Background subtraction adapted to PTZ cameras by keypoint density estimation

Constant Guillot[1]
constant.guillot@cea.fr
Maxime Taron[1]
maxime.taron@cea.fr
Patrick Sayd[1]
patrick.sayd@cea.fr
Quoc-Cuong Pham[1]
quoc-cuong.pham@cea.fr
Christophe Tilmant[2]
christophe.tilmant@lasmea.univ-bpclermont.fr
Jean-Marc Lavest[2]
J-Marc.Lavest@u-clermont1.fr

[1] CEA LIST
Laboratoire Vision et Ingénierie des Contenus,
BP 94, Gif-sur-Yvette,
F-91191 France
[2] LASMEA UMR 6602,
PRES Clermont Université/CNRS,
63177 Aubière cedex, France

Abstract

PTZ cameras have the ability to cover wide areas with an adapted resolution. In this article we propose a background subtraction algorithm suitable for a PTZ camera performing a guard tour. It relies on the estimation of a probability density function based on the matching of keypoints between the background and the current image. Its main interest consists in the resulting robustness to sudden illumination changes. Experiments show that our algorithm is more robust to sudden changes in illumination than one of the state of the art approaches in situations of low camera frame-rate. These properties allow us to successfully maintain up to date the background model of a wide scene made up of a collection of images.

1 Introduction

With the increasing use of CCTV cameras many efforts have been made to automate the analysis of video streams in order to improve their efficiency. Wide angle cameras can be used to monitor a wide scene, their interest is however limited by their low resolution when it comes to analysing the scene. Pan Tilt Zoom (PTZ) cameras have two rotation axis and a zoom function which enable focusing on a part of the scene at any suitable resolution. The obvious drawback of the PTZ sensor is its limited field of view.

In the case of a static camera, one of the usual approaches to issues such as tracking or object recognition is to build a background model. This model, which will have to be initialised and updated continuously, will allow the detection of objects of interest by estimating a distance to the current image. In the case of a PTZ camera, maintaining a whole

BMVC 2010 doi:10.5244/C.24.34

background model is a challenging problem since the necessary information is only rarely available.

In this article, we consider the case of a PTZ camera doing a guard tour over a wide area with the aim to detect objects of interest. The camera follows a predefined set of positions (pan, tilt, zoom) covering the area at an adapted resolution. For each of these positions we can consider that we are in the case of a static camera. The duration of a single tour can last up to tens of seconds. Such a duration constitutes a major difficulty since the background model will not be continuously updated and thus one can expect important disparities, especially in terms of illumination, between the model and the current image. Indeed, these difficulties are already an important issue in the common use of static cameras and can be expected to be magnified by the fact that we only have temporarily sparse information. We have developed a simple yet effective background subtraction algorithm based on the estimation of a density of non matching keypoints in the image. Although keypoints have been widely used in a number of computer vision problems, to our knowledge no background subtraction method based on keypoints has been published. Experiments show convincing results, especially in terms of robustness to sudden illumination variations.

This paper is composed as follows. Section 2 gives an overview of the existing previous work. Our algorithm is detailed in section 3, and experimental results are given in section 4. Conclusions and further work is section 5.

2 Previous work

There exist many background subtraction techniques in the literature, most of which are designed for static cameras. Stauffer and Grimson [13] first introduced a very popular statistical approach based on a mixture of Gaussian distributions as a model of the luminance of each pixel. This model is updated at each frame to take into account the variations of the background. This model inspired many methods such as Chen *et al.*'s [6]. Instead of having a pixel based approach they consider 8×8 blocks on top of which a local texture descriptor is computed and modelled by a mixture of Gaussian distributions. The descriptor encodes a contrast histogram and therefore significantly increases robustness to illumination variations. An overview of background subtraction methods based on mixtures of Gaussians is given by Bouwmans *et al.* [5].

These methods rely on colour information and are therefore sensitive to illumination changes. As a consequence a frequent update of the background model is required and these methods are not adapted to our case.

Solutions have also been proposed for the specific case of PTZ cameras. Most approaches are based on the creation of a mosaic of the scene background. The images from the camera are then registered on the mosaic, background subtraction is performed and then the background model is updated. The main difficulty stands in doing real time quality image registration. Bhat *et al.* [4] and Cucchiara *et al.* [7] register images on the background mosaic by estimating a translation. This registration model is however not adapted to wide scenes. Azzari *et al.* [1] and Robinault *et al.* [12] both estimate a homography. While Robinault detects moving objects using a mixture of gaussians as a background model, Azzari *et al.* use a model of the camera noise and only consider a difference between the background image and the current image.

