gousseau, and S. Masnou, *Adaptive blotches detection for film restoration*, in Proceedings of the IEEE International Conference on Image Processing, Hong-Kong, 2010, pp. 3317–3320.

[26] A. Bugeau and P. Pérez, *Detection and segmentation of moving objects in complex scenes*, Computer Vision Image Understanding, 113 (2009), pp. 459–476.

[27] F. Cao and P Bouthemy, *A general principled method for image similarity validation*, in Proceedings of the International Conference on Adaptive Multimedia Retrieval, Geneva, Switzerland, 2006, pp. 57–70.

[28] V. Caselles, J.L. Lisani, J.-M. Morel, and G. Sapiro, *Shape preserving local histogram modification*, IEEE Trans. Image Processing, 8 (1999), pp. 220–230.

[29] I. J. Cox, S. Roy, and S. L. Hingorani, *Dynamic histogram warping of image pairs for constant image brightness*, in Proceedings of the IEEE International Conference on Image Processing, vol. 2, Washington, DC, 1995, pp. 366–369.

[30] A. Criminisi, G. Cross, A. Blake, and V. Kolmogorov, *Bilayer segmentation of live video*, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, New-Yor, NY, 2006, pp. 53–60.

[31] A. Criminisi, P. Pérez, and K. Toyama, *Region filling and object removal by exemplar-based image inpainting*, IEEE Trans. Image Processing, 13 (2004), pp. 1200–1212.

[32] T. Crivelli, G. Piriou, P. Bouthemy, B. Cernuschi Frias, and J.F. Yao, *Simultaneous motion detection and background reconstruction with a mixed-state conditional Markov random field*, in Proceedings of the European Conference on Computer Vision, vol. 1, Marseille, France, 2008, pp. 113–126.

[33] M. Irani D. Glasner, S. Bagon, *Super-resolution from a single image*, in Proceedings of the IEEE International Conference on Computer Vision, Kyoto, Japan, 2009, pp. 349–356.

[34] J. Delon, *Movie and video scale-time equalization, application to flicker reduction*, IEEE Trans. Image Processing, 15 (2006), pp. 241–248.

[35] Y. Deng, Q. Yang, X. Lin, and X. Tang, *A symmetric patch-based correspondence model for occlusion handling*, in Proceedings of the IEEE International Conference on Computer Vision, vol. 2, Beijing, China, 2005, pp. 1316–1322.

[36] A. Desolneux, L. Moisan, and J.-M. Morel, *Meaningful alignments*, Int. J. Computer Vision, 40 (2000), pp. 7–23.

[37] F. Dibos and G. Koepler S. Pelletier, *Adapted windows detection of moving objects in video scenes*, SIAM J. Imaging Science, 2 (2009), pp. 1–19.

[38] M. Ebrahimi and E.R. Vrscay, *Examining the role of scale in the context of the non-local-means filter*, in Proceedings of the 5th international conference on Image Analysis and Recognition, Povoa de Varzim, Portugal, 2008, pp. 170–181.

[39] A.A. Efros and T.K. Leung, *Texture synthesis by non-parametric sampling*, in Proceedings of the IEEE International Conference on Computer Vision, vol. 2, Kerkyra, Greece, 1999, pp. 1033–1038.

[40] M. Elad and M. Aharon, *Image denoising via learned dictionaries and sparse representation*, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, New York, NY, 2006, pp. 895–900.

[41] A. Elgammal, D. Harwood, and L. Davis, *Non-parametric model for background subtraction*, in Proceedings of the European Conference on Computer Vision, vol. 2, Dublin, Ireland, 2000, pp. 751–767.

[42] R. Fransens, C. Strecha, and L. Van Gool, *A mean field EM-algorithm for coherent occlusion handling in MAP-estimation problems*, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, New-York, NY, 2006, pp. 300–307.

[43] W.T. Freeman, E.C. Pasztor, and O.T. Carmichael, *Learning low-level vision*, Int. J. Computer Vision, 40 (2000), pp. 25–47.

[44] D. Geman, S. Geman, C. Graffigne, and P. Dong, *Boundary detection by constrained optimization*, IEEE Trans. Pattern Analysis Machine Intelligence, 12 (1990), pp. 609–628.

[45] Z. Harmany, R. Willett, A. Singh, and R. Nowak, *Controlling the error in FMRI: hypothesis testing or set estimation ?*, in Proceedings of the IEEE International Symposium on Biomedical Imaging, Paris, France, 2008, pp. 552–555.

