

DISCRIMINANT RANDOM FIELD AND PATCH-BASED REDUNDANCY ANALYSIS FOR IMAGE CHANGE DETECTION

Charles Kervrann^(1,2), Jérôme Boulanger⁽³⁾, Thierry Pécot^(1,2,4) and Patrick Pérez⁽¹⁾

(1) INRIA Centre Rennes - Bretagne Atlantique, F-35042 Rennes,

(2) INRA, UR341 Mathématiques et Informatique Appliquées, F-78352 Jouy-en-Josas,

(3) Johann Radon Institute for Computational (RICAM), Austrian Academy of Sciences, A-4040 Linz, Austria,

(4) “Cell and Tissue Imaging Facility” - IBISA, Institut Curie, F-75248 Paris

ABSTRACT

To develop better image change detection algorithms, new models able to capture all the spatio-temporal regularities and geometries seen in an image pair are needed. In contrast to the usual pixel-wise methods, we propose a patch-based formulation for modeling semi-local interactions and detecting occlusions and other local or regional changes in an image pair. To this end, the image redundancy property is exploited to detect unusual spatio-temporal patterns in the scene. We first define adaptive detectors of changes between two given image patches and combine locally in space and scale such detectors. The resulting score at a given location is exploited within a discriminant Markov random field (DRF) whose global optimization flags out changes with no optical flow computation. Experimental results on several applications demonstrate that the method performs well at detecting occlusions and meaningful regional changes and is especially robust in the case of low signal-to-noise ratios.

1. INTRODUCTION

Change detection is of significant interest in a number of applications, such as video-surveillance (e.g. airports, museums, shops, ...), medical diagnosis, cell biology imaging and remote sensing. There has been a substantial amount of work to handle changes in an image pair. For a recent survey of pixel-wise change detection methods, see [1]. 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, etc. This can be achieved by using adaptive thresholds applied to image differencing. Thresholding is then central in most change detection methods and the problem to be addressed further is how to integrate the spatial (contextual) information from the individual pixels to cope with camera jitter or animated texture in the background. In [2, 3, 4, 5], the authors considered Markov random fields to capture the spatial correlation among nearby pixels.

In the area of video analysis, a sequence with no moving object is traditionally used to learn the statistics of the static background [6, 7, 8, 9, 10]. In [8], the authors proposed a mixture of Gaussians to approximate the probability distribution function (pdf) of the background and to test each pixel of the current frame against the learned pdf.

Occlusions generated by object and/or camera movements in video sequences (with stereo pairs as a particular case) form a special type of image changes whose handling is critical for motion and disparity estimation. Therefore, several methods have attempted to simultaneously detect motion discontinuities and to compute optical flow (e.g., [11]), or to detect the violation of motion consistency assumption (e.g., [12, 13]). They remain however more complex because of their compound nature, and are plagued by a chicken-and-egg problem (motions and disparities are often inaccurately estimated around occlusions when these are not localized beforehand, and localization of them is based on some estimation of displacements).

Our idea for better handling both occlusions and temporal changes is to exploit the abundant redundancy of the two images of the scene of interest, expect at location of changes and occlusions, without establishing explicit correspondences (or, equivalently, with no explicit estimation of displacements). Since computing occlusion is ambiguous, we assume that, at each patch in the first image, corresponds a small set of similar patches in the other image but not necessary an unique one [14, 15]. To detect the occlusions or changes occurring in two images, we propose to collect a set of binary decisions made at individual pixels in a local neighborhood. This amounts to counting the number of neighboring patches in the second image that are similar to the current patch in the first image. This count at a given pixel is used as score to define the unary potential of a DRF [16, 17]. Completed with standard Ising regularization, the global optimization of this DRF via min-cut/max-flow provides a binary mask of image changes. The approach is fully unsupervised and needs no statistical assumption on image and noise. In addition, unlike many subtraction meth-

ods [7, 9] which need a series of training images, our algorithm is able to produce satisfying results using an image pair only.

2. LOCAL CHANGE DETECTION

Unlike previous methods (kernel-based [7, 9, 5] or mixture of Gaussians-based [8, 18, 10]) 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 [19]. Background subtraction based on temporal information cannot be performed as described in [7, 8] since we are using two images. Nevertheless, our method can be compared to background subtraction methods based on image pairs [20, 10].

