



HAL
open science

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

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

► **To cite this version:**

Charles Kervrann, Jérôme Boulanger, Thierry Pecot, Patrick Pérez, Jean Salamero. Multiscale neighborhood-wise decision fusion for redundancy detection in image pairs. 2010. inria-00487051v1

HAL Id: inria-00487051

<https://inria.hal.science/inria-00487051v1>

Preprint submitted on 27 May 2010 (v1), last revised 9 Jan 2013 (v2)

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

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

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.

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 are 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, 99, 2, 86, 66, 80, 1, 49]. For a recent survey, see [83]. 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, 81, 45, 88, 90], cell biology imaging [78, 19] and remote sensing [23, 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 [23, 54, 77, 76, 92, 31, 96, 9, 10], several authors proposed to capture the spatial

^{*}Centre INRIA Rennes Bretagne Atlantique / INRA - Département de Mathématiques et Informatique Appliquées (MIA) (Charles.Kervran@inria.fr)

[†]Mathematical Imaging Group - Johann Radon Institute for Computational and Applied Mathematics (RICAM) (jerome.boulanger@ricam.oeaw.ac.at)

[‡]Centre INRIA Rennes Bretagne Atlantique / INRA - Département de Mathématiques et Informatique Appliquées (MIA) - UMR 144 CNRS, Institut Curie

[§]Centre INRIA Rennes Bretagne Atlantique / Technicolor Corporate Research Rennes Laboratory (Patrick.Perez@technicolor.com)

[¶]UMR 144 CNRS - PICT IBISA - Institut Curie (Jean.Salamero@curie.fr)

correlation among nearby pixels using Markov random fields (MRF). Energy minimization is usually performed by graph-cut algorithms [31, 96] or level set methods [76]. 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 [103], 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 [106, 41, 93, 72, 51, 9, 10]. Each pixel of the current frame is then tested against the learned probability distribution function (PDF). Stauffer and Grimson [93] 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, 94]. 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, 107]. 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, 89]. The most recent algorithms based on graph-cuts [102] or loopy belief propagation [105, 95] include a visibility label in the energy formulation to compute dense disparity [102]. 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 [108]. 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 [91]. The redundancy property captured by image patches was previously exploited for image segmentation [44, 58], image denoising [17, 25, 56, 7, 40, 87], image inpainting [32], defect detection in images [110] 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, 24]. To detect the occlusions or changes occurring in two images, we propose further to

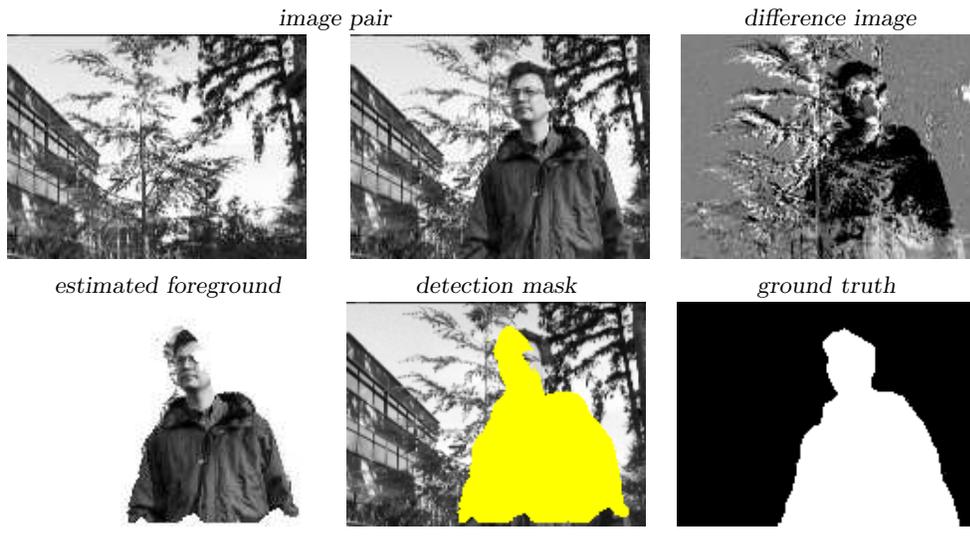


FIG. 1.1. Change detection in an image pair (see [100, 46]). Our algorithm correctly highlights the person who appeared in the second image, while ignoring temporal changes due to complex motion of tree branches in the background.

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 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 [75, 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, 101, 98, 88, 37, 84]. In the same spirit, Sabater *et al.* proposed recently in [89] 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, 101, 98, 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 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, 97].
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 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 previous methods (kernel-based [41, 72, 92] or mixture of Gaussian-based [93, 109, 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, 93, 31, 96, 33] 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 $B(\mathbf{x}) \subseteq \Omega$ 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 what follows, we will propose to make local decisions about the presence/absence of patches in the second image v that are similar to the reference in the first image u .

In [24, 39, 20], it has been confirmed that the L_2 distance is a good candidate to express the amount of dissimilarity between image patches. 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 enable to detect occlusions reliably for most studied situations.

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 :

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

where $g : \mathbb{R} \rightarrow \mathbb{R}^+$ 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 $W_n(\mathbf{t}) = \text{rect}_n(\mathbf{t}) \triangleq \mathbf{1}[\|\mathbf{t}\|_\infty \leq \frac{\sqrt{n}}{2}]$ where $\mathbf{1}[\cdot]$ 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, 22]) defined as :

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

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

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

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). 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 :

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

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

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

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

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 similarity measure [68] in the case of intensity-independent Gaussian noise with variance σ^2 where $\operatorname{erf}(\cdot)$ denotes the Gauss error function.

In the next sections, ϕ_{uv} will denote any of these similarity functions. Additional dissimilarity measures and functions $g(\cdot)$ based on the normalized cross-correlation [79] and more general probabilistic similarity measures [85, 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, 103] or dynamic histogram warping [30].

2.3. Parametric mixture models and first experiments. In Section 2.1, we have sketched a procedure based on patch comparisons, which tells whether a local change at the current pixel has occurred or not. To infer the thresholds for change detection, we propose here to study the usual image model (e.g. see [86, 2, 83]) of the form

$$(2.9) \quad \begin{aligned} u(\mathbf{x}) &= u_0(\mathbf{x}) + \epsilon(\mathbf{x}), \\ v(\mathbf{x}) &= v_0(\mathbf{x}) + \eta(\mathbf{x}) \end{aligned}$$

where u_0 and v_0 are the “true” images and the “errors” ϵ and η are i.i.d. (independent identically distributed) Gaussian zero-mean random variables with unknown variance σ^2 . In the absence of scene change, the noise being Gaussian in both images, the

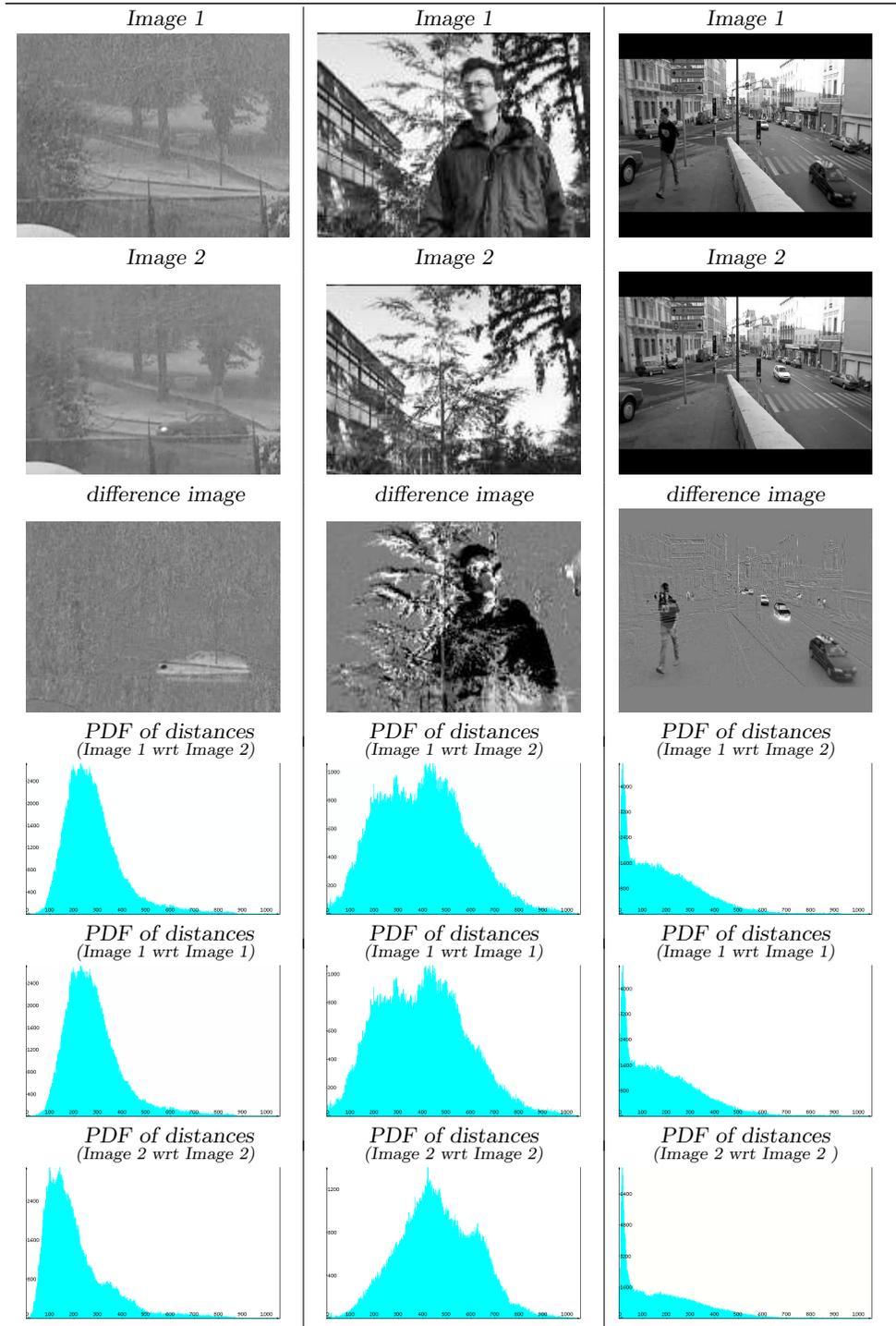


FIG. 2.1. PDFs of dissimilarity measures computed from image pairs and from a single image (see text). Left : snowy traffic scene ($n = 23 \times 23$, $|B(\mathbf{x})| = 3 \times 3$); middle : outdoor scene ($n = 23 \times 23$, $|B(\mathbf{x})| = 5 \times 5$); right : traffic scene ($n = 11 \times 11$, $|B(\mathbf{x})| = 5 \times 5$).

commonly-used distance $\phi_{uv}(\mathbf{x}, \mathbf{y})$ defined as

$$(2.10) \quad \phi_{uv}(\mathbf{x}, \mathbf{y}) = \frac{1}{2\sigma^2} \sum_{\mathbf{t} \in \mathbb{R}^2} \text{rect}_n(\mathbf{t}) (u(\mathbf{x} + \mathbf{t}) - v(\mathbf{y} + \mathbf{t}))^2$$

follows a central chi-squared distribution f_0 with n degrees of freedom: $\phi_{uv}(\mathbf{x}, \mathbf{y}) \sim \chi_n^2$. In the presence of scene change, under Gaussian noise assumptions, $\phi_{uv}(\mathbf{x}, \mathbf{y})$ is distributed according to the non-central chi-squared distribution f_1 with n degrees of freedom and an unknown spatially-varying parameter defined as

$$(2.11) \quad \varrho_{u_0v_0}(\mathbf{x}, \mathbf{y}) = \frac{1}{2\sigma^2} \sum_{\mathbf{t} \in \mathbb{R}^2} \text{rect}_n(\mathbf{t}) (u_0(\mathbf{x} + \mathbf{t}) - v_0(\mathbf{y} + \mathbf{t}))^2,$$

Furthermore, an usual way to evaluate the similarity of two matched patches consists in assuming that the distances $\phi_{uv}(\mathbf{x}, \mathbf{y})$ in a local neighborhood are independent and distributed according to the following two-component mixture distribution :

$$(2.12) \quad f(\phi_{uv}(\mathbf{x}, \mathbf{y})) = \pi_0 f_0(\phi_{uv}(\mathbf{x}, \mathbf{y})) + \pi_1 f_1(\phi_{uv}(\mathbf{x}, \mathbf{y}))$$

where π_0 (resp. $\pi_1 = 1 - \pi_0$) is the proportion of f_0 (resp. f_1). Since f_0 is perfectly known, i.e. $f_0 \equiv \chi_n^2$ with mean n and variance $2n$, we only need to estimate π_0 , σ and the non-centrality parameter $\varrho_{u_0v_0}(\mathbf{x}, \mathbf{y})$. Finally, the probability that a distance $\phi_{uv}(\mathbf{x}, \mathbf{y})$ is larger than a fixed threshold $\tau(\mathbf{x})$ is

$$(2.13) \quad \mathbf{P}(\phi_{uv}(\mathbf{x}, \mathbf{y}) \geq \tau(\mathbf{x})) = \pi_0 \int_{\tau(\mathbf{x})}^{\infty} f_0(\phi_{uv}(\mathbf{x}, \mathbf{y})) d\phi_{uv}(\mathbf{x}, \mathbf{y}) + \pi_1 \int_{\tau(\mathbf{x})}^{\infty} f_1(\phi_{uv}(\mathbf{x}, \mathbf{y})) d\phi_{uv}(\mathbf{x}, \mathbf{y}).$$

For ease of understanding, we evaluated experimentally the potential of this testing method for the entire image. Given an estimation of the noise variance σ^2 (e.g. [56]), we computed the PDFs of dissimilarity measures for several patch sizes and neighborhood sizes. As shown in Fig. 2.1, the observed PDFs are not stable for several tested image pairs and it was actually not possible to approximate the PDFs by mixtures of two components as described above. Even for small neighborhoods and homogeneous lighting conditions, the PDFs cannot be easily predicted, especially if the missing object areas are large. A more flexible strategy would be to give up the image model (2.9) and to approximate the PDFs by mixtures of Gaussians. Because it is too versatile, this modeling tends to capture all the observed variabilities and modes, which is not desirable for meaningful change detection. It is worth noting that considering small neighborhoods to estimate spatially-varying PDFs is also challenging and probably over-ambitious. Consequently, this idea was not investigated further.

In short, the aforementioned parametric modeling approach is generally too restrictive and not suitable to analyze any image pairs. The distortions observed in real images are not Gaussian distributed and known to be spatially non-stationary. Also, the neighboring and overlapping patches are not independent and the PDF of distances is potentially affected by the area of missing objects. Even by considering the same patch sizes and the same neighborhood sizes, the PDFs are not stable and bimodal. As a conclusion, the basic modeling (2.9) does not hold true in most of real tested situations. More interestingly, it turns out the PDFs of distances computed from a reference image u and a second image v look surprisingly similar to the PDFs

of distances computed from a single reference image u (with the same patch sizes and neighborhood sizes). For this, we examined the dissimilarities between a reference patch in u and its neighboring patches taken respectively in the reference image u and the image v (see Fig. 2.1). In the remainder of the paper, we exploit this result and propose an original non-parametric approach for detecting changes with no prior model on PDFs.

3. Neighborhood-wise decision. In this section, we describe the general probabilistic framework proposed for non-parametric change detection. In particular, we make a collective decision from the decisions obtained at neighboring points. Cooperation among neighboring points tends to enhance the ability to detect meaningful changes, namely, whether a change occurs or not within the search area $B(\mathbf{x})$.

3.1. Principle. 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 . Based on a dissimilarity measure $\phi_{uv}(\mathbf{x}, \mathbf{y})$ defined for instance as the L_2 distance between two patches (as explained in Section 2.2) :

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

we make a decision regarding the presence of a similar patch in the semi-local search area.

The collaborative neighborhood-wise decision for change detection is obtained through a fusion rule. An intuitive choice is to count the total number of positive decisions denoted S_N :

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

where $\tau(\mathbf{x})$ is a spatially-varying threshold. If the number of decisions exceeds the overall decision threshold $T \in \{1, \dots, N\}$ 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 :

$$(3.3) \quad D(\mathbf{x}) \triangleq \mathbb{1}[S_N(\mathbf{x}) \geq T].$$

This natural decision procedure that fuses local binary decisions has been already proposed in [49, 52] for change detection and in [75, 55] for analyzing distributed sensors in a wireless network.

3.2. Controlling the number of false alarms. In our framework, we declare that two patches are dissimilar if $\phi_{uv}(\mathbf{x}, \mathbf{y})$ exceeds a threshold $\tau(\mathbf{x})$, that is we select the hypothesis H_1 and the hypothesis H_0 on the contrary. The probability of false alarm, also called probability of Type I error, is defined as $\mathbf{P}(\text{decide } H_1 \text{ when } H_0 \text{ is true})$. This probability must be reduced as much as possible in most detection applications. In this section, we define the probabilities of false alarm corresponding to our modeling framework and we propose approximations required for computation.