[46] M. Heikkila and M. Pietikainen, *A textured-based method for modeling the background and detecting moving objects*, IEEE Trans. Pattern Analysis Machine Intelligence, 28 (2006), pp. 657–662.

[47] F. Heitz and P. Bouthemy, *Multimodal estimation of discontinuous optical flow using Markov Random Fields*, IEEE Trans. Pattern Analysis Machine Intelligence, 15 (1993), pp. 1217–1232.

[48] M.-O. Hongler, Y.L. de Meneses, A. Beyeler, and J. Jacot, *The resonant retina: exploiting vibration noise to optimally detect edges in an image*, IEEE Trans. Pattern Analysis Machine Intelligence, 25 (2003), pp. 1051–1062.

[49] S. Ince and J. Konrad, *Geometry-based estimation of occlusions from video frame pairs*, in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 2, Philadelphia, PA, 2005, pp. 933–936.

[50] H. Ishikawa and D. Geiger, *Occlusions, discontinuities, and epipolar lines in stereo*, in Proceedings of the European Conference on Computer Vision, vol. 1, Freiburg, Germany, 1998, pp. 232–248.

[51] P.-M. Jodoin, M. Mignotte, and J. Konrad, *Statistical background subtraction using spatial cues*, IEEE Trans. Circuits and Systems for Video Technology, 17 (2007), pp. 1758–1763.

[52] P.-M. Jodoin, M. Mignotte, and C. Rosenberger, *Segmentation framework based on label field fusion*, IEEE Trans. Image Processing, 16 (2007), pp. 2535–2550.

[53] J.N. Kapur, P.K. Sahoo, and A.K.C. Wong, *A new method for gray-level picture thresholding using the entropy of the histogram*, Computer Vision Graphical Image Processing, 29 (1985), pp. 273–285.

[54] T. Kasetkasem and P.K. Varshney, *An image change detection algorithm based on Markov random field models*, IEEE Trans. Geoscience and Remote Sensing, 40 (2002), pp. 1815–1823.

[55] N. Katenka, E. Levina, and G. Michailidis, *Local vote decision fusion for target detection in wireless sensor networks*, IEEE Trans. Signal Processing, 56 (2008), pp. 329–338.

[56] C. Kervrann and J. Boulanger, *Optimal spatial adaptation for patch-based image denoising*, IEEE Trans. Image Processing, 15 (2006), pp. 2866–2878.

[57] C. Kervrann, J. Boulanger, T. Pécot, and P. Pérez, *Discriminant random field and patch-based redundancy analysis for image change detection*, in Proceedings of the IEEE Workshop on Machine Learning for Signal Processing, Grenoble, France, 2009, pp. 1–6.

[58] C. Kervrann and F. Heitz, *A Markov random field model-based approach to unsupervised texture segmentation using local and global spatial statistics*, IEEE Trans. Image Processing, 4 (1995), pp. 856–862.

[59] P. Kohli and P.H.S. Torr, *Measuring uncertainty in graph cut solutions: Efficiently computing min-marginal energies using dynamic graph cuts*, in Proceedings of the European Conference on Computer Vision, vol. 2, Graz, Austria, 2006, pp. 30–43.

[60] V. Kolmogorov and R. Zabih, *Computing visual correspondence with occlusions via graph cuts*, in Proceedings of the IEEE International Conference on Computer Vision, vol. 2, Vancouver, Canada, 2001, pp. 508–515.

[61] ———, *Multi-camera scene reconstruction via graph cuts*, in Proceedings of the European Conference on Computer Vision, vol. 3, Copenhagen, Denmark, 2002, pp. 82–96.

[62] S. Kumar and M. Hebert, *Discriminative random fields*, Int. J. Computer Vision, 68 (2006), pp. 179–201.

[63] T.K. Leung and J. Malik, *Detecting, localizing and grouping repeated scene elements from an image*, in Proceedings of the European Conference on Computer Vision, vol. 1, Cambridge, UK, 1996, pp. 546–555.