In order to describe our estimation 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 non-local image features able to capture local geometries and spatial regularities. Our idea is to guess a n -dimensional patch $\underline{u}(\mathbf{x})$ in u from patches $\underline{v}(\mathbf{x}_i)$ taken in the (fixed size) semi-local neighborhood $B(\mathbf{x}) \subseteq \Omega$ with $N = |B(\mathbf{x})|$ elements observed at point \mathbf{x}_i in the second image v . For the sake of simplicity, a vectorized image patch $\underline{v}(\mathbf{x})$ is defined as the pixels ordered lexicographically in the patch. We assume that $\underline{u}(\mathbf{x}) \equiv \underline{v}(\mathbf{x}_i)$, for some $\mathbf{x}_i \in B(\mathbf{x})$ if no scene change occurs at pixel \mathbf{x} .

Given an appropriate patch-similarity measure $\Phi(\cdot, \cdot)$, we first propose to make individual decisions about the presence/absence of patches $\underline{v}(\mathbf{x}_i)$ in the second image v , similar to $\underline{u}(\mathbf{x})$ in the first image u . Individual decisions at neighboring points are then combined in a single score. Cooperation among neighboring points tends to enhance the ability to detect meaningful changes; namely, whether a change occurs or not within the search area $B(\mathbf{x})$.

More formally, consider a central patch $\underline{u}(\mathbf{x})$ in u at location \mathbf{x} to be compared to neighboring patches $(\underline{v}(\mathbf{x}_i))_{\mathbf{x}_i \in B(\mathbf{x})}$ in the second image v . Based on similarity

$$z(\mathbf{x}_i) = \Phi(\underline{u}(\mathbf{x}), \underline{v}(\mathbf{x}_i))$$

between patches, each pixel \mathbf{x}_i makes a decision $d_0(\mathbf{x}_i) \doteq \mathbf{1}(z(\mathbf{x}_i) \geq \tau(\mathbf{x})) \in \{0, 1\}$, where $\mathbf{1}(\cdot)$ is the indicator function, regarding the presence of a similar patch in the semi-local search area. The score is non-negative and the smaller the value of score is, the more similar are the patches. The decision depends on whether the score exceeds a spatially-varying threshold $\tau(\mathbf{x})$.

The collaborative neighborhood-wise decision for change detection is obtained through a fusion rule. After collecting

the set of individual decisions $\{d_0(\mathbf{x}_i) : \mathbf{x}_i \in B(\mathbf{x})\}$, the total number of positive decisions $S_N(\mathbf{x}) = \sum_{\mathbf{x}_i \in B(\mathbf{x})} d_0(\mathbf{x}_i)$ made by individual pixels is used as the main information at location \mathbf{x} : the final system should favor a change flagging at location with large score, and no change detection otherwise. This is mitigated with classic spatial regularization within a compound energy function associated to a conditional (or discriminant) random field.

To compute 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 one single image*, by examining the highest scores in very small neighborhoods $b(\mathbf{x})$ (e.g., 3×3 square window). More precisely, we postulate that all positive decisions $d_0(\mathbf{x}_i)$ correspond to scores $z(\mathbf{x}_i)$ higher than the highest score $\tau(\mathbf{x}) \doteq \tau_u(\mathbf{x})$ at pixel \mathbf{x} computed from a single image u and defined as

$$\tau_u(\mathbf{x}) \doteq \max \left(\sup_{\mathbf{y} \in b(\mathbf{x})} \Phi(\underline{u}(\mathbf{x}), \underline{u}(\mathbf{y})), \tau_0 \right). \quad (1)$$

The idea is to artificially generate perturbations by examining a local neighborhood and to estimate the minimal thresholds for robust detection. The nearby patches are assumed to be perturbed configurations of the central patch.

Considering only one training image for change detection has been already suggested in [10] but a mixture of Gaussians was necessary to derive a unique decision threshold for the whole image. By introducing a minimal value τ_0 in (1) defined as the average of the lowest scores computed over the image domain Ω , we increase the robustness to low signal-to-noise ratios.