Let $P_{fa}(\mathbf{x}, T) \triangleq \mathbf{P}(S_N(\mathbf{x}) \geq T | H_0)$ be the probability that there exists a collection of at least T pixels declared in favor of a detection under H_0 :

$$(3.4) \quad P_{fa}(\mathbf{x}, T) = \sum_{k=T}^N \sum_{\mathcal{X}^k \in \mathcal{X}_N^k} \left(\prod_{\mathbf{y} \in \mathcal{X}^k} \mathbf{P}(\phi_{uv}(\mathbf{x}, \mathbf{y}) \geq \tau(\mathbf{x}) | H_0) \right) \left(\prod_{\mathbf{y} \in \bar{\mathcal{X}}^k} (1 - \mathbf{P}(\phi_{uv}(\mathbf{x}, \mathbf{y}) \geq \tau(\mathbf{x}) | H_0)) \right)$$

where $N = |B(\mathbf{x})|$ is the search window size, \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. As the number of pixels increases in $B(\mathbf{x})$, the computation burden of (3.4) becomes daunting. In what follows, we assume that the probabilities $\mathbf{P}(\phi_{uv}(\mathbf{x}, \mathbf{y}) \geq \tau(\mathbf{x})|H_0)$ are identical¹ for all the pixels $\mathbf{y} \in B(\mathbf{x})$, i.e.

$$(3.5) \quad \mathbf{P}(\phi_{uv}(\mathbf{x}, \mathbf{y}) \geq \tau(\mathbf{x})|H_0) \triangleq p_{fa}(\mathbf{x}, \tau(\mathbf{x})).$$

In our change detection scenario, most patches overlap since $B(\mathbf{x})$ will be small. Hence, the decision statistic $S_N(\mathbf{x})$ is the sum of *dependent* Bernoulli random variables. Nevertheless, from the Central Limit Theorem (CLT) for dependent random variables [13] we know that $S_N(\mathbf{x})$ is asymptotically Gaussian. Indeed, the steady-state expectation of $S_N(\mathbf{x})$ under hypothesis H_0 is ($\mathbf{E}[\cdot]$ denotes the mathematical expectation)

$$(3.6) \quad \mathbf{E}[S_N(\mathbf{x})|H_0] = Np_{fa}(\mathbf{x}, \tau(\mathbf{x})),$$

which is also the expectation of the sum of independent random variables ; in [74] it has been also shown that the steady-state variance of $S_N(\mathbf{x})$ is upper-bounded as

$$(3.7) \quad \text{Var}[S_N(\mathbf{x})|H_0] \leq Np_{fa}(\mathbf{x}, \tau(\mathbf{x}))(1 - p_{fa}(\mathbf{x}, \tau(\mathbf{x}))).$$

It turns out that this upper-bound is also the variance for the sum of independent random variables. Therefore, $S_N(\mathbf{x})$ can be approximated as the sum of independent and identically distributed variables if N is large enough. Finally, the probability $P_{fa}(\mathbf{x}, T)$ defined in (3.4) can be approximated by the right tail of the binomial distribution², even if the neighbor pixels share common contextual information :

$$(3.9) \quad P_{fa}(\mathbf{x}, T) \triangleq \sum_{k=T}^N \binom{N}{k} (p_{fa}(\mathbf{x}, \tau(\mathbf{x})))^k (1 - p_{fa}(\mathbf{x}, \tau(\mathbf{x})))^{N-k}.$$

In practice, there is no way to derive an explicit form of $p_{fa}(\mathbf{x}, \tau(\mathbf{x}))$ at location \mathbf{x} . To overcome this difficulty, we propose to examine computable upper-bounds of $P_{fa}(\mathbf{x}, T)$ if there exist. First, applying the Chebyshev's concentration inequality which states that for a random variable Z , $\mathbf{P}(|Z - \mathbf{E}[Z]| \geq t) \leq \text{Var}[Z]/t^2$, $\forall t > 0$, the probability of false alarm is upper-bounded as follows :

$$(3.10) \quad \begin{aligned} P_{fa}(\mathbf{x}, T) &\leq \mathbf{P}(|S_N(\mathbf{x}) - Np_{fa}(\mathbf{x}, \tau(\mathbf{x}))| \geq (T - Np_{fa}(\mathbf{x}, \tau(\mathbf{x})))|H_0) \\ &\leq \frac{\text{Var}[S_N(\mathbf{x})|H_0]}{(T - Np_{fa}(\mathbf{x}, \tau(\mathbf{x})))^2} = \frac{Np_{fa}(\mathbf{x}, \tau(\mathbf{x}))(1 - p_{fa}(\mathbf{x}, \tau(\mathbf{x})))}{(T - Np_{fa}(\mathbf{x}, \tau(\mathbf{x})))^2}. \end{aligned}$$

The Chebyshev's inequality yields a bound which is known not to be small enough. Indeed, we need to know if $S_N(\mathbf{x})$ is strongly *concentrated* around $\mathbf{E}[S_N(\mathbf{x})|H_0]$. A

¹An approximation is given by $p_{fa}(\mathbf{x}, \tau(\mathbf{x})) \approx \frac{1}{N} \sum_{\mathbf{y} \in B(\mathbf{x})} \mathbf{P}(\phi_{uv}(\mathbf{x}, \mathbf{y}) \geq \tau(\mathbf{x})|H_0)$

²When N is large enough, (3.9) is usually calculated (from Central Limit Theorem) by using the Laplace-DeMoivre approximation :

$$(3.8) \quad \lim_{N \rightarrow \infty} P_{fa}(\mathbf{x}, T) \approx 1 - \text{CDF} \left(\frac{T - Np_{fa}(\mathbf{x}, \tau(\mathbf{x}))}{\sqrt{Np_{fa}(\mathbf{x}, \tau(\mathbf{x}))(1 - p_{fa}(\mathbf{x}, \tau(\mathbf{x})))}} \right)$$

where $\text{CDF}(\cdot)$ denotes the standard normal cumulative distribution function.

typical concentration result is the Chernoff's bound [29] on the tail of the binomial distribution :

LEMMA 3.1. *From the Chernoff's theorem [29], if $t(\mathbf{x}) \triangleq T - Np_{fa}(\mathbf{x}, \tau(\mathbf{x}))$ with $t(\mathbf{x}) > 0$, then the probability of false alarm corresponding to the tail of the binomial distribution is bounded as*

$$P_{fa}(\mathbf{x}, T) \leq e^{-2Nt(\mathbf{x})^2}.$$

Unfortunately, the Chebyshev's and Chernoff's upper bounds depend on the knowledge on $p_{fa}(\mathbf{x}, \tau(\mathbf{x}))$ which is not accessible for our purpose.

To compute an upper-bound, we can use the following lemma :

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

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

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 $\mathbf{E} \left[e^{S_N(\mathbf{x})} \right]$ can be empirically estimated as :

$$(3.11) \quad \mathbf{E} \left[e^{S_N(\mathbf{x})} \right] \approx \frac{1}{|V(\mathbf{x})|} \sum_{\mathbf{y} \in V(\mathbf{x})} e^{S_N(\mathbf{y})}$$

where $V(\mathbf{x}) \subseteq \Omega$ stands for an arbitrary set of pixels around \mathbf{x} .

Actually, since it is unrealistic to choose the thresholds $\tau(\mathbf{x})$ to achieve a desired local probability of false alarm, we will investigate strategies for estimating these spatially-varying parameters required to compute (3.11). In the next section, we present first several dissimilarity measures to be considered for robust change detection. We present also a conventional parametric testing method and adopt an usual Gaussian image modeling. The idea is to derive the thresholds $\tau(\mathbf{x})$ from the probability distribution functions of dissimilarity measure approximated by mixtures of Gaussians. At last, we will address the setting of parameters T , the patch size n and the sizes of the search window $B(\mathbf{x})$ and the neighborhood $V(\mathbf{x})$.

4. Multiscale probabilistic decision framework. The multiscale approach, which uses the property of patch repetitions across scales, is recommended traditionally to analyze several spatial contexts [43, 71, 38, 82, 34]. In our case, the multiscale strategy we adopt enables to reduce the number of false alarms.

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 $S_N(\mathbf{x}) = \sum_{\mathbf{y} \in B(\mathbf{x})} \mathbf{1}[\phi_{uv}(\mathbf{x}, \mathbf{y}) \geq \tau(\mathbf{x})]$, we make a neighborhood-wise decision $D(\mathbf{x}) = \mathbf{1}[S_N(\mathbf{x}) \geq T] \in \{0, 1\}$ 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

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

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 $\tau(\mathbf{x})$, thus in agreement in the neighborhood $B(\mathbf{x})$. By setting $T = N$ also justified in [95, 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 we define the set of binary random variables at each location \mathbf{x} as : $\{D_1(\mathbf{x}), \dots, D_L(\mathbf{x})\}$.

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 Bernoulli 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 Bernoulli 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 $D_1(\mathbf{x}), D_2(\mathbf{x}), \dots, D_L(\mathbf{x})$ 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 (3.9) for a given patch size $n_\ell = (2\ell + 1)^2$. The probability of success is different for each Bernoulli variable $D_\ell(\mathbf{x})$ and we define $\lambda_T(\mathbf{x}) \triangleq \sum_{\ell=1}^L P_{fa,\ell}(\mathbf{x}, T)$. Then $\sum_{\ell=1}^L D_\ell(\mathbf{x})$ tends to a Poisson law $\text{Po}(\lambda_T(\mathbf{x}))$ in distribution as $n_\ell \rightarrow \infty$ according to the Chen-Stein method [6] :

THEOREM 4.1. (Arratia (1989) [6]) *Let \mathcal{I} be a finite index set and let $\{X_a : a \in \mathcal{I}\}$ be indicator random variables. For a given index $a \in \mathcal{I}$, let \mathcal{B}_a a set of indices b such that X_a and X_b are dependent. Let $Y = \sum_{a \in \mathcal{I}} X_a$ and let Z be a Poisson random variable with mean $\lambda = \mathbf{E}[Y] = \sum_{a \in \mathcal{I}} \mathbf{E}[X_a]$. Then the total variation distance between Y and Z is at most $Q_1 + Q_2 + Q_3$ where*

$$\begin{aligned} Q_1 &\triangleq \sum_{a \in \mathcal{I}} \sum_{b \in \mathcal{B}_a} \mathbf{E}[X_a] \mathbf{E}[X_b], \\ Q_2 &\triangleq \sum_{a \in \mathcal{I}} \sum_{b \in \mathcal{B}_a, b \neq a} \mathbf{E}[X_a X_b], \\ Q_3 &\triangleq \sum_{a \in \mathcal{I}} \mathbf{E}[|\mathbf{E}[X_a | X_b : b \notin \mathcal{B}_a] - \mathbf{E}[X_a]|]. \end{aligned}$$

Q_1 measures the total size of the dependence neighborhoods, Q_2 measures how many dependent pairs are likely to arise, and Q_3 measures how far X_a is from being independent of $\{X_b : b \notin \mathcal{B}_a\}$.

Consequently, the probability of false alarm denoted as $P_{FA}(\mathbf{x}, L)$ at location \mathbf{x} is given by the Poisson tail with parameter $\lambda_T(\mathbf{x}) = \sum_{\ell=1}^L P_{fa,\ell}(\mathbf{x}, T)$:

$$(4.2) \quad P_{FA}(\mathbf{x}, L) \triangleq \mathbf{P} \left(\sum_{\ell=1}^L D_\ell(\mathbf{x}) > k_D(\mathbf{x}) | H_0 \right) = 1 - \sum_{k=0}^{k_D(\mathbf{x})} (\lambda_T(\mathbf{x}))^k \frac{e^{-\lambda_T(\mathbf{x})}}{k!}$$

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, we consider that the probability $P_{fa,\ell}(\mathbf{x}, N)$ can be bounded at each pixel as (if $T = N$)

$$(4.3) \quad P_{fa,\ell}(\mathbf{x}, N) \leq e^{-N} \mathbf{E} \left[e^{S_{N,\ell}(\mathbf{x})} \right] \approx \frac{1}{|V(\mathbf{x})|} \sum_{\mathbf{y} \in V(\mathbf{x})} e^{S_{N,\ell}(\mathbf{y}) - N}$$

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_ℓ , $1 \leq \ell \leq L$ and $V(\mathbf{x})$ stands for a neighborhood around \mathbf{x} . 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 $V(\mathbf{x})$ in order to both capture long range information and to detect unusual events when comparing two images. In the special case when $V(\mathbf{x}) \equiv \Omega$ and $\lambda_N(\mathbf{x}) \equiv \lambda_N = \sum_{\ell=1}^L P_{fa,\ell}(N)$ adopted in our experiments, we get

$$(4.4) \quad \begin{cases} P_{FA}(\mathbf{x}, L) &= 1 - \sum_{k=0}^{k_D(\mathbf{x})} (\lambda_N)^k \frac{e^{-\lambda_N}}{k!}, \\ \lambda_N &= \sum_{\ell=1}^L \frac{e^{-N}}{|\Omega|} \sum_{\mathbf{y} \in \Omega} e^{S_{N,\ell}(\mathbf{y})}. \end{cases}$$

Finally, we consider that a change occurs at pixel if

$$(4.5) \quad P_{FA}(\mathbf{x}, L) \leq \alpha(\mathbf{x}),$$

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 $0 < \alpha(\mathbf{x}) < 1$ and close to 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, 101, 88, 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, 101, 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 $H_\varepsilon(u, v) : \Omega \rightarrow \{0, 1\}$ be the final multiscale change detection map defined as :

$$(4.6) \quad H_\varepsilon(u, v)(\mathbf{x}) = \begin{cases} 1 & \text{if } P_{FA}(\mathbf{x}, L) \leq \varepsilon/|\Omega|, \\ 0 & \text{otherwise.} \end{cases}$$

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

Proof.

$$\mathbf{E} \left[\sum_{\mathbf{x} \in \Omega} H_\varepsilon(u, v)(\mathbf{x}) | H_0 \right] = \sum_{\mathbf{x} \in \Omega} \mathbf{E}[H_\varepsilon(u, v)(\mathbf{x}) | H_0]$$

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

□

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 :

$$(4.7) \quad \alpha(\mathbf{x}) = \alpha \triangleq \max \left(\frac{\varepsilon}{|\Omega|}, \inf_{\mathbf{x} \in \Omega} P_{FA}(\mathbf{x}, L) \right).$$

As mentioned earlier, the Bonferonni 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 [11], 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 multi-scale change detection procedure.

5.1. Estimation of spatially-varying thresholds. To apply the algorithm summarized in Fig.5.1, 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.* The PDF of distance is no longer considered and we examine the highest distances in very small neighborhoods $b(\mathbf{x}) \subseteq \Omega$.

To derive the adaptive thresholds for change detection, we postulate that all positive decisions correspond to 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

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

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 $\mathbf{E}[\|\mathbf{s}(\mathbf{x})\|_2^2] = \sigma_s^2$. 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 :

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,*

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

Proof. If we consider the usual Euclidean distance to evaluate the dissimilarity between image patches, we have

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

From the Chebyshev's inequality, we get $\mathbf{P}(\|\mathbf{s}(\mathbf{x})\|_2 \geq r_b) \leq \frac{\sigma_{\mathbf{s}}^2}{r_b^2} \leq \frac{nK^2\sigma_{\mathbf{s}}^2}{\tau_u(\mathbf{x})}$.

□

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 $r_b \rightarrow 0$ then $\tau(\mathbf{x}) \rightarrow 0$ and a change is likely detected at location \mathbf{x} in the image. From (5.1), it turns out that if the images u and v are two noisy versions of the same scene, no meaningful change detection occurs with a very high probability. 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 :

$$(5.2) \quad \begin{aligned} \tau(\mathbf{x}) &= \min(\tau_u(\mathbf{x}), \tau_v(\mathbf{x})), \\ \phi_{uv}(\mathbf{x}, \mathbf{y}) &\triangleq \min(\phi_{uv}(\mathbf{x}, \mathbf{y}), \phi_{vu}(\mathbf{x}, \mathbf{y})). \end{aligned}$$

5.2. Implementation. Our final algorithm is presented in Fig. 5.1 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 $|b(\mathbf{x})| \equiv |B(\mathbf{x})|$, 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 evaluate the performance of the proposed algorithm in section 7.