[64] K.P. Lim, A. Das, and M.N. Chong, *Estimation of occlusion and dense motion fields in a bidirectional bayesian framework*, IEEE Trans. Pattern Analysis Machine Intelligence, 24

(2002), pp. 712–718.
[65] S. Lim, A. Mittal, L.S. Davis, and N. Paragios, *Fast illumination invariant background subtraction using two views: error analysis, sensor placement and applications*, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, San-Diego, CA, 2005, pp. 1071–1078.
[66] J.L. Lisani and J.-M. Morel, *Detection of major changes in satellite images*, in Proceedings of the IEEE International Conference on Image Processing, vol. 1, Barcelona, Spain, 2003, pp. 941–944.
[67] L. Lu and G.D. Hager, *Dynamic background/foreground segmentation from images and videos using random patches*, in Proceedings of the Annual Conference on Neural Information Processing Systems, Vancouver, B.C. Canada, 2006, pp. 1–8.
[68] Y. Matsushita and S. Lin, *A probabilistic intensity similarity measure based on noise distribution*, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, MN, 2007, pp. 1–8.
[69] Y. Matsushita, K. Nishino, K. Ikeuchi, and M. Sakauchi, *Illumination normalization with time-dependent intrinsic images for video surveillance*, IEEE Trans. Pattern Analysis Machine Intelligence, 26 (2004), pp. 1336–1347.
[70] E. Mémin and P. Pérez, *Hierarchical estimation and segmentation of dense motion fields*, Int. J. Computer Vision, 46 (2002), pp. 129–155.
[71] M. Mignotte, *Nonparametric multiscale energy-based model and its application in some imagery problems*, IEEE Trans. Pattern Analysis Machine Intelligence, 26 (2004), pp. 184–197.
[72] A. Mittal and N. Paragios, *Motion-based background subtraction using adaptive kernel density estimation*, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, vol. 2, Washington, DC, 2004, pp. 302–309.
[73] P. Monasse and F. Guichard, *Fast computation of a contrast-invariant image representation*, IEEE Trans. Image Processing, 9 (2000), pp. 860–872.
[74] R. Niu, P.K. Varshney, M. Moore, and D. Klamer, *Decision fusion in a wireless sensor network with a large number of sensors*, in Proceedings of the International Conference on Information Fusion, Stockholm, Sweden, 2004, pp. 1–7.
[75] N. Paragios and R. Deriche, *Geodesic active contours and level sets for the detection and tracking of moving objects*, IEEE Trans. Pattern Analysis Machine Intelligence, 22 (2000), pp. 266–280.
[76] N. Paragios and G. Tziritas, *Adaptive detection and localization of moving objects in image sequences*, Signal Processing: Image Communication, 14 (1999), pp. 277–296.
[77] T. Pécot, C. Kervrann, S. Bardin, B. Goud, and J. Salamero, *Patch-based Markov models for event detection in fluorescence bioimaging*, in Proceedings of the International Conference on Medical Image Computing and Computer Assisted Intervention, vol. 2, New York City, NY, 2008, pp. 95–103.
[78] J. Pilet, C. Strecha, and P. Fua, *Making background subtraction robust to sudden illumination changes*, in Proceedings of the European Conference on Computer Vision, vol. 4, Marseille, France, 2008, pp. 567–580.
[79] G. Piriou, P. Bouthemy, and J-F. Yao, *Extraction of semantic dynamic content from videos with probabilistic motion models*, in Proceedings of the European Conference on Computer Vision, vol. 3, Prague, Czech Republic, 2004.
[80] S. Prima, D.L. Arnold, and D.L. Collins, *Multivariate statistics for detection of ms activity in serial multimodal mr images*, in Proceedings of the International Conference on Medical Image Computing and Computer Assisted Intervention, vol. 2, Montreal, Canada, 2003, pp. 663–670.
[81] M. Protter, M. Elad, H. Takeda, and P. Milanfar, *Generalizing the non-local-means to super-resolution reconstruction*, IEEE Trans. Image Processing, 18 (2009), pp. 36–51.
[82] R.J. Radke, S. Andra, O. Al Kofahi, and B. Roysam, *Image change detection algorithms: a systematic survey*, IEEE Trans. Image Processing, 14 (2005), pp. 294–307.
[83] A. Robin, L. Moisan, and S. Le Hgarat-Mascle, *An a-contrario approach for sub-pixel change detection in satellite imagery*, IEEE Trans. Pattern Analysis Machine Intelligence, 32 (2010), pp. 1977–1993.
[84] A. Roche, G. Malandain, and N. Ayache, *Unifying maximum likelihood approaches in medical image registration*, Int. J. Imaging Systems and Technology, 11 (2000), pp. 71–80.
[85] P.L. Rosin, *Thresholding for change detection*, Computer Vision Image Understanding, 86 (2002), pp. 79–95.
[86] S. Roth and M.J. Black, *Fields of experts: A framework for learning image priors*, in

Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, vol. 2, San Diego, CA, 2005, pp. 860–867.

[87] F. Rousseau, F. Blanc, J. De Seze, L. Rumbach, and J.-P. Armpach, *An a contrario approach for outliers segmentation: application to multiple sclerosis in MRI*, in Proceedings of the IEEE International Symposium on Biomedical Imaging, Paris, France, 2008, pp. 9–12.