It is also worth noting that the method 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 scores from u and v as follows:

$$\begin{aligned} \tau(\mathbf{x}) &\doteq \min(\tau_u(\mathbf{x}), \tau_v(\mathbf{x})), \\ z(\mathbf{x}_i) &\doteq \min(\Phi(\underline{u}(\mathbf{x}), \underline{v}(\mathbf{x}_i)), \Phi(\underline{v}(\mathbf{x}), \underline{u}(\mathbf{x}_i))). \end{aligned} \quad (2)$$

As this stage of system design, several size parameters require our attention: the size N of search window B , the size of window b involved in the computation of location-dependent threshold τ and the size n of the patches.

The size N of the search window depends on the amplitude of the motions (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) in the applications with still cameras. This means we are testing similarity in very local

neighborhoods. Accordingly, we choose $|b(\mathbf{x})| \equiv |B(\mathbf{x})|$, which limits the number of the 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 the background of second image is cluttered.

As for the size of patches, we chose not to commit to a single one, which would be difficult to select, and consider instead a range of different sizes $n_\ell = (2\ell + 1)^2$, $\ell = 1 \dots L$. This is key to the robustness of the approach. We will denote

$$S_{N,\ell}(\mathbf{x}) = \sum_{\mathbf{x}_i \in B(\mathbf{x})} \mathbf{1}(\Phi(\underline{u}(\mathbf{x}), \underline{v}(\mathbf{x}_i)) \geq \tau(\mathbf{x})) \quad (3)$$

where \underline{u} and \underline{v} image patches of size n_ℓ . Intuitively, the number of scales is related to the area of change regions in the image pairs.

3. GLOBAL COST FUNCTION WITH RESULTS

The sequence of scores $S_{N,\ell}(\mathbf{x})$, $\ell = 1 \dots L$ at pixel \mathbf{x} are the key information to derive the unary potential of the global cost function. A change should be more likely to be detected at pixel \mathbf{x} when scores at all scales are large (ideally all equal to N), and more likely to be flagged with no change when these scores are low (ideally all equal to 0).

We define the global cost function of the binary labeling $d \in \{0, 1\}^\Omega$, with $d(\mathbf{x}) = 1$ indicating a change at pixel \mathbf{x} , as follows:

$$\mathcal{J}(d) = \sum_{\mathbf{x} \in \Omega} \sum_{\ell=1}^L \left| \frac{1}{N} S_{N,\ell}(\mathbf{x}) - d(\mathbf{x}) \right| + \beta \sum_{\langle \mathbf{x}, \mathbf{y} \rangle} [1 - \mathbf{1}(d(\mathbf{x}) = d(\mathbf{y}))]. \quad (4)$$

The first terms tends to label a pixel as an occluded pixel if the number of positive decisions is high for a large number of scales. The second term enforces piecewise spatial smoothness through Ising prior. Balance parameter $\beta > 0$ is adjusted by the user by trials, while experimenting the optimization algorithm. This parameter has significant effects on the qualitative properties of the minimizer.

The global minimizer of this cost function, amounting to the MAP estimate of the conditional random field with \mathcal{J} as its energy function, is obtained using efficient min-cut/max-flow algorithm [21].

To evaluate our patch-based method, we conducted experiments on a variety of image pairs (including some with illumination changes and motion variations) and applications including video surveillance and blotch detection in old digitized movies. We report a few of them, using the Euclidean distance to define the similarity measure Φ except in the case of illumination changes where patches are first normalized to have zero-mean. We also set $B(\mathbf{x}) \equiv b(\mathbf{x})$ as 3×3 pixels windows.