6. Theoretical properties. In this section, we examine some properties of the proposed detector.

6.1. Detection for low signal-to-noise ratios. Our multiscale change detection method can actually deal with low signal-to-noise ratios. Actually, the method

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)).

o 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

$$S_{N,\ell}(\mathbf{x}) = \sum_{\mathbf{y} \in B(\mathbf{x})} \mathbf{1}[\phi_{uu,\ell}(\mathbf{x}, \mathbf{y}) \geq \tau_\ell(\mathbf{x})].$$

3. Compute $P_{fa,\ell}(N) = \frac{1}{|\Omega|} \sum_{\mathbf{y} \in \Omega} e^{S_{N,\ell}(\mathbf{y}) - N}$.

o Compute $\lambda_N = \sum_{\ell=1}^L P_{fa,\ell}(N)$.

o The final decision at pixel $\mathbf{x} \in \Omega$ is defined as :

$$H_\varepsilon(u, v)(\mathbf{x}) = \mathbf{1}[P_{FA}(\mathbf{x}, L) \leq \alpha]$$

with $\alpha = \max(\varepsilon/|\Omega|, \inf_{\mathbf{x} \in \Omega} P_{FA}(\mathbf{x}, L))$ and $P_{FA}(\mathbf{x}, L) = 1 - \sum_{k=0}^{k_D(\mathbf{x})} (\lambda_N)^k \frac{e^{-\lambda_N}}{k!}$ where $1 \leq (k_D(\mathbf{x}) + 1) \leq L$ is the actual number of changes detected at location \mathbf{x} for the set of L patch sizes.

FIG. 5.1. *Asymmetric change detection algorithm between images u and v .*

is designed to avoid the detection of undesirable patterns due to noise, as we shall see in the experiments. In this section, we study the idealized situation : $u = u_0 + \epsilon$ and $v = v_0 + \eta$ where the underlying images u_0 and v_0 are piecewise constant and ϵ and η Gaussian noises with known variance σ^2 .

PROPOSITION 6.1. *Define*

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

where $\mathbf{y} \in B(\mathbf{x}) \subseteq \Omega$ and assume $\mathbf{P}(\phi_{uv}(\mathbf{x}, \mathbf{y}) \geq \tau(\mathbf{x}) | H_0) \equiv p_{fa}(\mathbf{x}, \tau(\mathbf{x})), \forall \mathbf{y} \in B(\mathbf{x})$. If $\sigma \rightarrow +\infty$, $0 < u_0 < +\infty$ and n large enough, we have

$$\mathbf{P}(S_N(\mathbf{x}) = N | H_0) \leq C^{-nN} \quad \text{with some } 1 < C < +\infty.$$

Proof. If $\sigma \rightarrow +\infty$, the contents of u_0 and v_0 are drowned in the noise and we have $\phi_{uv}(\mathbf{x}, \mathbf{y}) \approx \sum_{\mathbf{t} \in \mathbb{R}^2} \text{rect}_n(\mathbf{t}) (\epsilon(\mathbf{x} + \mathbf{t}) - \eta(\mathbf{y} + \mathbf{t}))^2, \mathbf{y} \in B(\mathbf{x})$. It follows that $\phi_{uv}(\mathbf{x}, \mathbf{y}) \sim 2\sigma^2 \chi_n^2$, $\mathbf{E}[\phi_{uv}(\mathbf{x}, \mathbf{y})] = 2n\sigma^2$ and $\text{Var}[\phi_{uv}(\mathbf{x}, \mathbf{y})] = 8n\sigma^4$ since ϵ and η are

i.i.d. Gaussian variables. Moreover, if σ tends to infinity, we have

$$\begin{aligned} \tau(\mathbf{x}) &\approx \max \left(\bar{\tau}(\mathbf{x}), \sup_{\mathbf{y} \in b(\mathbf{x})} \sum_{\mathbf{t} \in \mathbb{R}^2} \text{rect}_n(\mathbf{t}) (\epsilon(\mathbf{x} + \mathbf{t}) - \epsilon(\mathbf{y} + \mathbf{t}))^2 \right) \\ &\geq \max \left(\bar{\tau}(\mathbf{x}), \frac{1}{|b(\mathbf{x})|} \sum_{\mathbf{y} \in b(\mathbf{x})} \sum_{\mathbf{t} \in \mathbb{R}^2} \text{rect}_n(\mathbf{t}) (\epsilon(\mathbf{x} + \mathbf{t}) - \epsilon(\mathbf{y} + \mathbf{t}))^2 \right) \approx \max(\bar{\tau}(\mathbf{x}), 2n\sigma^2). \end{aligned}$$

Since $|b(\mathbf{x})|$ is constant $\forall \mathbf{x} \in \Omega$, we have

$$\begin{aligned} \bar{\tau} &= \frac{1}{|\Omega|} \sum_{\mathbf{x} \in \Omega} \inf_{\mathbf{y} \in b(\mathbf{x})} \phi_{uu}(\mathbf{x}, \mathbf{y}) \\ &\leq \frac{1}{|\Omega|} \sum_{\mathbf{x} \in \Omega} \frac{1}{|b(\mathbf{x})|} \sum_{\mathbf{y} \in b(\mathbf{x})} \phi_{uu}(\mathbf{x}, \mathbf{y}) = \frac{1}{|b(\mathbf{x})|} \sum_{\mathbf{y} \in b(\mathbf{x})} \frac{1}{|\Omega|} \sum_{\mathbf{x} \in \Omega} \phi_{uu}(\mathbf{x}, \mathbf{y}) \\ &\approx \frac{1}{|b(\mathbf{x})|} \sum_{\mathbf{y} \in b(\mathbf{x})} \mathbf{E} \left[\sum_{\mathbf{t} \in \mathbb{R}^2} \text{rect}_n(\mathbf{t}) (\epsilon(\mathbf{x} + \mathbf{t}) - \epsilon(\mathbf{y} + \mathbf{t}))^2 \right] = 2n\sigma^2. \end{aligned}$$

It follows that $\tau(\mathbf{x}) \geq 2n\sigma^2$.

It is worth noting that $(\phi_{uv}(\mathbf{x}, \mathbf{y}) - n)/\sqrt{2n} \sim \mathcal{N}(0, 1)$ if n is large enough. Define $\tau(\mathbf{x}) \triangleq 2n\sigma^2\delta(\mathbf{x})$ subject to $\inf_{\mathbf{x} \in \Omega} \delta(\mathbf{x}) > \frac{1}{2\sigma^2}$. From the exponential Chebychev's applied to normal variables, we have

$$\begin{aligned} \mathbf{P}(\phi_{uv}(\mathbf{x}, \mathbf{y}) \geq \tau(\mathbf{x}) | H_0) &= \mathbf{P} \left(\frac{\phi_{uv}(\mathbf{x}, \mathbf{y}) - n}{\sqrt{2n}} > \frac{\tau(\mathbf{x}) - n}{\sqrt{2n}} | H_0 \right) \leq \exp \left(- \frac{(\tau(\mathbf{x}) - n)^2}{4n} \right) \\ &\leq \left(\exp \left(\frac{(2\sigma^2\delta(\mathbf{x}) - 1)^2}{4} \right) \right)^{-n} \\ &\leq \left(\exp \left(\frac{(2\sigma^2 \inf_{\mathbf{x} \in \Omega} \delta(\mathbf{x}) - 1)^2}{4} \right) \right)^{-n}. \end{aligned}$$

It follows that

$$\mathbf{P}(S_N(\mathbf{x}) = N | H_0) = (\mathbf{P}(\phi_{uv}(\mathbf{x}, \mathbf{y}) \geq \tau(\mathbf{x}) | H_0))^N \leq C^{-nN},$$

with $C = \exp \left(\frac{(2\sigma^2 \inf_{\mathbf{x} \in \Omega} \delta(\mathbf{x}) - 1)^2}{4} \right) > 1$ and we obtain the required assertion. \square

This bound is small provided that nN is large enough. Consequently, with a probability of at least $1 - C^{-nN}$, $u(\mathbf{x})$ coincides with $v(\mathbf{x})$ at location \mathbf{x} if $\sigma \rightarrow +\infty$ and for a given patch size n . From (3.6)-(3.10), it follows that $\lambda_N = \sum_{\ell=1}^L P_{fa,\ell}(N) \rightarrow 0$ and consequently $P_{FA}(\mathbf{x}, L) \rightarrow 1, \forall \mathbf{x} \in \Omega$ if σ is very large. Hence, no change occurs at location \mathbf{x} since $P_{FA}(\mathbf{x}, L) \rightarrow 1 > \alpha$ based on the aforementioned assumptions and the two images are ‘‘similar’’. In short, no change is expected if the two images are corrupted by white Gaussian noise whose variance σ^2 is the same and tends to infinity.

6.2. Detection for high signal-to-noise ratios. Now we discuss the situation when $\sigma \rightarrow 0$. We intend to show that if the contrast of the image is sufficiently large compared with the level of noise then we typically obtain positive decisions for all pixels $\mathbf{y} \in B(\mathbf{x})$. To simplify the presentation, we suppose that u_0 and v_0 are piecewise constant images in regions but have different values : $u_0(\mathbf{x}) = a_1$ and $v_0(\mathbf{x}) = a_2, \forall \mathbf{x} \in A(\mathbf{x})$ subject to $a_1 \neq a_2$ where $A(\mathbf{x}) \subset \Omega$ is a set of pixels. We assume that $A(\mathbf{x})$ is a sufficiently large region and we impose $|a_1 - a_2|^2 > \tau(\mathbf{x})/n$

when σ tends to 0. This implies that the neighborhood $b(\mathbf{x})$ belongs completely to $A(\mathbf{x})$.

PROPOSITION 6.2. *Assume $u_0(\mathbf{x}) = a_1$ and $v_0(\mathbf{x}) = a_2, \forall \mathbf{x} \in A(\mathbf{x})$ subject to $a_1 \neq a_2$ and $b(\mathbf{x}) \subset A(\mathbf{x}) \subset \Omega$ such that $|a_1 - a_2|^2 > \tau(\mathbf{x})/n$. It follows that*

$$\lim_{\sigma \rightarrow 0} (\mathbf{P}(S_N(\mathbf{x}) = N | H_0) = (\mathbf{P}(\phi_{uv}(\mathbf{x}, \mathbf{y}) \geq \tau(\mathbf{x}) | H_0))^N) \rightarrow 1.$$

Proof. If $\sigma \rightarrow 0$, the geometrical structure and texture of u and v dominate over noise, and we have $u(\mathbf{x}) \approx a_1$ and $v(\mathbf{x}) \approx a_2$. Accordingly, we have

$$\phi_{uv}(\mathbf{x}, \mathbf{y}) = \sum_{\mathbf{t} \in \mathbb{R}^2} \text{rect}_n(\mathbf{t}) (u(\mathbf{x} + \mathbf{t}) - v(\mathbf{y} + \mathbf{t}))^2 \approx n|a_1 - a_2|^2 > \tau(\mathbf{x}), \forall \mathbf{y} \in B(\mathbf{x}).$$

Hence all the decisions at locations $\mathbf{y} \in B(\mathbf{x})$ are positive. With these assumptions, the probability $(\mathbf{P}(\phi_{uv}(\mathbf{x}, \mathbf{y}) \geq \tau(\mathbf{x}) | H_0))^N$ tends to 1 and the assertion is proved. \square

The result of proposition 6.2 leads to the following conclusion. Based on the the previous assumptions, $\lambda_N = \sum_{\ell=1}^L P_{fa,\ell}(N) \rightarrow L$ for any patch size and $P_{FA}(\mathbf{x}, L) \rightarrow 0, \forall \mathbf{x} \in \Omega$ if σ tends to 0. Therefore, since $P_{FA}(\mathbf{x}, L) \leq \alpha$ holds true for any value $\alpha > 0$, a change occurs at location \mathbf{x} . In short, if the two images are piecewise constant and the regions have different values and areas large enough, a change is likely detected at each pixel in the image.

6.3. 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 have³

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

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

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

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

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

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 $\nu \in [0, 2\pi]$:

$$H_\varepsilon(R_\nu \circ u, R_\nu \circ v) = R_\nu \circ H_\varepsilon(u, v).$$

³*Sketch of proof :* Assume with loss of generality $\bar{\tau} = 0$. Hence, we have

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

- **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. 5.1 is then modified using (5.2) 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, blotch detection in old digitized movies and symmetry detection in MRI imaging. In this section, we present results on 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. 5.1. 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. $B(\mathbf{x}) \equiv b(\mathbf{x})$) except in Section 6.5. 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 visually assessed by comparison with the ground truth [100]. The method is robust to undesirable motions in the background corresponding to trees shaken by the wind. In Fig. 7.1, 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 ($T = N$ for the highest (hot) value). In these two typical examples, the masks are quite regularized with no hole because of the sliding window process and patch overlapping. Figure 7.1 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. 7.2-7.5, other examples of change detection are shown for video-surveillance scenarios for which the

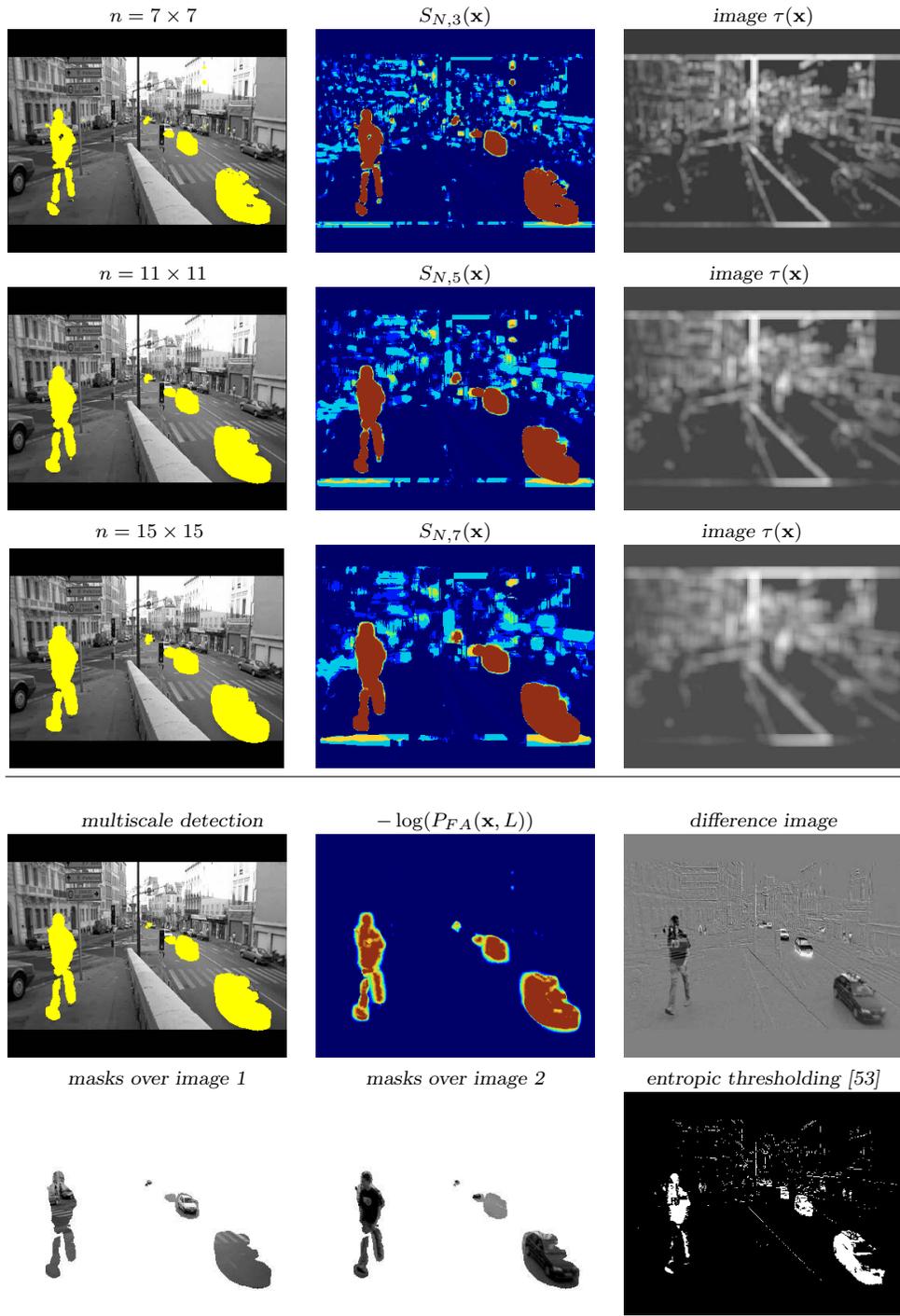


FIG. 7.1. The images $S_{N,\ell}(\mathbf{x})$ correspond to the number of positive decisions in 3×3 search windows $B(\mathbf{x})$ for several patch sizes and scales : $n_3 = 7 \times 7$, $n_5 = 11 \times 11$ and $n_7 = 15 \times 15$ respectively. The level sets corresponding to $T = N$ are superimposed 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].