[88] N. Sabater, A. Almansa, and J.-M. Morel, *Rejecting wrong matches in stereovision*, Tech. Report 28, CMLA, 2008.

[89] H.J. Seo and P. Milanfar, *A non-parametric approach to automatic change detection in MRI images of the brain*, in Proceedings of the IEEE International Symposium on Biomedical Imaging, Boston, MA, 2009, pp. 245–248.

[90] E. Shechtman and M. Irani, *Matching local self-similarities across images and videos*, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, MN, 2007, pp. 1–8.

[91] Y. Sheikh and M. Shah, *Bayesian object detection in dynamic scenes*, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, San Diego, CA, 2005, pp. 74–79.

[92] C. Stauffer and W.E.L. Grimson, *Adaptive background mixture models for real time tracking*, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Ft. Collins, CO, 1999, pp. 246–252.

[93] C. Strecha, R. Fransens, and L. Van Gool, *A probabilistic approach to large displacement optical flow and occlusion detection*, in Proceedings of the ECCV Workshop Statistical Methods in Video Processing, Prague, Czech Republic, 2004, pp. 71–82.

[94] J. Sun, Y. Li, S.-B. Kang, and H.-Y. Shum, *Symmetric stereo matching for occlusion handling*, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, San Diego, CA, 2005, pp. 299–406.

[95] J. Sun, W. Zhang, X. Tang, and H.-Y. Shum, *Background cut*, in Proceedings of the European Conference on Computer Vision, vol. 2, Graz, Austria, 2006, pp. 628–641.

[96] M. Szummer, P. Kohli, and D. Hoiem, *Learning CRFs using graph cuts*, in Proceedings of the European Conference on Computer Vision, vol. 2, Marseille, France, 2008, pp. 582–595.

[97] S. Tilie, L. Laborelli, and I. Bloch, *A contrario false alarms removal for improving blotch detection in digitized films restoration*, in Proceedings of the EURASIP Conference focused on Speech and Image Processing, Multimedia Communications and Services, Maribor, Slovenia, 2007, pp. 410–413.

[98] D. Toth, T. Aach, and V. Metzler, *Bayesian spatio-temporal motion detection under varying illumination*, in Proceedings of the European Signal Processing Conference, Tampere, Finland, 2000, pp. 2081–2084.

[99] K. Toyama, J. Krumm, B. Brumitt, and B. Meyers, *Wallflower: principles and practice of background maintenance*, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, Ft. Collins, CO, 1999, pp. 255–261.

[100] T. Veit, F. Cao, and P. Bouthemy, *Probabilistic parameter-free motion detection*, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, Washington, DC, 2004, pp. 715–721.

[101] Y. Wei and L. Quan, *Asymmetrical occlusion handling using graph cut for multi-view stereo*, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, vol. 2, San-Diego, CA, 2005, pp. 902–909.

[102] P. Weiss, A. Fournier, L. Blanc-Féraud, and G. Aubert, *On the illumination invariance of the level lines under directed light: Application to change detection*, SIAM J. Imaging Sciences, 4 (2011), pp. 448–471.

[103] O. Williams, M. Isard, and J. MacCormick, *Estimating disparity and occlusions in stereo video sequences*, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, vol. 2, San-Diego, CA, 2005, pp. 250–257.

[104] C.R. Wren, A. Azarbayejani, T. Darrell, and A. Pentland, *Pfinder: real-time tracking of the human body*, IEEE Trans. Pattern Analysis Machine Intelligence, 19 (1997), pp. 780–785.

[105] J. Xiao, H. Cheng, H.S. Sawhney, C. Rao, and M. Isnardi, *Bilateral filtering-based optical flow estimation with occlusion detection*, in Proceedings of the European Conference on Computer Vision, vol. 1, 2006, pp. 211–224.

[106] J. Xiao and M. Shah, *Motion layer extraction in the presence of occlusion using graph cuts*, IEEE Trans. Pattern Analysis Machine Intelligence, 27 (2005), pp. 1644–1659.

[107] Z. Zivkovic and F. van der Heijden, *Efficient adaptive density estimation per image pixel*

for the task of background subtraction, Pattern Recognition Letters, 27 (2006), pp. 773–780.

[108] M. Zontak and I. Cohen, *Kernel-based detection of defects on semiconductor wafers*, in Proceedings of the IEEE Workshop on Machine Learning for Signal Processing, Grenoble, France, 2009, pp. 1–6.