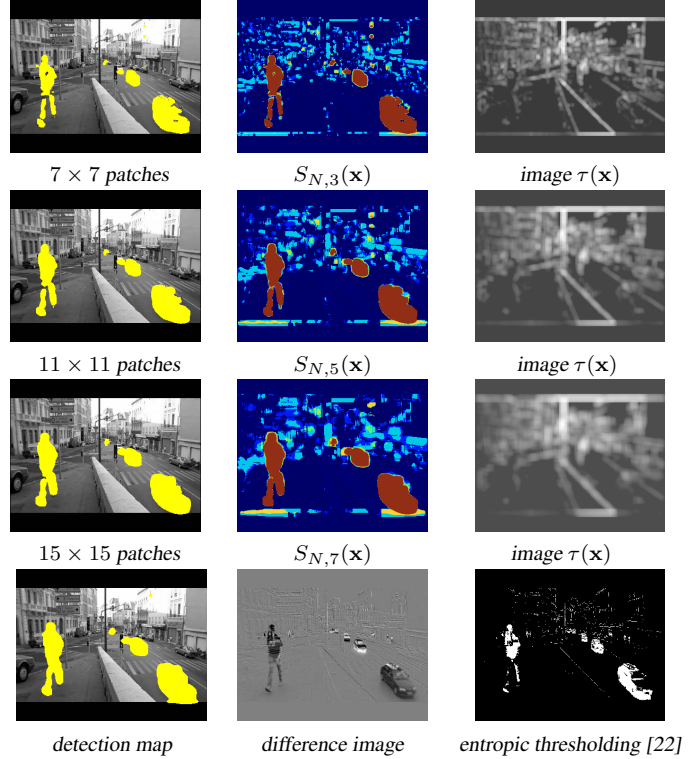


Fig. 1. The images $S_{N,\ell}(\mathbf{x})$ correspond to the number of positive decisions in a local neighborhood $B(\mathbf{x})$ for 7×7 , 11×11 and 15×15 patches. The level sets corresponding to $T = N$ are superimposed on the original image for each patch size (left column). The last row shows the detection results obtained by the complete approach with patch size ranging from 3×3 to 23×23 pixels ($L = 11$) and $\beta = 1$. For comparison, the entropic thresholding method of [22] is applied to obtained the last result.

In the first example in Fig. 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 $L = 11$. We also examined the detection maps and the counting values $S_{N,\ell}(\mathbf{x})$ 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 Figure 2, other examples of change detection are shown for video-surveillance scenarios where illumination conditions are unchanged.

In Fig. 3, the image pair is composed of two consecutive frames of an old movie. We used small patches ($L = 3$ pixels) to detect all the blotches known to suddenly appear in the image at random locations [23, 24] and considered small neighborhoods (3×3 square windows). Blotches are bright or dark small regions caused by dirt or damage processes due to ageing effects and bad film quality. Generally, they have different intensity values from the original im-

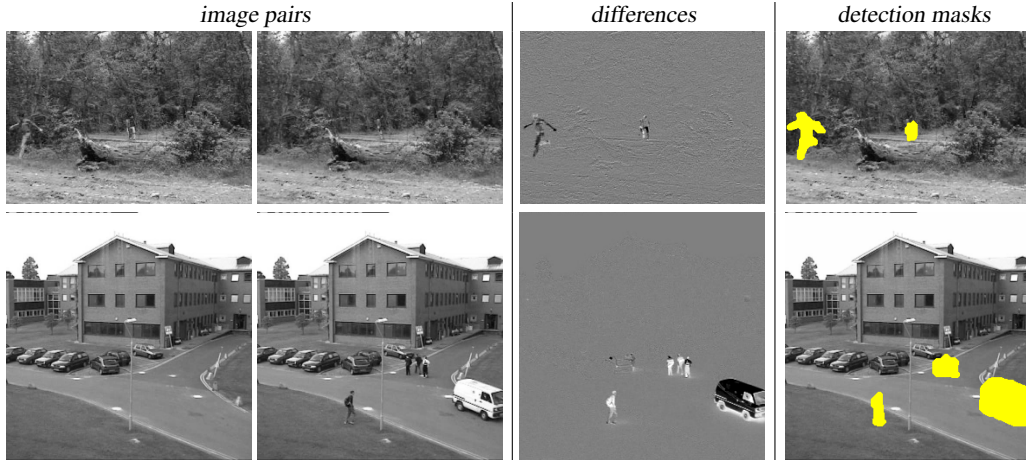


Fig. 2. Change detection results of our method on two outdoor image pairs (top: $L = 5$; bottom: $L = 3$).