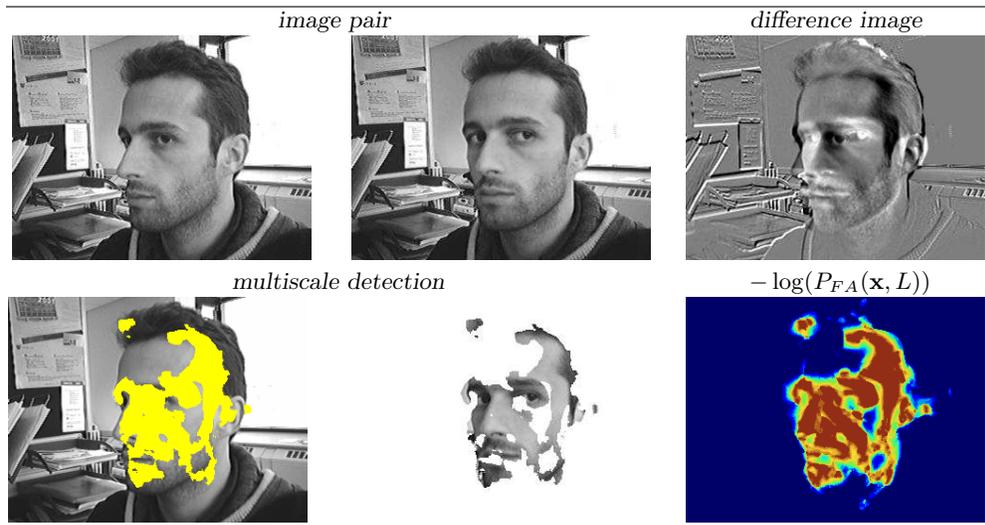


FIG. 7.2. Occlusion detection in an image pair using 5×5 search windows ($L = 15$).

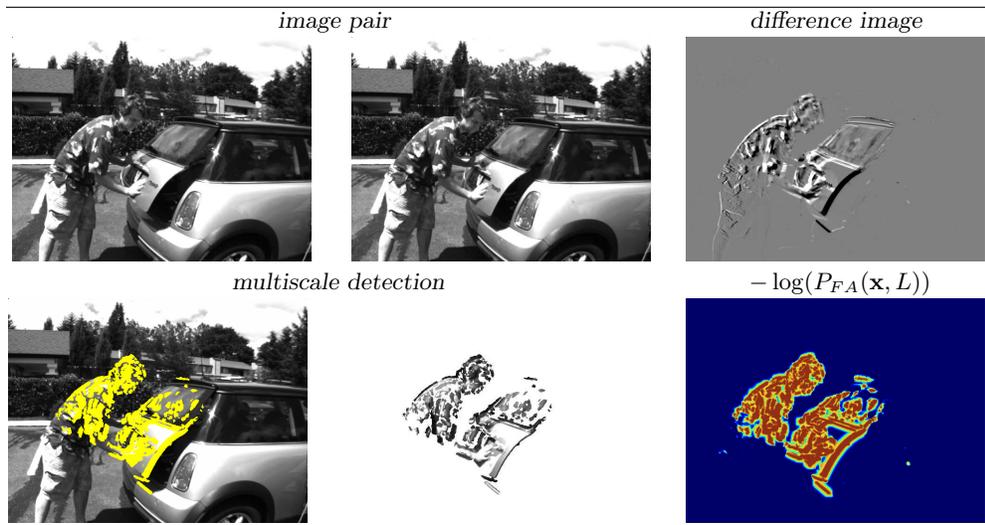


FIG. 7.3. Occlusion detection in an image pair using 5×5 search windows ($L = 15$) (see <http://vision.middlebury.edu/flow/data/>).

illumination conditions are relatively stable.

In Fig. 7.6 and Fig. 7.7, the image pairs are 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 [12, 98]. 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

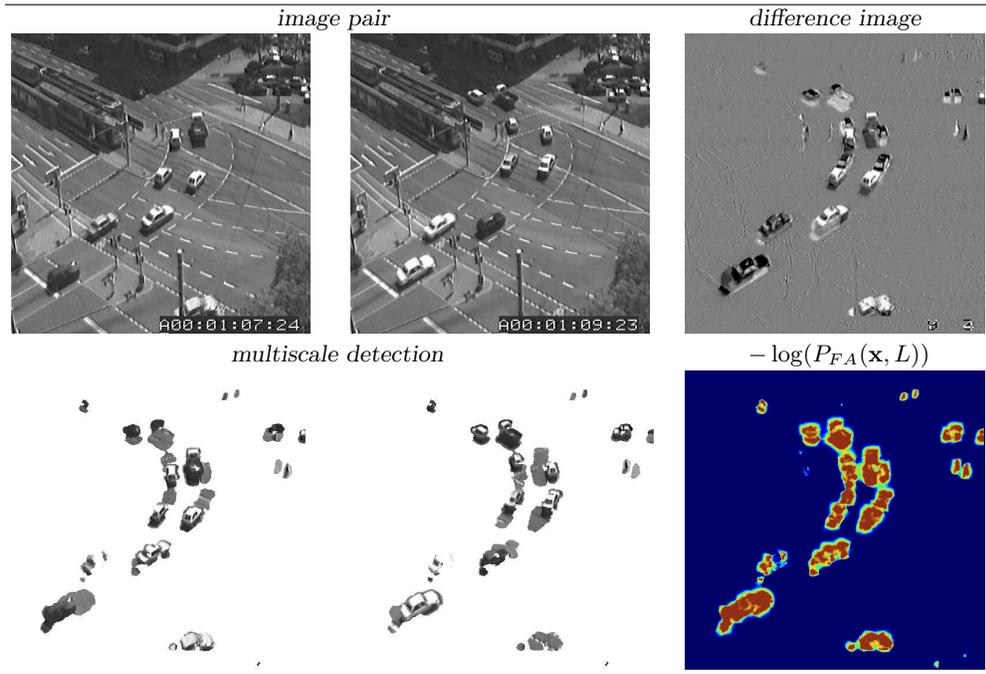


FIG. 7.4. Occlusion detection in a traffic scene (images at time $t = 1$ and $t = 50$) using 5×5 search windows ($L = 12$) (see http://i21www.ira.uka.de/image_sequences/).

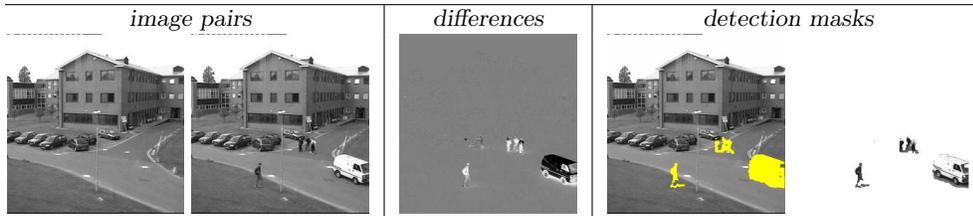


FIG. 7.5. Change detection results by using the L_2 dissimilarity measure (2.5). Top : $L = 5, \hat{\lambda}_N = 0.1$; bottom : $L = 3, \hat{\lambda}_N = 0.13$ (PETS 2006 dataset).

corrupted as well but the temporal coherence is not preserved. In this experiments, we considered small neighborhoods (3×3 square windows) and we found $\hat{\lambda}_N = 0.050$ and $\hat{\lambda}_N = 0.047$ respectively for the two tested images. We analyzed the images in both directions and the algorithm given in Fig. 5.1 has been then modified accordingly using (5.2). 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 repairing and 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 shown in Fig. 2.1. Figs. 7.8-7.9 show 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.

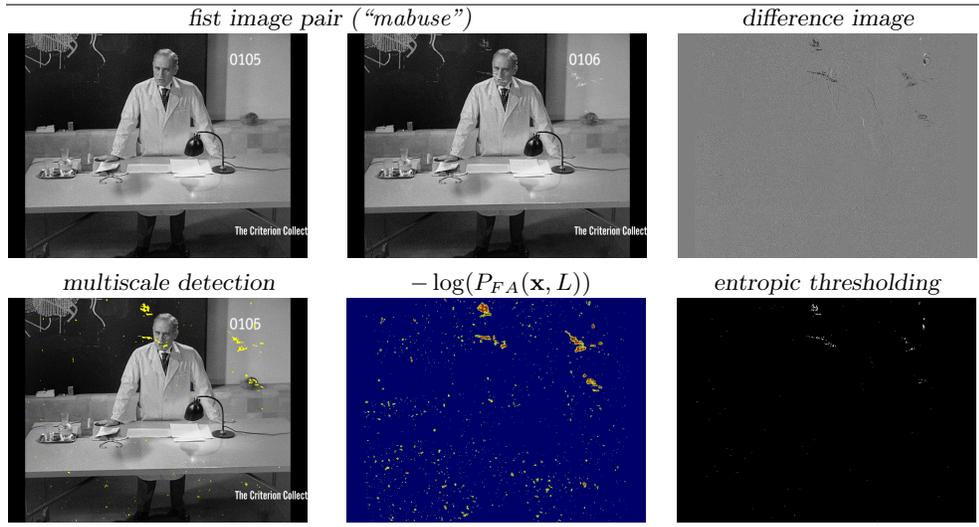


FIG. 7.6. Blotch detection in an image pair from an old movie ($L = 3$). The histogram of the absolute difference image was thresholded with the Kapur's entropic method [53].

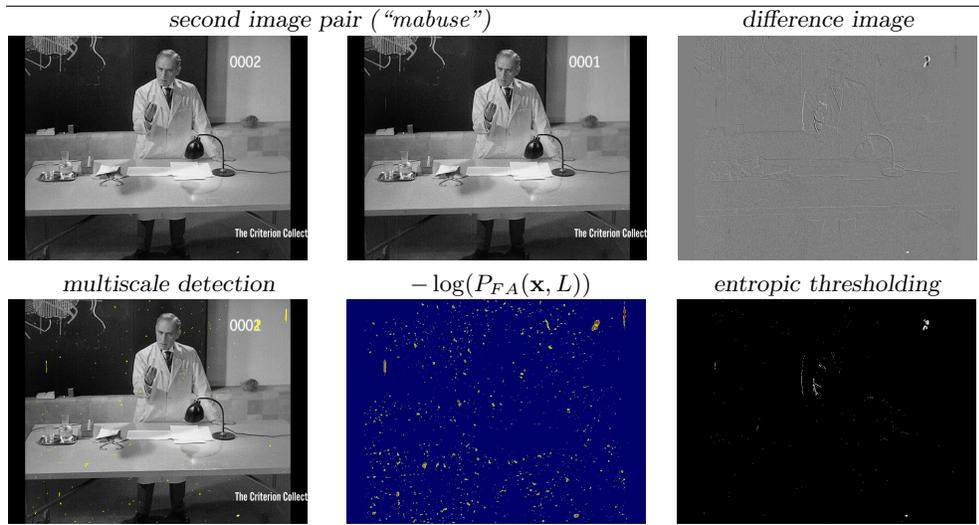


FIG. 7.7. Blotch detection in an image pair from an old movie ($L = 3$). The histogram of the absolute difference image was thresholded with the Kapur's entropic method [53].

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. 7.10 shows a snowy traffic scene with low contrast. We applied the usual algorithm to detect the car in the background. 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

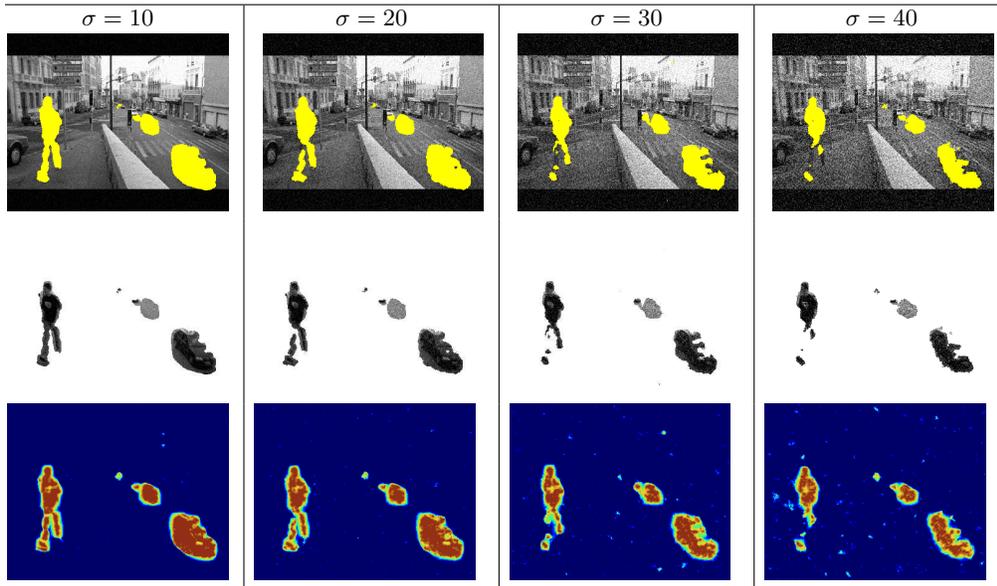


FIG. 7.8. *Robustness to white Gaussian noise. The two first rows show the detection results and the third row shows the probabilities of false alarm ($-\log(P_{FA}(\mathbf{x}, L))$).*

are not the same in the image as assumed in Section 3.2, we proposed successfully this simple modification. In Fig. 7.10, 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 a first example (Fig. 7.11), 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 (2.3) since illumination 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). In this application, we have considered also the dissimilarity measure (2.4) (correlation coefficient) since it represents an alternative to (2.3). In practice, the dissimilarity measure (2.3) is more flexible since we can adapt the photometry invariance by adjusting ρ unlike the correlation criterion (2.4) which is parameter free.

7.2. Comparisons with state-of-the-art methods. We compare our approach to commonly-used and more recent background subtraction methods [93, 41, 109, 33, 79, 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 the algorithm. We considered the usual “RECALL” and “PRECISION” rates defined in the range $[0, 1]$ as :

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

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

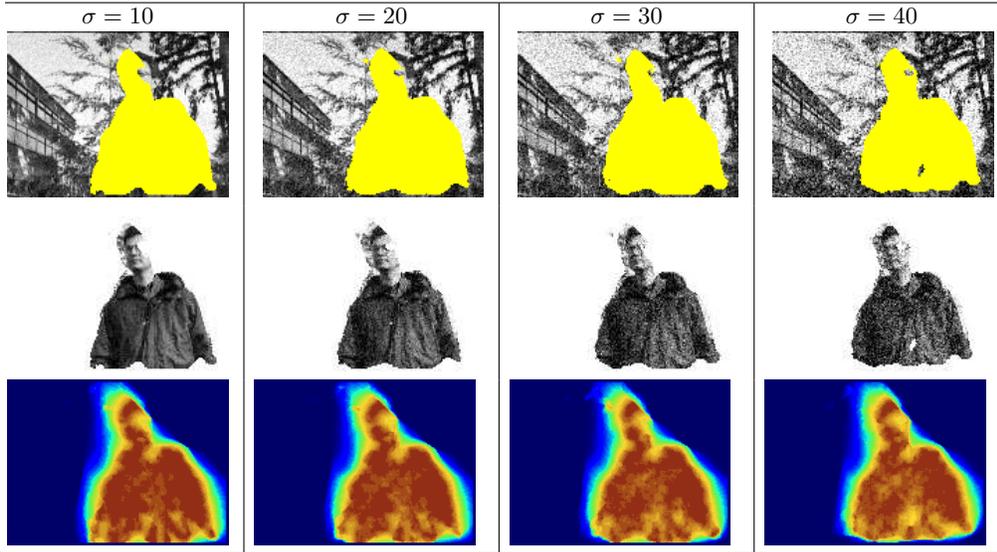


FIG. 7.9. *Robustness to white Gaussian noise. The two first rows show the detection results and the third row shows the probabilities of false alarm ($-\log(P_{FA}(\mathbf{x}, L))$).*

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 7.12 and 7.13 show two image pairs extracted from the “light switch” and “bootstrapping” benchmarks described in [100]. 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 a single reference image when compared to the state-of-the-art methods exploiting series of temporal images [100] or color information [79]. By setting $\rho = 1$, the method is relatively robust to sudden illumination (see Fig. 7.12, $N = 9, L = 10, \hat{\lambda}_N = 3.04$) and shadow effects (see Fig. 7.13, ($N = 9, L = 15, \hat{\lambda}_N = 4.69$)) 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. 7.12 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 [100, 46]. The potential of our method shown in Fig. 7.14 is satisfying when compared to the state-of-the-art background subtraction methods which can handle illumination changes and moving objects [100, 41, 93, 72, 109, 33, 46, 79]. Our method did the same job using two input images and sometimes outperformed several methods [93, 41, 2, 46, 79] which require a long image sequence as input. Nevertheless, background subtraction cannot be used to detect changes

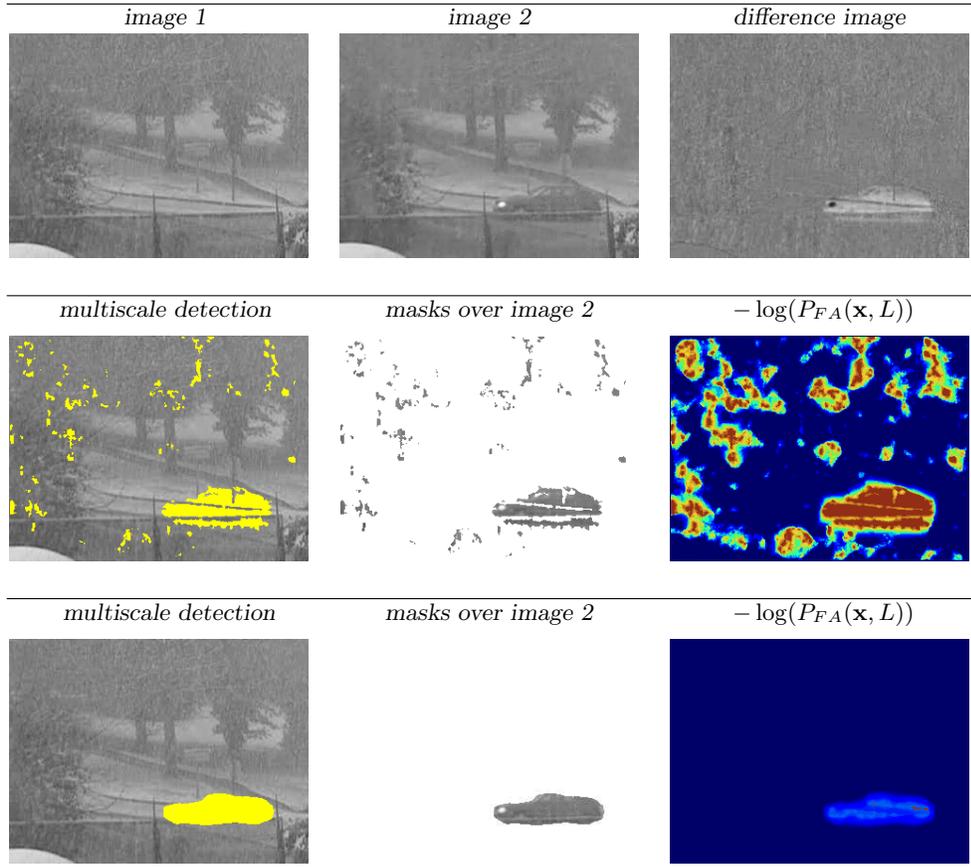


FIG. 7.10. Car detection in a snowy traffic scene. The second row corresponds to detection performed as previously ; the third row shows the detection results by considering a unique threshold (see text).