The drawback of these methods is that there is no global update of the background model, thus their is no warranty that the model of an area which has not been visited for a while

will be usable. As a consequence these methods are adapted for tracking or moving object detection where an updated background is needed only in the neighbourhood of objects of interest.

The algorithm which is the most similar to ours is Trichet *et al.* [14]. Keypoints are labelled as foreground or background to improve the accuracy of a tracking application. A tracked object is defined by its bounding box and the label is given according to four features: the label of the matched keypoint, colour, motion and position in the bounding box. Thus only inlier keypoints will be used to estimate the position of the target. However even if some keypoints are labelled as foreground or background, they do not perform background subtraction properly speaking.

3 Our method

We are now going to introduce our background subtraction algorithm (see fig 1). The main idea consists in estimating a density of non matching keypoints between the background image and the current image. We assume that keypoints which cannot be matched from one image to the other belong to objects of interest.

As explained above, using a mosaic to build a background model presents several drawbacks. Since our application (object detection) does not explicitly require the creation of a mosaic, we choose to use a collection of registered images as a background model. Each image corresponding to a predefined position of the PTZ camera in the guard tour.

3.1 Registering images

The first stage of our method is the extraction of keypoints in both the background and current images. We use a Harris keypoint detector with a low threshold to have points spread everywhere in the images. We use SURF [2] as a local descriptor neighbourhood of the keypoint. It is based on a histogram of oriented gradients and known to be robust to illumination variations. Due to mechanical imprecision on the pan tilt and zoom parameters, the most stable keypoints, the one with the highest Harris scores, are used to register the current image on the background image by robustly estimating a homography. This preliminary registration does not require to compute extra keypoints since we use the same keypoints as those used in the following background subtraction phase.

3.2 Matching keypoints

Once image registration is done, keypoints are matched in order to process background subtraction. During this phase we do not simply match the keypoints of the background image with keypoints of the current images, but for each image we consider the union of background and current image keypoints.

If we only used keypoints from the background image there could be areas where there is no keypoint in the model, in saturated zones for instance. Matching keypoints from this model to the same positions of the current image, an object of interest appearing in such area would not be detected. For a similar reason using keypoints of the current image to match them on the background image would not be sufficient. In addition to avoiding false negatives considering the union of the keypoints also helps to avoid many false positives. Indeed Harris keypoints are not very stable and it often happens that a keypoint is detected in

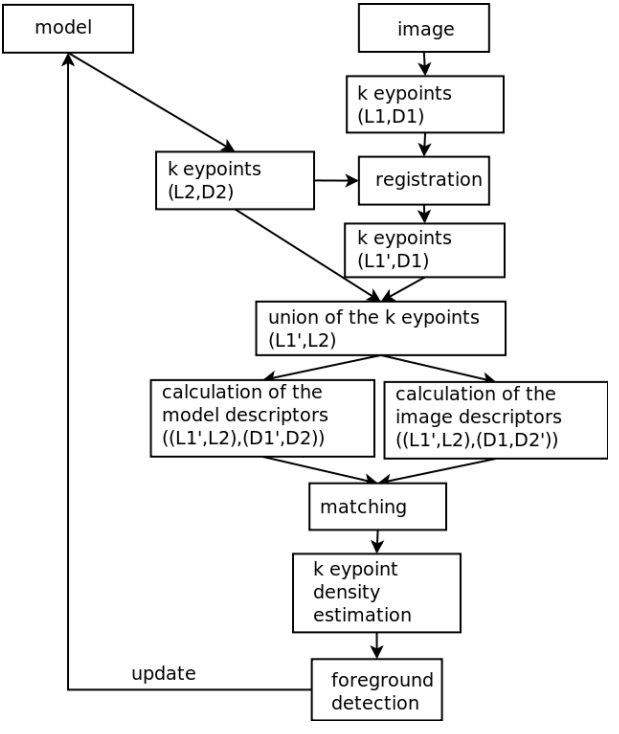

Figure 1: Background subtraction algorithm used at each position of the PTZ camera during the guard tour. $L1$ and $L2$ represent lists of keypoints' coordinate. $L1'$ is the list of registered keypoints computed from $L1$. $D1$ and $D2$ represent two lists of keypoint descriptors.

the background image but not in the current image. This is shown on the left image of figure 2 where only the blue keypoints are detected on both background and current image. With no matching candidate these particular keypoints would not be matched and then would be considered as belonging to an object of interest.Thanks to the preliminary registration and the use of the union of both keypoint lists, it is possible to limit the number of matching candidates by choosing a small search window. In the case of a static background we can choose to have only one matching candidate per keypoint, the one at the same location. If the background is dynamic, typical case of waving trees, we can use a wider search window which will allow to integrate part of the background movements and prevent false positives.