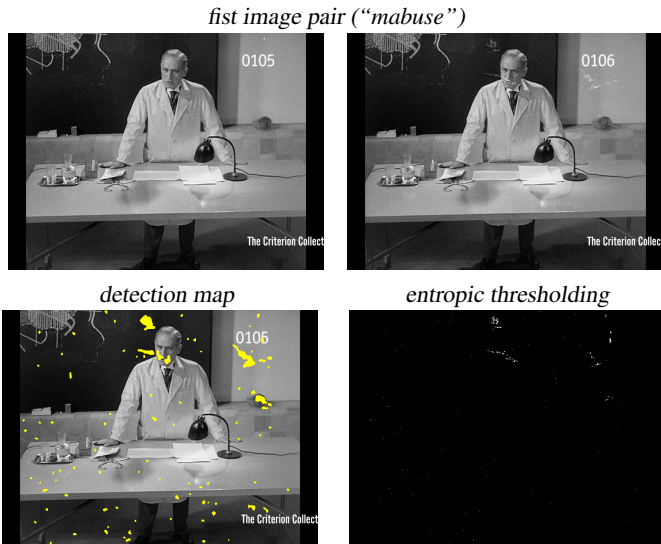


Fig. 3. Blotch detection in an image pair taken from an old movie ($L = 3$). The histogram of the absolute value of the difference image was analyzed for automatically thresholding [22].

age contents. In that case, we analyzed the images in both directions. These detected regions can be repaired further by inpainting methods (e.g. [25]) if successfully detected. Too many alarms is not desirable since repairing and inpainting methods may be fallible. The experiments we presented demonstrate that the patch-based method works well for tested image pairs.

We now show a few results on the WallFlower background subtraction benchmark [26], demonstrating the robustness of the approach in presence of sudden illumination changes, shadows and specularities. On the tested im-

ages (with ground truths), the performance of baseline techniques are limited as reported in [26, 27]. The performance of our method shown in Fig. 4 is satisfactory when compared to the state-of-the-art background subtraction methods which can handle illumination changes and moving objects [26, 7, 8, 9, 18, 28, 27, 20]. Our method did the same job using two input images and sometimes outperformed methods [8, 7, 29, 27, 20] which require a long image sequence as input. Nevertheless, background subtraction cannot be used to detect changes between two images only (background model must be learned beforehand). Note that motion in the background makes detecting changes by baseline methods very challenging (see [26]).

Results in Fig.5 focus on the problem of detecting occlusions as in [30, 31, 32]. 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 $B(\mathbf{x})$), no change is detected. As shown in examples in Fig. 5, 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 $B(\mathbf{x})/2$ pixels. If $B(\mathbf{x})$ is large enough, the detector is potentially invariant to a wide range of movements, including those caused by camera displacement, to the extent of only detecting the two first types of events. In this example, the input images are highly redundant. Then, it makes sense to examine the situation with 3×3 search windows $B(\mathbf{x})$. A limited number of patch sizes ($L = 3$) is more suitable since occluded areas are small. Similar textured patches are actually found in the second image because of texture redundancy. If $B(\mathbf{x})$ is

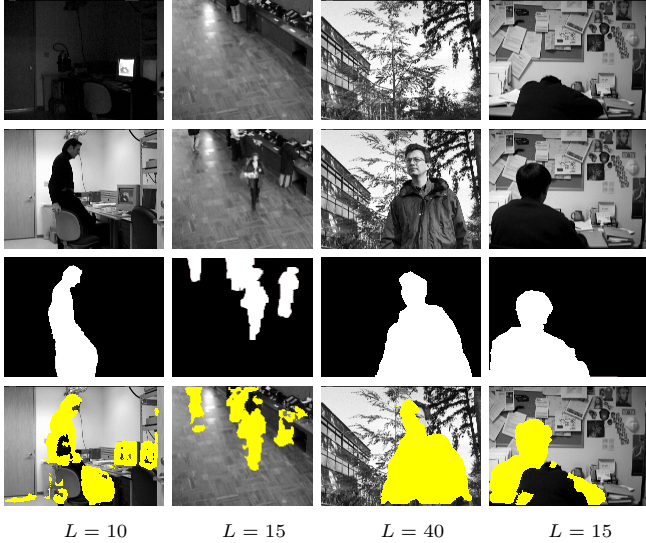


Fig. 4. Change detection results of our method for the test image pairs presented in [26, 27]. First rows: image pairs; Third row: ground truths; Fourth row: our detection results.