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 [100]).

In Fig. 7.15 (third row), the masks of moving objects are estimated by considering respectively 50 frames and a mixture of three Gaussians [93] (third row, middle), and 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. 7.15 because of the uniform appearance 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 frame ($t = 84$) to an arbitrarily chosen reference frame (Fig. 7.15, 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 to the reference frame showing only the background. Additionally we observe resid-

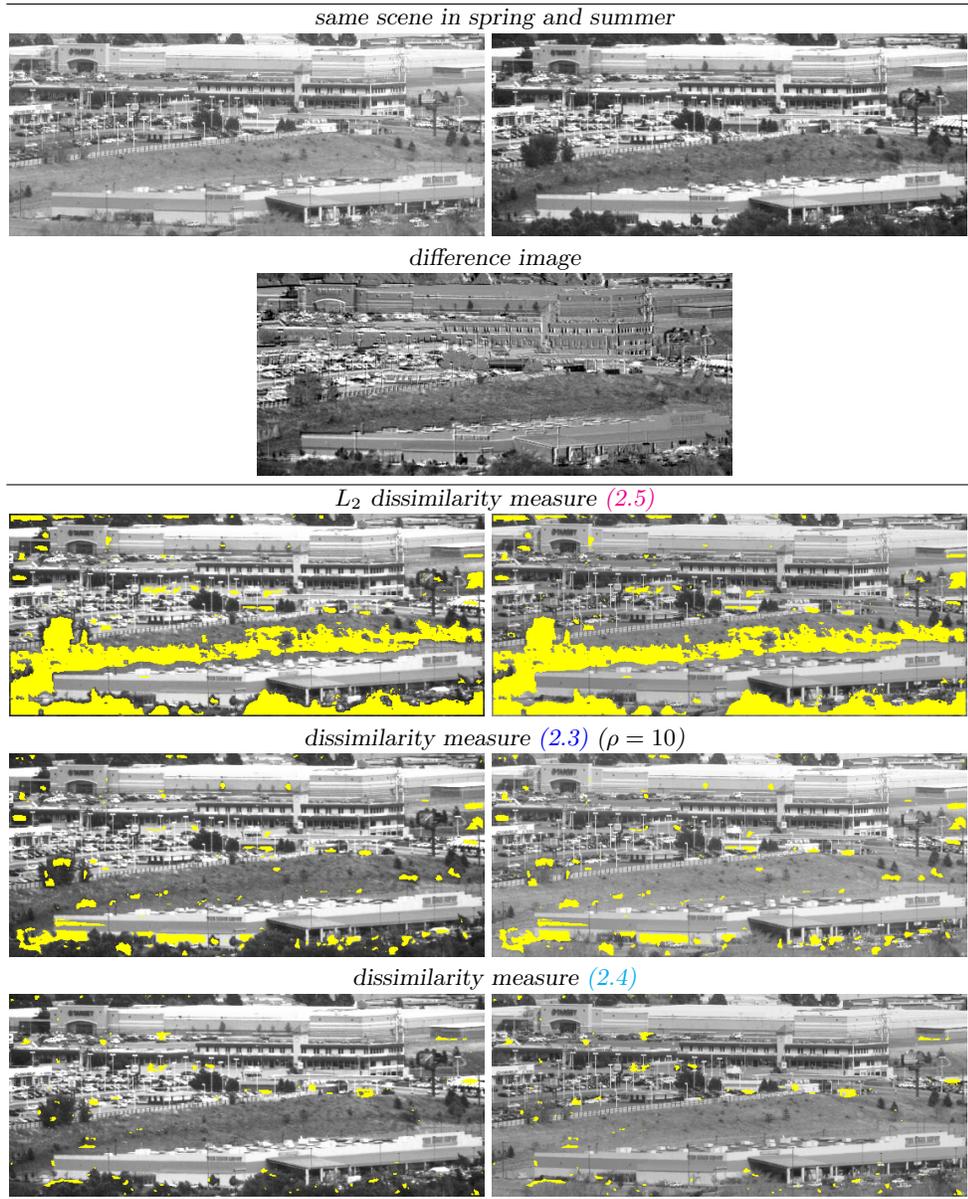


FIG. 7.11. Change detection with several dissimilarity measures in an outdoor scene at two different dates in the year (spring and summer) (5×5 search windows and $L = 7$).

ual 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, 33] is shown in Fig. 7.16.

In Fig. 7.18, we compare our results to those obtained by Zivkovic & van der Heijden [109] and Pilet *et al.* [79] on three images taken from an image sequence

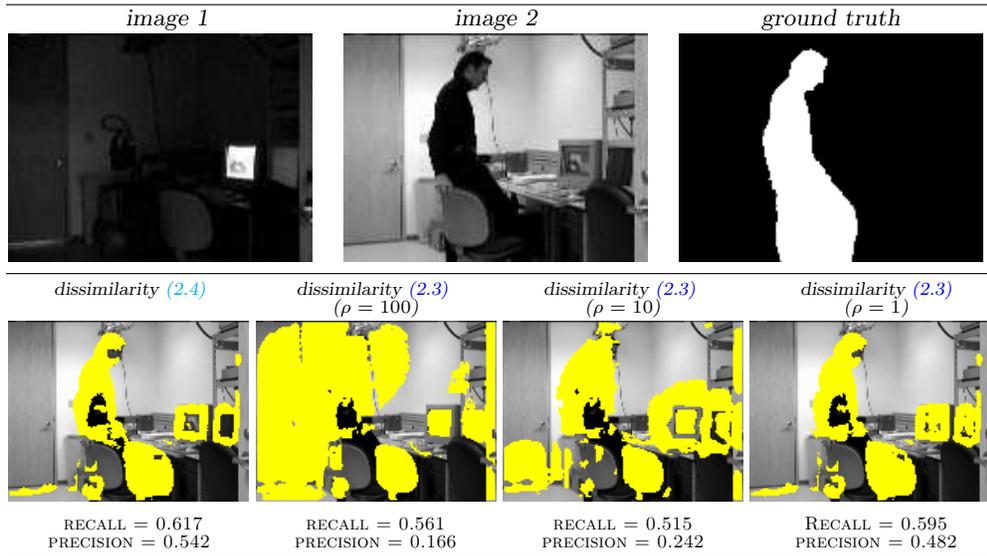


FIG. 7.12. Robustness to sudden illumination changes with two different dissimilarity measures and different values of ρ ($\varepsilon = 100$) (see [100, 46, 79] for comparison).

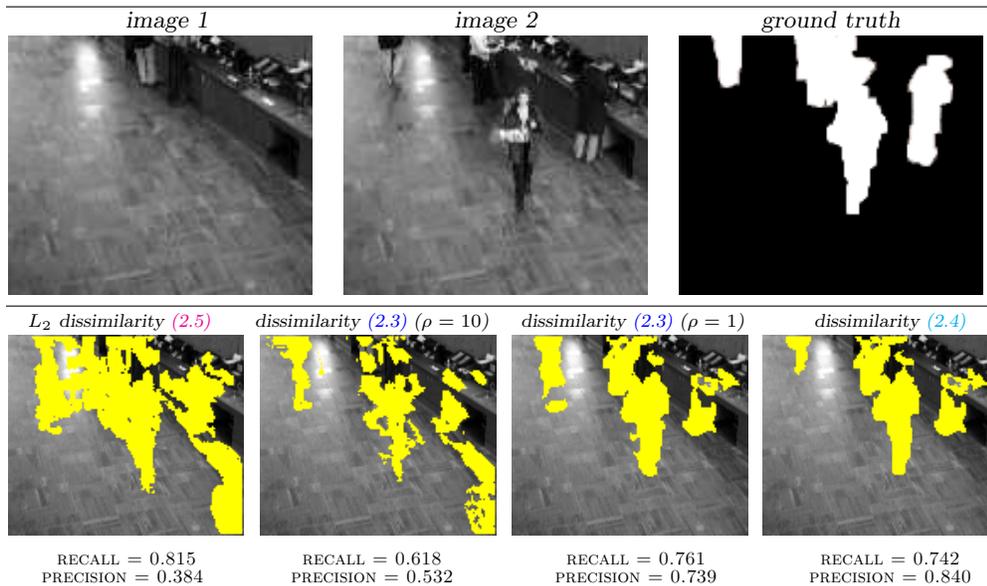


FIG. 7.13. Robustness to shadows and specularities with the dissimilarity measures (2.3) ($\varepsilon = 10$) and (2.5) and (2.4) ($\varepsilon = 1$.) (see [100, 46] for comparison).

from the PETS 2006 dataset. Unlike these two methods, our method exploits only the reference image shown on top row of Fig. 7.18. The persons are reliably detected and the results are very similar to those obtained in [79]. Finally, we compared our method to the three layer MRF method [9, 10] applied to an image pair for which the ground truth is available. In Fig. 7.17, our algorithm missed mainly a non-contrasted

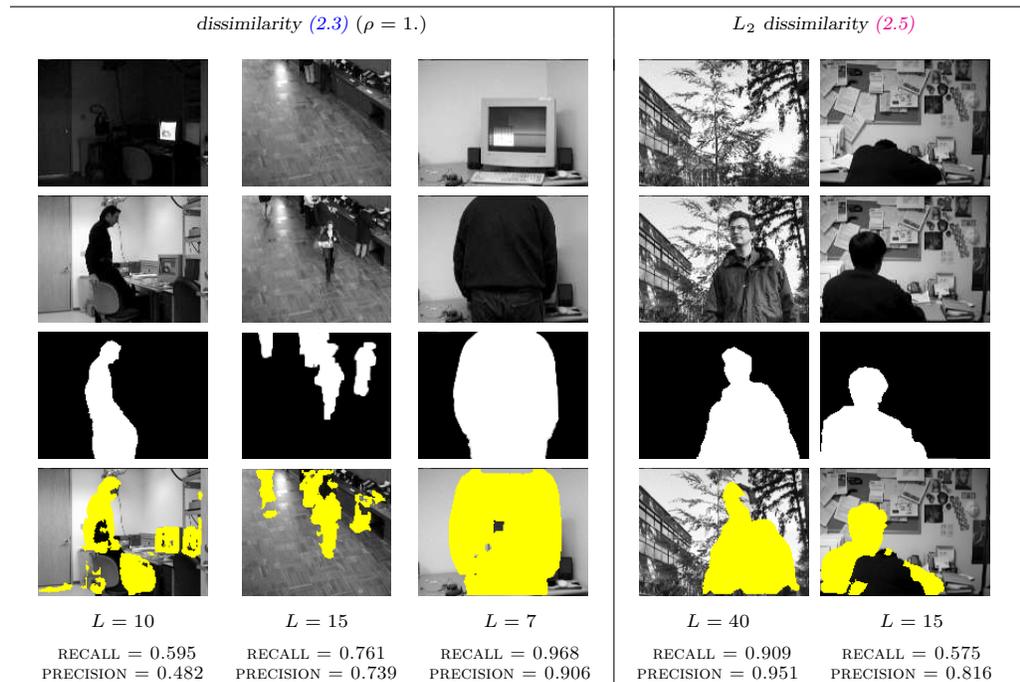


FIG. 7.14. Change detection masks computed by the proposed method for the image pairs tested in [100, 46] (see also [79]). Two first rows : image pairs ; third row : ground truths ; fourth row : our detection results with the dissimilarity measures (2.3) and (2.5).

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. 7.19-7.20 focus on the problem of detecting space-time occlusions in consecutive frames as in [64, 108, 52]. On the classic “flower garden” sequence, 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. 7.19-7.20, 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

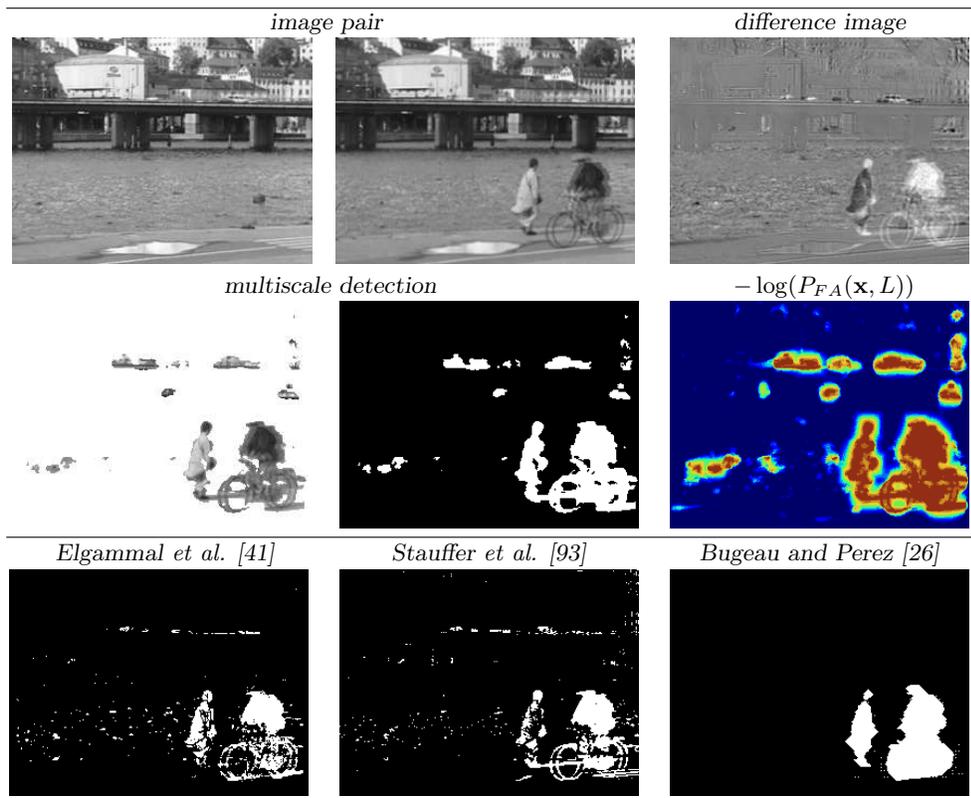


FIG. 7.15. Comparison of background subtraction methods with our approach that utilizes a single reference frame showing the background with no moving object.

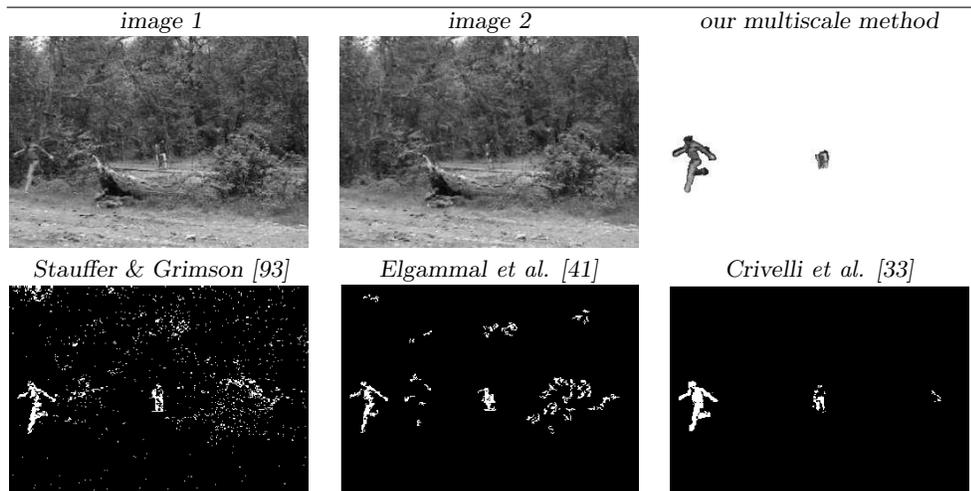


FIG. 7.16. Comparison with three background subtraction methods using several frames from the analyzed image sequence [41, 33] and our multiscale change detection method that uses only a single reference frame shown on top left (L_2 dissimilarity measure : $L = 5, \hat{\lambda}_N = 0.1$).

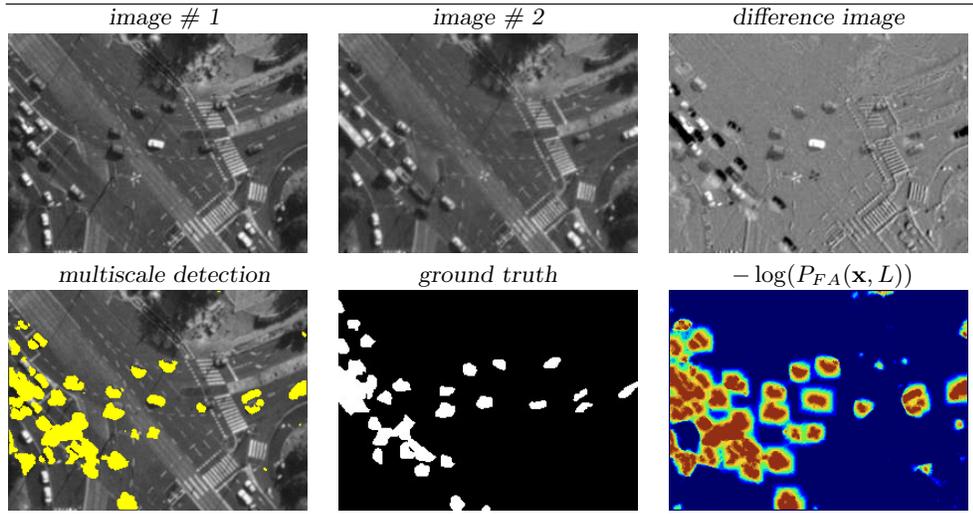


FIG. 7.17. Occlusion detection in a traffic scene using 3×3 search windows ($L = 12$, RECALL = 0.759, PRECISION = 0.616) (<http://web.eee.sztaki.hu/bcsaba/aerialObjectMotionBenchmark.htm>) (see [9, 10]).

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.

Let Θ the set of configurations $\theta = (\theta(\mathbf{x}))_{\mathbf{x} \in \Omega}$ with $\theta(\mathbf{x}) \in \{0, 1\}$ for all $\mathbf{x} \in \Omega$. First, we consider the following cost functional [104] :

$$(7.1) \quad J_{uv}^{entropy}(\theta) = \sum_{\mathbf{x} \in \Omega} c(\mathbf{x})(t_e - u(\mathbf{x}) + v(\mathbf{x})) + \sum_{\langle \mathbf{x}, \mathbf{y} \rangle} \psi(\theta(\mathbf{x}), \theta(\mathbf{y})).$$

where $c(\mathbf{x}) = (2 \times \mathbb{1}[\theta(\mathbf{x}) = 1] - 1) \in \{-1, 1\}$. The first term measures the distance between $(u(\mathbf{x}) - v(\mathbf{x}))$ and a threshold t_e and weights the distances at each location \mathbf{x} by plus or minus one according to whether $\theta(\mathbf{x}) = 1$ or $\theta(\mathbf{x}) = 0$ as proposed in [104]. Thus, pixels where $\theta(\mathbf{x}) = 1$ but $(u(\mathbf{x}) - v(\mathbf{x})) < t_e$ will contribute positively to the global energy. The value t_e is the prior threshold automatically found according to the (entropy-based) Kapur’s method [53] but other methods (EM algorithms) could be also investigated. The second term takes the form of an Ising MRF (Markov Random Field) model. It is defined on pairs $\langle \mathbf{x}, \mathbf{y} \rangle$ of neighboring pixels (with respect to

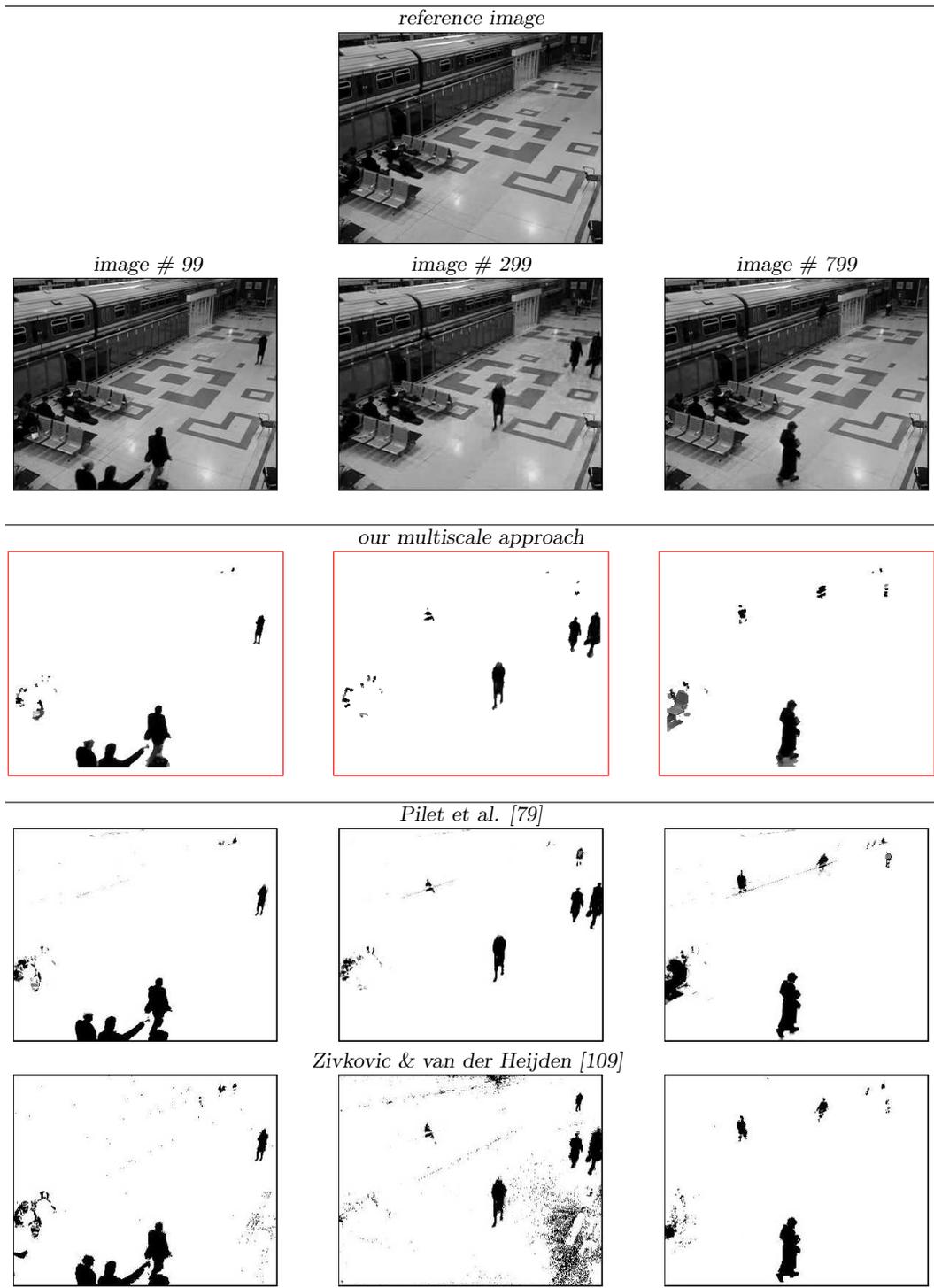


FIG. 7.18. Detection on three images from the PETS 2006 dataset. The methods [109, 79] exploit the image sequence unlike our method which uses only a single reference image (top row).

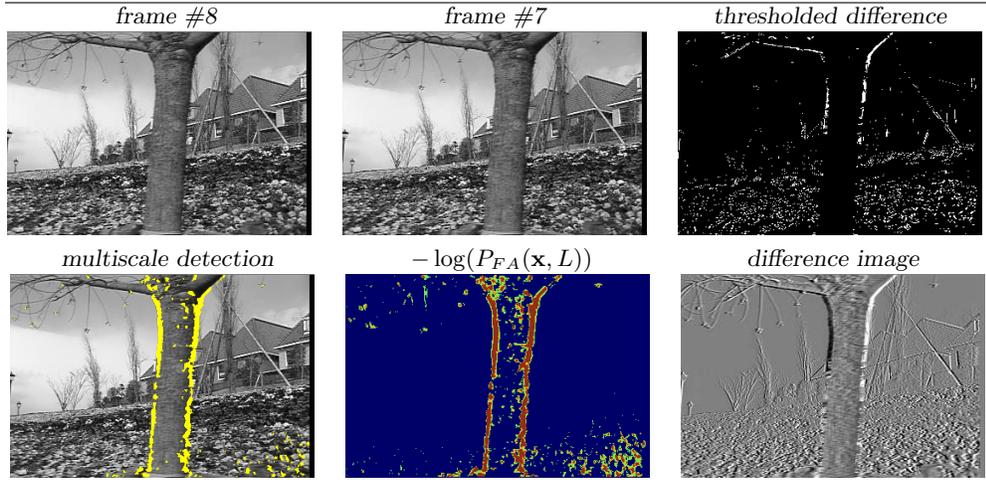


FIG. 7.19. Detection of spatio-temporal discontinuities in “flower garden” (3×3 neighborhoods, 3×3 search windows, $L = 3$). The discontinuity areas (superimposed on the original frame #7) correspond to the highest level set of the image $-\log(P_{FA}(\mathbf{x}, L))$.

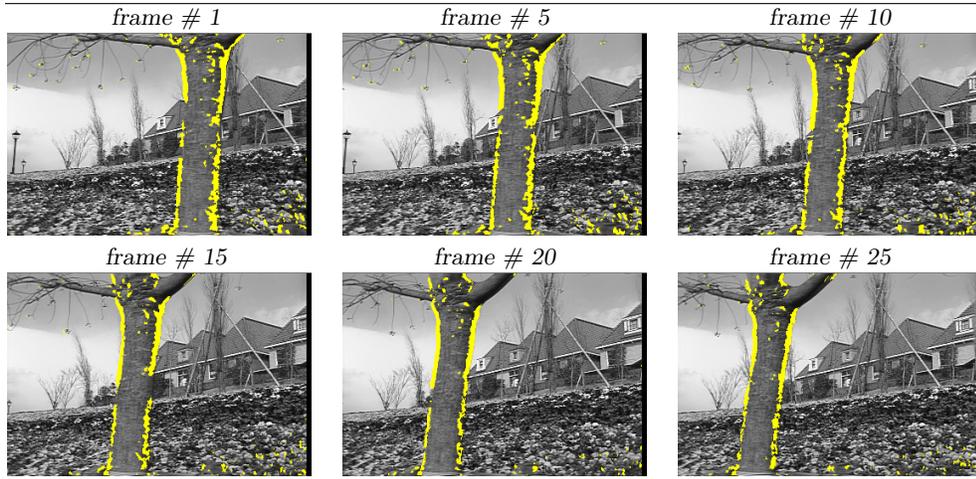


FIG. 7.20. Space-time discontinuities in “flower garden”.

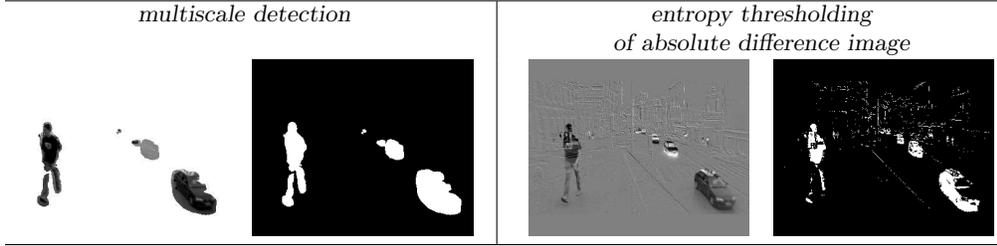
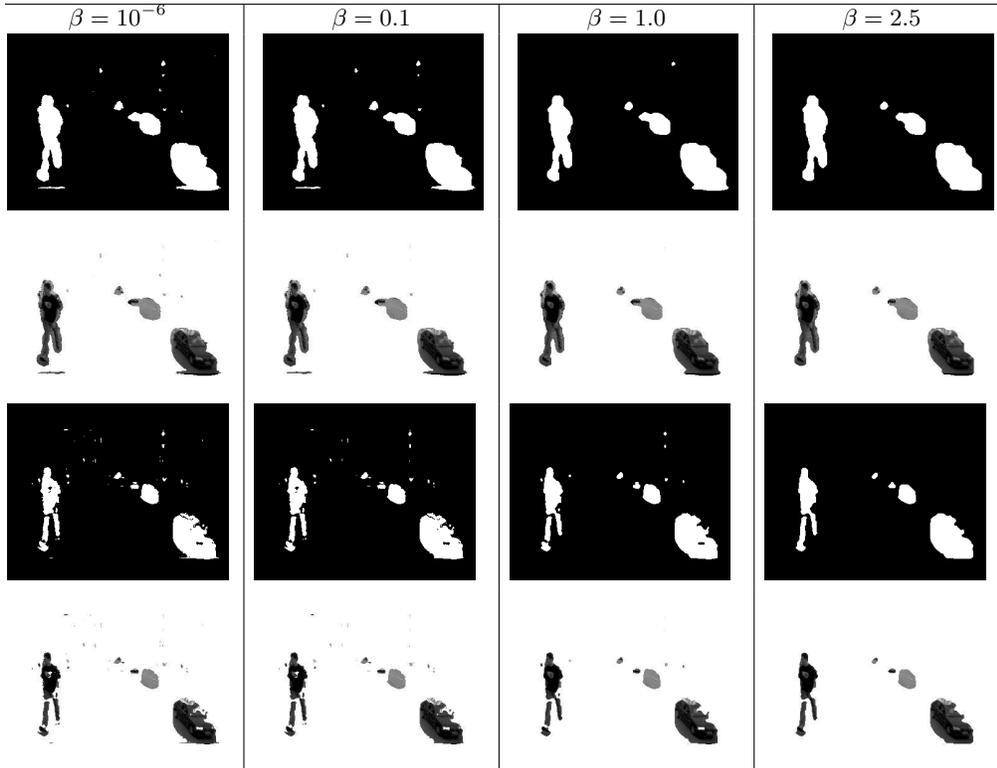
4-neighborhood system) as

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

where $\beta > 0$ is the balance parameter manually adjusted. This parameter has significant effects on the qualitative properties of the minimizer as shown in Fig. 7.23.

For more objective comparisons, we also consider the following patch-based cost functional [57] :

$$(7.3) \quad J_{uv}^{count}(\theta) = \sum_{\mathbf{x} \in \Omega} \sum_{\ell=1}^L a_{\ell} \left| \frac{1}{N} S_{N,\ell}(\mathbf{x}) - \theta(\mathbf{x}) \right| + \sum_{\langle \mathbf{x}, \mathbf{y} \rangle} \psi(\theta(\mathbf{x}), \theta(\mathbf{y}))$$

FIG. 7.21. Our multiscale detector ($L = 11$) applied to the image pair shown in Fig. 7.1.FIG. 7.22. Detection results for different values of β which controls the patch-based energy functional $J_{uv}^{count}(\theta)$ (image pair shown in Fig. 7.1). First and second rows : $a'_\ell = 1, \ell = 1, \dots, L$; third and fourth rows : $a'_\ell = (2\ell + 1)^{-1}, \ell = 1, \dots, L$.

where the first term 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 and $a_\ell = \frac{a'_\ell}{\sum_{\ell=1}^L a'_\ell}$ 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 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 is similar to the regularization term (Ising model) introduced in $J_{uv}^{count}(\theta)$.