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 will correspond to occluded regions.

4. CONCLUSION

We have described a non-parametric patch-wise change detector. The method robustly detects areas in images where the redundancy property captured by image patches does not hold. Our patch-based approach is robust to many types of variations, such as local appearance change, residual motion and scale variation. In this approach, local and independent decisions “0” or “1” for nearby patches are first made for different patch sizes, based on an adaptive location-dependent threshold. Final change detection map is obtained by minimizing a DRF energy which combines these local detection informations with spatial regularization.

This approach is capable of extracting clean occlusion and change masks. An important feature of the approach is that image motion does not have to be computed explicitly. 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.

In all experiments, gray level intensity values are used for matching although color images could be considered in future work. Moreover, we did not address the stereo problem since the displacements are traditionally large. This could be investigated in future work.

5. REFERENCES

- [1] R.J. Radke, S. Andra, O. Al Kofahi, and B. Roysam, “Image change detection algorithms: a systematic survey,” *IEEE Trans. Image Processing*, vol. 14, no. 3, pp. 294–307, March 2005.
- [2] D.F. Prieto L. Bruzzone, “Automatic analysis of the difference image for unsupervised change detection,” *IEEE T. Geoscience and Remote Sensing*, vol. 38, no. 3, pp. 1171–1182, 2000.
- [3] P.K. Varshney T. Kasetkasem, “An image change detection algorithm based on markov random field models,” *IEEE T. Geoscience and Remote Sensing*, vol. 40, no. 8, pp. 1815–1823, 2002.
- [4] 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.*, vol. 22, no. 3, pp. 266–280, 2000.
- [5] Y. Sheikh and M. Shah, “Bayesian object detection in dynamic scenes,” in *CVPR*, San Diego, CA, 2005, pp. 74–79.
- [6] C.R. Wren, A. Azarbayejani, T. Darrell, and A. Pentland, “Pfinder: real-time tracking of the human body,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 780–785, 1997.
- [7] A. Elgammal, D. Harwood, and L. Davis, “Non-parametric model for background subtraction,” in *ECCV*, Dublin, Ireland, 2000, pp. 751–767.
- [8] C. Stauffer and W.E.L. Grimson, “Adaptive background mixture models for real time tracking,” in *CVPR*, Ft. Collins, CO, 1999, pp. 246–252.
- [9] Anurag Mittal and Nikos Paragios, “Motion-based background subtraction using adaptive kernel density estimation,” in *CVPR*, 2004, pp. 302–309.
- [10] P.-M. Jodoin, M. Mignotte, and J. Konrad, “Statistical background subtraction using spatial cues,” *IEEE Trans. Circuits and Systems for Video Technology*, vol. 17, no. 12, pp. 1758–1763, October 2007.
- [11] Etienne Mémin and Patrick Pérez, “Hierarchical estimation and segmentation of dense motion fields,” *Int. J. Comput. Vision*, vol. 46, no. 2, pp. 129–155, 2002.
- [12] F. Heitz and P. Bouthemy, “Multimodal estimation of discontinuous optical flow using markov random fields,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 15, no. 12, pp. 1217–1232, 1993.
- [13] Jiangjian Xiao, Hui Cheng, Harpreet S. Sawhney, Cen Rao, and Michael Isardi, “Bilateral filtering-based optical flow estimation with occlusion detection,” in *ECCV (I)*, 2006, pp. 211–224.

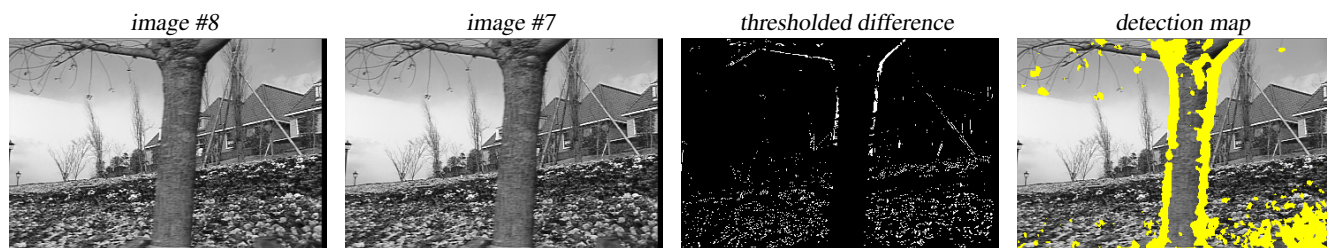


Fig. 5. Detection of spatio-temporal discontinuities in ‘flower garden’ (3×3 neighborhoods $\mathcal{N}(\mathbf{x})$, 3×3 search windows $B(\mathbf{x})$, $L = 3$, $\beta = 0.1$).

- [14] J. Boulanger, Ch. Kervrann, and P. Bouthemy, “Space-time adaptation for patch based image sequence restoration,” *IEEE Trans. on Pattern Anal. Mach. Intell.*, vol. 29, no. 6, pp. 1096–1102, 2007.
- [15] A. Buades, B. Coll, and J.M. Morel, “Nonlocal image and movie denoising,” *Int. J. Comput. Vision*, vol. 76, no. 2, pp. 123–139, 2008.
- [16] S. Kumar and M. Hebert, “Discriminative random fields,” *Int. J. Comput. Vision*, vol. 68, no. ???, pp. 179–201, 2006.
- [17] L. Lu and G.D. Hager, “Dynamic background/foreground segmentation from images and videos using random patches,” in *Neural Information Processing and System (NIPS’06)*, Vancouver, B.C. Canada, December 2006.
- [18] Z. Zivkovic and F. van der Heijden, “Efficient adaptive density estimation per image pixel for the task of background subtraction,” *Pattern Recog. Letters*, vol. 27, no. 7, pp. 773–780, 2006.
- [19] S. Lim, A. Mittal, L.S. Davis, and N. Paragios, “Fast illuminationinvariant background subtraction using two views: error analysis, sensor placement and applications,” in *CVPR*, 2005, pp. 1071–1078.
- [20] Julien Pilet, Christoph Strecha, and Pascal Fua, “Making background subtraction robust to sudden illumination changes,” in *ECCV (4)*, 2008, pp. 567–580.
- [21] Vladimir Kolmogorov and Ramin Zabih, “Computing visual correspondence with occlusions via graph cuts,” in *ICCV*, Vancouver, Canada., 2001, pp. 508–515.
- [22] 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 Process.*, vol. 29, no. 3, pp. 273–285, 1985.
- [23] J. Biemond, P.M.B. van Roosmalen, and R.L. Lagendijk, “Improved blotch detection by postprocessing,” in *Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP’99)(6)*, Washington, DC, 1999, pp. 3101–3104.
- [24] S. Tilie, L. Laborelli, and I. Bloch, “A contrario false alarms removal for improving blotch detection in digitized films restoration,” in *EC-SIPMCS and IWSSIP*, Maribor, Slovenia, June 2007, pp. 410–413.
- [25] Alexei A. Efros and Thomas K. Leung, “Texture synthesis by non-parametric sampling,” in *ICCV (2)*, Kerkyra, Greece, 1999, pp. 1033–1038.
- [26] K. Toyama, J. Krumm, B. Brumitt, and B. Meyers, “Wallflower: principles and practice of background maintenance,” in *CVPR (1)*, Ft. Collins, CO, 1999, pp. 255–261.
- [27] M. Pietikainen M. Heikkila, “A textured-based method for modeling the background and detecting moving objects,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 4, pp. 657–662, 2006.
- [28] 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 *ECCV (1)*, Marseille, France, 2008, pp. 113–126.
- [29] Til Aach, Lutz Dumbgen, Rudlof Mester, and Daniel Toth, “Bayesian illumination-invariant motion detection,” in *ICIP (3)*, Thessaloniki, Greece, 2001, pp. 640–643.
- [30] 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.*, vol. 24, no. 5, pp. 712–718, 2002.
- [31] Jiangjian Xiao and Mubarak Shah, “Motion layer extraction in the presence of occlusion using graph cuts,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 10, pp. 1644–1659, 2005.
- [32] P.-M. Jodoin, M. Mignotte, and C. Rosenberger, “Segmentation framework based on label field fusion,” *IEEE Trans. Image Processing*, vol. 16, no. 10, pp. 2535–2550, October 2007.