On the image pair shown in Fig. 7.1, we present the detection results obtained

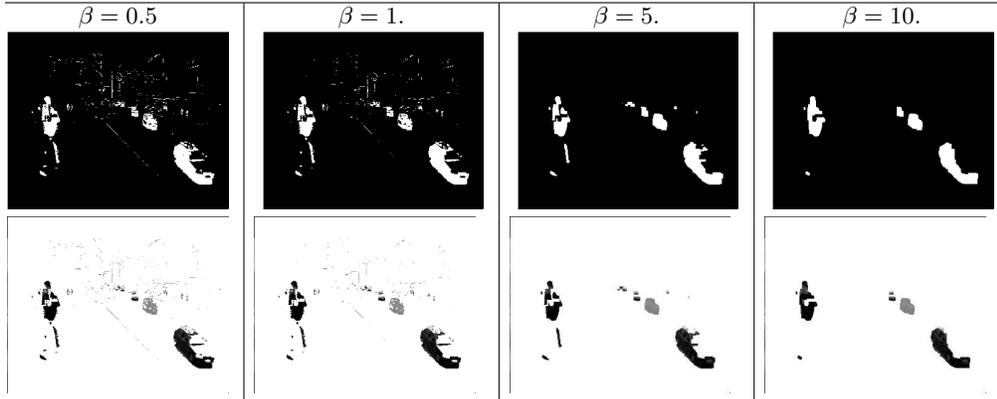
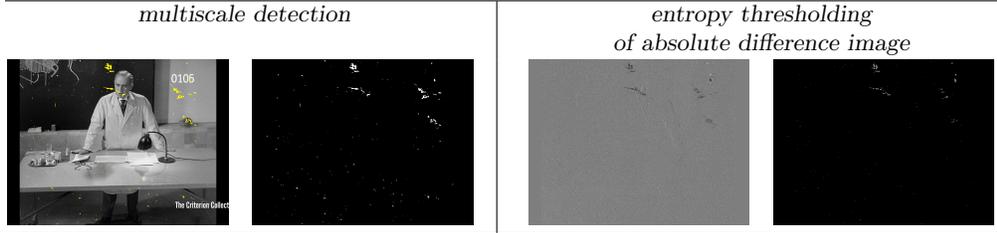
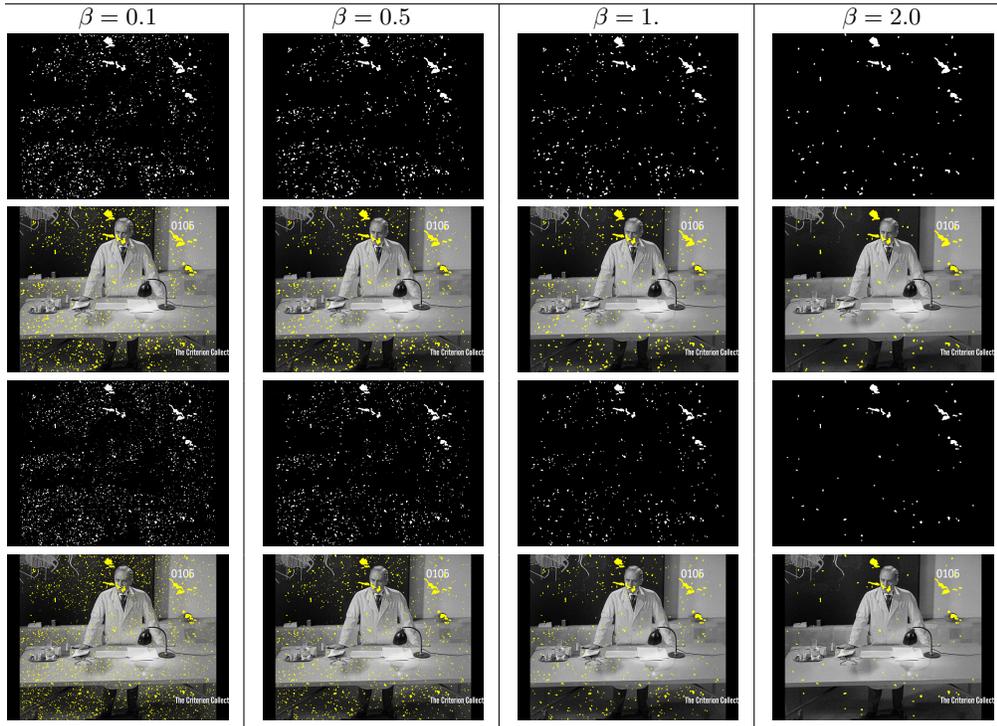


FIG. 7.23. Detection results for different values of β which controls the entropy-based energy functional (7.1) (image pair shown in Fig. 7.1).

respectively by automatic thresholding of the absolute image difference (Fig. 7.21 (right)) and by MRF-based labeling and graph cut minimization [21] of the aforementioned energy (Fig. 7.23). In this experiment, the threshold t_e is over-estimated and the expected and meaningful occluded objects are not recovered. In the two other studied examples, it was not possible to find appropriate thresholds (even set manually) to extract the occluded regions. Accordingly, we focused on the second cost functional $J_{uv}^{count}(\theta)$ since it exploits equally the counting maps $\{S_{N,\ell}\}$ obtained for different patch sizes $n_\ell, l = 1, \dots, L$. In Fig. 7.22, more regular masks are estimated by increasing β but small undesirable objects are also extracted for all tested β values. Moreover, rough blobs are generally isolated as it is confirmed on additional examples and applications shown in Figs. 7.25-7.27. The object boundaries are better delineated using our multiscale framework (Figs. 7.24-7.26). Actually imposing more regularity by increasing β tends to simultaneously decrease the perimeter of number of connected components and the number of connected components. Intuitively β is an hyperparameter which can be thought as controlling the number of false alarms.

Nevertheless, a compromise is not easily achieved to produce results visually similar to those obtained using our testing method which globally analyzes the occurrence of dependent counts for different patch sizes (see Figs. 7.26, 7.24 and 7.27). It turns out that the cost functional (7.3) is not equivalent to our probabilistic detection approach. In addition, a reliability measure expressed in terms of false alarm rates is provided by using our approach unlike the energy-based counterpart. To overcome these difficulties, more sophisticated cost functionals and Conditional Random Fields (CRF) models including edge terms could be designed as proposed in [59, 96, 52, 97]. Considering the maps of probability of false alarm within a CRF framework (see for instance [33]) is also an opportunity we plan to investigate in future work.

7.5. Asymmetry detection. The algorithm has been also applied to a single 2D MRI image where a tumor is observed on the left hand side of the brain (Fig. 7.28). In this experiment, we flipped one original image at its vertical axis to get a second image. We run the detector on the left half of the original image. In Fig. 7.28, the detected pixels on the left image correspond to regions with no correspondence on the right hand side on the original image. We tested several search window sizes in order to extract the meaningful regions (Fig. 7.29). The algorithm given in Fig. 5.1 is

FIG. 7.24. Our multiscale detector ($L = 3$) applied to the image pair shown in Fig. 7.6.FIG. 7.25. Detection results for different values of β which controls the patch-based energy functional $J_{uv}^{count}(\theta)$ (image pair shown in Fig. 7.6). First and second rows : $a_\ell = 1, \ell = 1, \dots, L$; third and fourth rows : $a_\ell = (2\ell + 1)^{-1}, \ell = 1, \dots, L$.

modified using (5.2) to analyze the dissimilarity in both directions. In this experiment $|b(\mathbf{x})| = 3 \times 3$ and $|B(\mathbf{x})|$ varies from 3×3 to 15×15 . It is worth noting that the human brain is only approximatively symmetric, and point-to-point comparisons between the left and right hand sides do not yield satisfactory results ; many occluded pixels are detected because of asymmetry/deformation. In Fig. 7.29, our approach produces similar detection masks if the search windows is higher than 5×5 pixels. In future work, we plan to compare these results to those obtained by more dedicated methods [88, 90] in medical imaging.

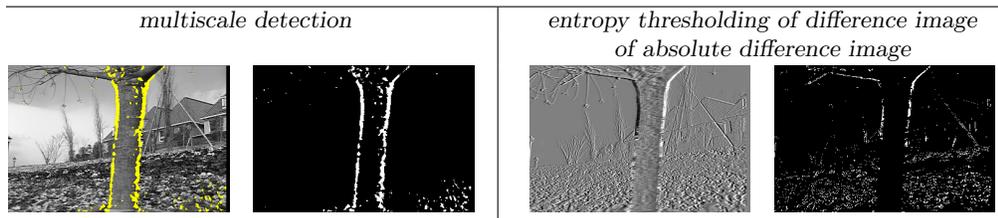


FIG. 7.26. Our multiscale detector ($L = 3$) applied to the image pair (“flower garden”) shown in Fig. 7.19.

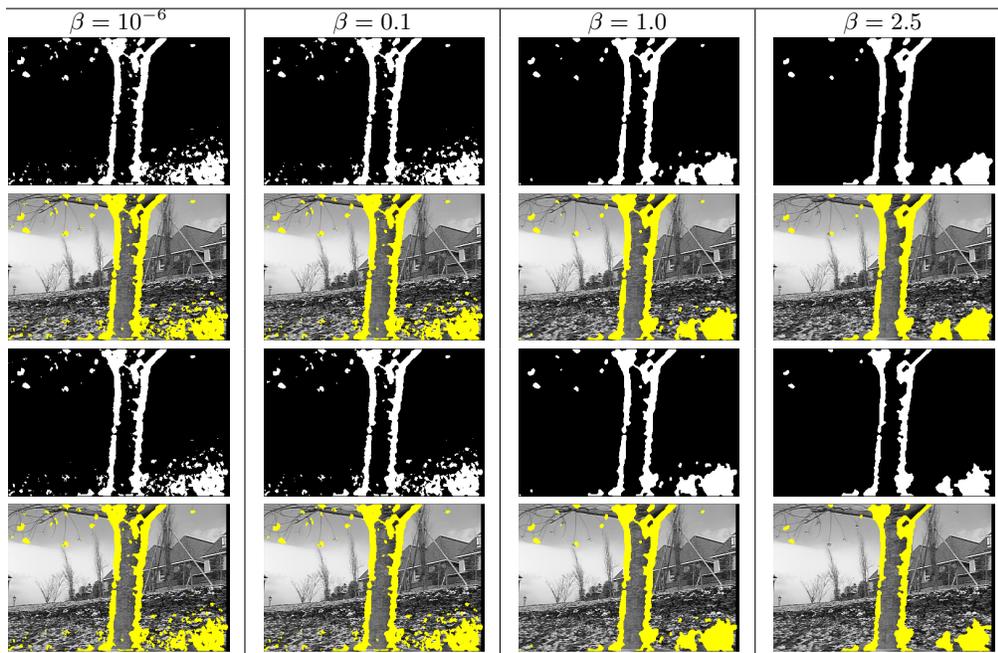


FIG. 7.27. Detection results for different values of β which controls the patch-based energy functional $J_{uv}^{\text{count}}(\theta)$ (image pair shown in Fig. 7.19). First and second rows : $a_\ell = 1, \ell = 1, \dots, L$; third and fourth rows : $a_\ell = (2\ell + 1)^{-1}, \ell = 1, \dots, L$.

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 in 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

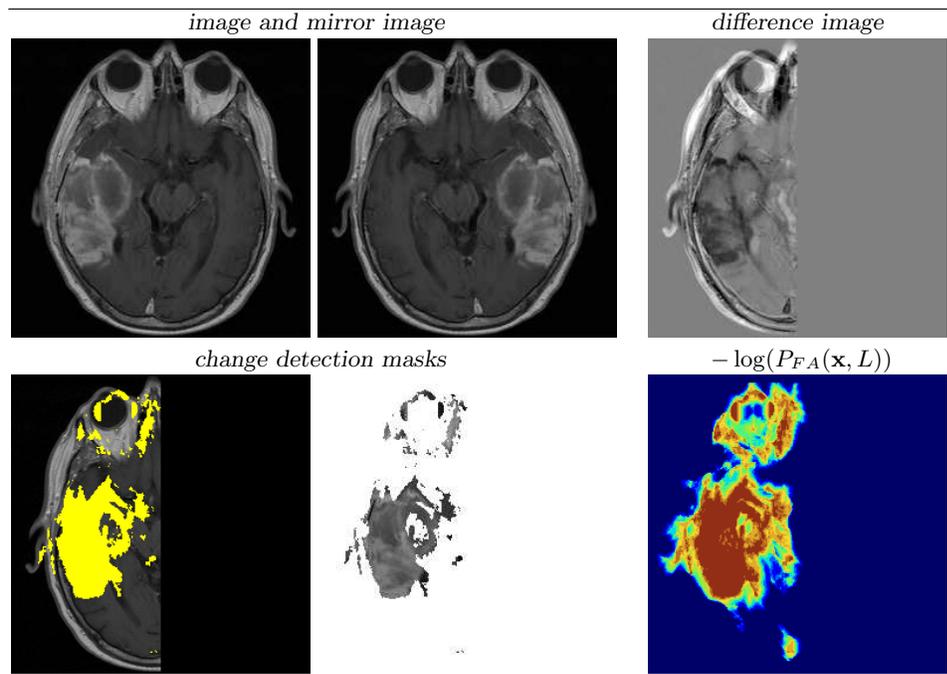


FIG. 7.28. *Asymmetry detection in a MRI image ($L = 10$, 15×15 search windows).*

an image sequence with static camera, our method, using only two input images, does as well as and sometimes better than methods [93, 41, 2, 79] 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 in future work. We did not address the stereo problem yet since the displacements are traditionally large.

We are convinced that all these technical contributions are meaningful in detection theory and image analysis. In future works, we plan to adapt this framework to the comparison of several images instead of two images.

REFERENCES

- [1] T. AACH AND A. CONDURACHE, *Transformation of adaptive thresholds by significance invariance for change detection*, in Proc. of SSP 2005, Bordeaux, July 17-20 2005.
- [2] T. AACH, L. DUMBGEN, R. MESTER, AND D. TOTH, *Bayesian illumination-invariant motion detection*, in Proc. of ICIP'01 (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 Proc. of CVPR'06 (1), Washington, DC, USA, 2006, pp. 798–805.
- [5] L. ALVAREZ, R. DERICHE, T. PAPADOPOULOU, AND J. SANCHEZ, *Symmetrical dense optical flow estimation with occlusions detection*, in Proc. of ECCV'02 (1), Copenhagen, Denmark,

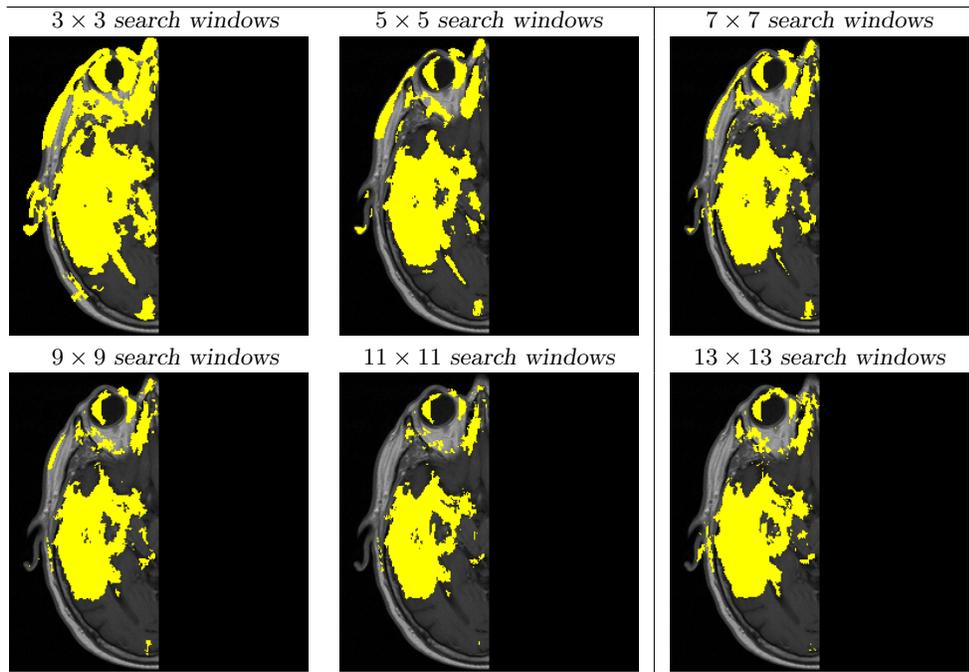


FIG. 7.29. Asymmetry detection in a MRI image for different search window sizes ($L = 10$).

- 2002, pp. 721–735.
- [6] R. ARRATIA, L. GOLDSTEIN, AND L. GORDON, *Two moments suffice for poisson approximations: the chen-stein method*, *Annals of Probabilities*, 17 (1989), pp. 9–25.
- [7] S.P. AWATE AND R.T. WHITAKER, *Unsupervised, information-theoretic, adaptive image filtering for image restoration*, *IEEE Trans. Pattern Anal. Mach. Intell.*, 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, p. 4155.
- [9] C. BENEDEK, T. SZIRÁNYI, Z. KATO, AND J. ZERUBIA, *A multi-layer MRF model for object-motion detection in unregistered airborne image-pairs*, in *Proc. of ICIP'07*, vol. 6, San Antonio, Texas, USA, September 2007, pp. 141–144.
- [10] Cs. BENEDEK, T. SZIRÁNYI, Z. KATO, AND J. ZERUBIA, *Detection of object motion regions in aerial image pairs with a multi-layer markovian model*, *IEEE Trans. Image Processing*, (2009). In Press.
- [11] Y. BENJAMINI AND Y. HOCHBERG, *Controlling false discovery rate: a practical and powerful approach to multiple testing*, *J. Royal Stat. Soc. B*, 57 (1995), pp. 289–300.
- [12] J. BIEMOND, P.M.B. VAN ROOSMALEN, AND R.L. LAGENDIJK, *Improved blotch detection by postprocessing*, in *Proc. of ICASSP'99* (6), Washington, DC, 1999, pp. 3101–3104.
- [13] P. BILLINGSLEY, *Probability and Measure*, John Wiley and Sons, New-York, NY, 1979.
- [14] M.J. BLACK AND D.J. FLEET, *Probabilistic detection and tracking of motion boundaries*, *Int. J. Comput. Vision*, 38 (2000), pp. 231–245.
- [15] O. BOIMAN AND M. IRANI, *Detecting irregularities in images and in video*, in *Proc. of ICCV'05* (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. BOSCH, 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 CH. KERVRANN, *Repetitive and transient event detection in fluorescence video-microscopy*, in Focus on Microscopy (FOM'08), Osaka-Awaji, Japan, April 2008.
- [20] J. BOULANGER, C. KERVRANN, AND P. BOUTHEMY, *Space-time adaptation for patch based image sequence restoration*, IEEE Trans. on Pattern Anal. Mach. Intell., 29 (2007), pp. 1096–1102.
- [21] Y. BOYKOV, O. VEKSLER, AND R. ZABIH, *Markov random fields with efficient approximations*, in Proc. of CVPR'98, Santa Barbara, CA, 1998, pp. 648–655.
- [22] T. BROX, O. KLEINSCHMIDT, AND D. CREMERS, *Efficient nonlocal means for denoising of textural patterns*, IEEE Trans. Image Processing, 17 (2008), pp. 1083–1092.
- [23] 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.
- [24] A. BUADES, B. COLL, AND J.M. MOREL, *Nonlocal image and movie denoising*, Int. J. Comput. Vision, 76 (2008), pp. 123–139.
- [25] A. BUADES, B. COLL, AND J. M. MOREL, *A review of image denoising algorithms, with a new one*, Multiscale Modeling and Simulation, 4 (2005), pp. 490–530.
- [26] A. BUGEAU AND P. PÉREZ, *Detection and segmentation of moving objects in complex scenes*, Comput. Vis. Image Underst., 113 (2009), pp. 459–476.
- [27] F. CAO AND P. BOUTHEMY, *A general principled method for image similarity validation*, in Proc. of AMR'2006, Geneva, Switzerland, July 2006, pp. 57–70.
- [28] V. CASELLES, J.L. LISANI, J.-M. MOREL, AND G. SAPIRO, *Shape preserving local histogram modification*, IEEE Trans. on Image Processing, 8 (1999), pp. 220–230.
- [29] H. CHERNOFF, *A measure of the asymptotic efficiency for tests of a hypothesis based on the sum of observables*, Annals Math. Stat., 23 (1952), pp. 493–509.
- [30] I. J. COX, S. ROY, AND S. L. HINGORANI, *Dynamic histogram warping of image pairs for constant image brightness*, in Proc. of ICIP'95, vol. 2, Washington, DC, USA, 1995, pp. 366–369.
- [31] A. CRIMINISI, G. CROSS, A. BLAKE, AND V. KOLMOGOROV, *Bilayer segmentation of live video*, in Proc. of CVPR'06 (1), Washington, DC, USA, 2006, pp. 53–60.
- [32] 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.
- [33] 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 Proc. of ECCV'08 (1), Marseille, France, 2008, pp. 113–126.
- [34] M. IRANI D. GLASNER, S. BAGON, *Super-resolution from a single image*, in Proc. of ICCV'09, Kyoto, Japan, 2009.
- [35] Y. DENG, Q. YANG, X. LIN, AND X. TANG, *A symmetric patch-based correspondence model for occlusion handling*, in Proc. of ICCV'05 (2), Beijing, China, 2005, pp. 1316–1322.
- [36] A. DESOLNEUX, L. MOISAN, AND J.-M. MOREL, *Meaningful alignments*, Int. J. Comput. Vision, 40 (2000), pp. 7–23.
- [37] F. DIBOS AND G. KOEPLER S. PELLETTIER, *Adapted windows detection of moving objects in video scenes*, Imaging Science, 2 (2009), pp. 1–19.
- [38] M. EBRAHIMI AND EDWARD R. VRSCAY, *Examining the role of scale in the context of the non-local-means filter*, in LNCS Volume 5112, Image analysis and Recognition, vol. 5112, Springer Berlin, Heidelberg, 2008, pp. 170–181.
- [39] A.A. EFROS AND T.K. LEUNG, *Texture synthesis by non-parametric sampling*, in Proc. of ICCV'99 (2), Kerkyra, Greece, 1999, pp. 1033–1038.
- [40] M. ELAD AND M. AHARON, *Image denoising via learned dictionaries and sparse representation*, in Proc. of CVPR'06 (1), New York, NY, 2006, pp. 895–900.
- [41] A. ELGAMMAL, D. HARWOOD, AND L. DAVIS, *Non-parametric model for background subtraction*, in Proc. of ECCV'00, 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 prob.*, in Proc. of CVPR'06, Washington, DC, USA, 2006, pp. 300–307.
- [43] W.T. FREEMAN, E.C. PASZTOR, AND O.T. CARMICHAEL, *Learning low-level vision*, Int. J. Comput. Vision, 40 (2000), pp. 25–47.
- [44] D. GEMAN, S. GEMAN, C. GRAFFIGNE, AND P. DONG, *Boundary detection by constrained optimization*, IEEE Trans. Pattern Anal. Mach. Intell., 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 Proc. of ISBI'08, 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 Anal. Mach. Intell., 28 (2006), pp. 657–662.

- [47] F. HEITZ AND P. BOUTHEMY, *Multimodal estimation of discontinuous optical flow using Markov random fields*, IEEE Trans. Pattern Anal. Mach. Intell., 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 Anal. Mach. Intell., 25 (2003), pp. 1051–1062.
- [49] S. INCE AND J. KONRAD, *Geometry-based estimation of occlusions from video frame pairs*, in Proc. of ICASSP'05 (2), Philadelphia, PA, 2005, pp. 933–936.
- [50] H. ISHIKAWA AND D. GEIGER, *Occlusions, discontinuities, and epipolar lines in stereo*, in Proc. of ECCV'98 (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*, Comp. Vis. Graph. 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. MICHAELIDIS, *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 Proc. of MLSP'09, Grenoble, France, September 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. on 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 Proc. of ECCV'06 (2), 2006, pp. 30–43.
- [60] V. KOLMOGOROV AND R. ZABIH, *Computing visual correspondence with occlusions via graph cuts*, in Proc. of ICCV'01, Vancouver, Canada., 2001, pp. 508–515.
- [61] ———, *Multi-camera scene reconstruction via graph cuts.*, in Proc. of ECCV'02 (3), Copenhagen, Denmark, 2002, pp. 82–96.
- [62] S. KUMAR AND M. HEBERT, *Discriminative random fields*, Int. J. Comput. Vision, 68 (2006), pp. 179–201.
- [63] T.K. LEUNG AND J. MALIK, *Detecting, localizing and grouping repeated scene elements from an image*, in Proc. of ECCV'96 (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. on Pattern Anal. Mach. Intell., 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 Proc. of CVPR'05, 2005, pp. 1071–1078.
- [66] J.L. LISANI AND J.-M. MOREL, *Detection of major changes in satellite images*, in Proc. of ICIP'03 (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 Proc. of NIPS'06, Vancouver, B.C. Canada, December 2006.
- [68] Y. MATSUSHITA AND S. LIN, *A probabilistic intensity similarity measure based on noise distribution*, in Proc. of CVPR'07, Minneapolis, Minnesota, 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 Anal. Mach. Intell., 26 (2004), pp. 1336–1347.
- [70] E. MÉMIN AND P. PÉREZ, *Hierarchical estimation and segmentation of dense motion fields*, Int. J. Comput. Vision, 46 (2002), pp. 129–155.
- [71] M. MIGNOTTE, *Nonparametric multiscale energy-based model and its application in some imagery problems*, IEEE Trans. Pattern Anal. Mach. Intell., 26 (2004), pp. 184–197.

- [72] A. MITTAL AND N. PARAGIOS, *Motion-based background subtraction using adaptive kernel density estimation*, in Proc. of CVPR'04, 2004, pp. 302–309.
- [73] P. MONASSE AND F. GUICHARD, *Fast computation of a contrast-invariant image representation*, IEEE Trans. Image Processing, 9 (200), pp. 860–872.
- [74] L.W. NEMSICK AND E. GERANIOTIS, *Adaptive multichannel detection of frequency-hopping signals*, IEEE Trans. Communications, 40 (1992), pp. 1502–1511.
- [75] R. NIU, P.K. VARSHNEY, M. MOORE, AND D. KLAMER, *Decision fusion in a wireless sensor network with a large number of sensors*, in Proc. Int. Conf. Information Fusion (FUSION'04), Stockholm, Sweden, 2004.
- [76] N. PARAGIOS AND R. DERICHE, *Geodesic active contours and level sets for the detection and tracking of moving objects*, IEEE Trans. Pattern Anal. Mach. Intell., 22 (2000), pp. 266–280.
- [77] N. PARAGIOS AND G. TZIRITAS, *Adaptive detection and localization of moving objects in image sequences*, Signal Processing: Image Communication, 14 (1999), pp. 277–296.
- [78] T. PÉCOT, C. KERVRANN, S. BARDIN, B. GOUD, AND J. SALAMERO, *Patch-based Markov models for event detection in fluorescence bioimaging*, in Proc. of MICCAI'08, vol. 2, New York City, USA, September 2008, pp. 95–103.
- [79] J. PILET, C. STRECHA, AND P. FUA, *Making background subtraction robust to sudden illumination changes*, in Proc. of ECCV'08 (4), 2008, pp. 567–580.
- [80] G. PIRIOU, P. BOUTHEMY, AND J-F. YAO, *Extraction of semantic dynamic content from videos with probabilistic motion models*, in Proc. of ECCV'04, Prague, Czech Republic, May 2004.
- [81] S. PRIMA, D.L. ARNOLD, AND D.L. COLLINS, *Multivariate statistics for detection of MS activity in serial multimodal MR images*, in Proc. of MICCAI'03, R.E. Ellis and T.M. Peters, eds., vol. 2878 of Lecture Notes in Computer Science, Montreal, Canada, November 2003, Springer, pp. 663–670.
- [82] 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.
- [83] R.J. RADKE, S. ANDRA, O. AL KOFABI, AND B. ROYSAM, *Image change detection algorithms: a systematic survey*, IEEE Trans. Image Processing, 14 (2005), pp. 294–307.
- [84] A. ROBIN, L. MOISAN, AND S. LE HGARAT-MASCLE, *An a-contrario approach for sub-pixel change detection in satellite imagery*, IEEE Trans. Pattern Anal. Mach. Intell., (2009).
- [85] 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.
- [86] P.L. ROSIN, *Thresholding for change detection*, Comp. Vis. Image Underst., 86 (2002), pp. 79–95.
- [87] S. ROTH AND M.J. BLACK, *Fields of experts: A framework for learning image priors*, in Proc. of CVPR'05, San Diego, CA, 2005, pp. 860–867.
- [88] 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 Proc. of ISBI'08, Paris, France, 2008, pp. 9–12.
- [89] N. SABATER, A. ALMANSA, AND J.-M. MOREL, *Rejecting wrong matches in stereovision*, Tech. Report 2008-28, CMLA, 2008.
- [90] H.J. SEO AND P. MILANFAR, *A non-parametric approach to automatic change detection in mri images of the brain*, in Proc. of ISBI'09, Boston, MA, 2009.
- [91] E. SHECHTMAN AND M. IRANI, *Matching local self-similarities across images and videos*, in Proc. of CVPR'07, Minneapolis, Minnesota, 2007, pp. 1–8.
- [92] Y. SHEIKH AND M. SHAH, *Bayesian object detection in dynamic scenes*, in Proc. of CVPR'05, San Diego, CA, 2005, pp. 74–79.
- [93] C. STAUFFER AND W.E.L. GRIMSON, *Adaptive background mixture models for real time tracking*, in Proc. of CVPR'99, Ft. Collins, CO, 1999, pp. 246–252.
- [94] C. STRECHA, R. FRANSENS, AND L. VAN GOOL, *A probabilistic approach to large displacement optical flow and occlusion detection*, in Proc. of ECCV Workshop SMVP, Prague, Czech Republic, 2004, pp. 71–82.
- [95] J. SUN, Y. LI, S.-B. KANG, AND H.-Y. SHUM, *Symmetric stereo matching for occlusion handling*, in Proc. of CVPR'05, San Diego, CA, 2005, pp. 299–406.
- [96] J. SUN, W. ZHANG, X. TANG, AND H.-Y. SHUM, *Background cut*, in Proc. of ECCV'06 (2), Graz, Austria, 2006, pp. 628–641.
- [97] M. SZUMMER, P. KOHLI, AND D. HOIEM, *Learning crfs using graph cuts*, in Proc. of ECCV'08 (2), 2008, pp. 582–595.
- [98] S. TILIE, L. LABORELLI, AND I. BLOCH, *A contrario false alarms removal for improving blotch*

- detection in digitized films restoration*, in Proc. of EC-SIPMCS and IWSSIP, Maribor, Slovenia, June 2007, pp. 410–413.
- [99] D. TOTH, T. AACH, AND V. METZLER, *Bayesian spatio-temporal motion detection under varying illumination*, in European Signal Processing Conference, M. Gabbouj and P. Kosmanen, eds., Tampere, Finland, September 3-8 2000, EURASIP, pp. 2081–2084.
 - [100] K. TOYAMA, J. KRUMM, B. BRUMITT, AND B. MEYERS, *Wallflower: principles and practice of background maintenance*, in Proc. of CVPR'99 (1), Ft. Collins, CO, 1999, pp. 255–261.
 - [101] T. VEIT, F. CAO, AND P. BOUTHEMY, *Probabilistic parameter-free motion detection*, in Proc. of CVPR'04 (1), Washington, DC, June 2004, pp. 715–721.
 - [102] Y. WEI AND L. QUAN, *Asymmetrical occlusion handling using graph cut for multi-view stereo*, in Proc. of CVPR'05, Washington, DC, USA, 2005, pp. 902–909.
 - [103] P. WEISS, A. FOURNIER, L. BLANC-FRAUD, AND G. AUBERT, *On the illumination invariance of the level lines under directed light. application to change detection*, Research Report 6612, INRIA, 2008.
 - [104] R. WILLETT AND R.D. NOWAK, *Minimax optimal level-set estimation*, IEEE Trans. Image Processing, 16 (2007), pp. 2965–2979.
 - [105] O. WILLIAMS, M. ISARD, AND J. MACCORMICK, *Estimating disparity and occlusions in stereo video sequences*, in Proc. of CVPR'05, San-Diego, CA, 2005, pp. 250–257.
 - [106] C.R. WREN, A. AZARBAYEJANI, T. DARRELL, AND A. PENTLAND, *Pfinder: real-time tracking of the human body*, IEEE Trans. Pattern Anal. Mach. Intell., 19 (1997), pp. 780–785.
 - [107] J. XIAO, H. CHENG, H.S. SAWHNEY, C. RAO, AND M. ISNARDI, *Bilateral filtering-based optical flow estimation with occlusion detection*, in Proc. of ECCV'06 (1), 2006, pp. 211–224.
 - [108] J. XIAO AND M. SHAH, *Motion layer extraction in the presence of occlusion using graph cuts*, IEEE Trans. Pattern Anal. Mach. Intell., 27 (2005), pp. 1644–1659.
 - [109] Z. ZIVKOVIC AND F. VAN DER HELDEN, *Efficient adaptive density estimation per image pixel for the task of background subtraction*, Pattern Recog. Letters, 27 (2006), pp. 773–780.
 - [110] M. ZONTAK AND I. COHEN, *Kernel-based detection of defects on semiconductor wafers*, in Proc. of MLSP'09, Grenoble, France, September 2009, pp. 1–6.