3.3 Keypoint density estimation

Once the matching step is done, non matching keypoints are assumed to be part of the foreground. We use a non-parametric estimation method to estimate a density of non-matching keypoints to obtain a dense information from this very sparse one. We consider that the non matching keypoints are observations of a random variable following an unknown probability density function d. This pdf d is then estimated with a Kernel Density Estimation method [10].

Let $(p_1, \ldots, p_N)$ be the N non matching keypoints, then

$$\hat{d}_h(x) = \frac{1}{Nh} \sum_{i=1}^{N} K\left(\frac{\|x - p_i\|_{img}}{h}\right), \tag{1}$$

with K being a kernel function, is an estimation of the value of d to the pixel x. The parameters h is a smoothing parameter which specifies the influence of each observation on its neighbourhood. We have chosen to use a Gaussian kernel

$$K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}. \tag{2}$$

Rather than using $\hat{d}_h$ we consider $N\hat{d}_h$ which is invariant to the number of keypoints detected in the image. Thus the same threshold is used on sequences with distinct resolutions. Pixels where $N\hat{d}_h > s$ are classified as foreground, the others are considered as background, with s being the detection threshold. In practise we have chosen to take $s = 0.08$ and $h = 11$.

Figure 2: Left: Union of keypoints detected on background and current image. Blue keypoints are detected on both images while the red ones are detected on only one image. Center: Non matching keypoints between the background image and the current view. Right: The associated density probability function. Blue is for low, green for medium and red for high density values.

3.4 Background update

In order to take into account the variations of the background an updating stage is necessary. The idea is here to replace the pixels of the background image by pixels from the current image, at any location that is not labelled as foreground. Indeed since we have a temporarily spare information, information at time $n-1$ becomes obsolete at time n. Let bg_n and img_n be the background image and the current image at step n, and x be a pixel. We apply the following updating rule:

If $\hat{d}_h(x) > s$,

$$bg_n(x) = bg_{n-1}(x). \tag{3}$$

Else,

$$bg_n(x) = bg_{n-1}(x)\frac{N\hat{d}_h(x)}{s} + img_n(x)\left(1 - \frac{N\hat{d}_h(x)}{s}\right). \tag{4}$$

The aim of equation 4 is to smooth the background image at the frontier of the objects of interest, where $N\hat{d}_h$ is still close to s. This prevents from introducing salient artificial edges in the background model in case of important changes in the illumination. This could lead to the creation of artificial keypoints and therefore to false positives in subsequent frames. Pixels from the background corresponding to pixels where a foreground object is detected are left unchanged (equation 3).

4 Experimental results

We compared our algorithm to Chen *et al.* [6] which has been proved to be a very competitive state of the art method [9], and to Stauffer and Grimson [13]. In a first part evaluation is done on several sequences acquired during a guard tour over an area of 40×40 meters square. The tour is composed of 25 positions and lasts for 30 seconds, therefore we can consider that each view is equivalent to a static camera with a frame rate of one image per thirty seconds. We use only 25 images to learn the background model. If we had used more images, the first tour used for background model learning would be so long that the model would be totally outdated for the detection tours. Since the shift between images may be important because of both wind and mechanical imprecision of the PTZ, all images are registered before being treated by Chen *et al.*'s and Stauffer's algorithm in order to have a fair comparison.

Figure 3 shows the detection results from three different views of the guard tour. Images are 768×576 and an average of 3500 keypoints are detected in them. The main difficulties on these sequences consist in the moving shadows and trees. Objects of interest are well detected even if one can notice the existence of false negatives in some homogeneous areas. Detected blobs tend to be wider than the objects of interest. This phenomenon is due to the smoothing parameter h and is hard to avoid. Indeed a lower h would lead to more false negatives at the centre of homogeneous areas. Even if there is no sudden illumination variation in this sequence, our algorithm outperforms Chen's in low contrast areas. This shows the importance of using a discriminative descriptor and confirms that SURF is an adapted choice.

We also tested our algorithm on a sequence showing sudden illumination changes and which we will later refer to as the *Light change* sequence. Figure 4 shows two sets of three consecutive frames from an outdoor static camera with a low frame-rate (one image every 20 seconds). The resolution is 640×480 and an average of 1500 keypoints are detected. As expected Chen's background subtraction algorithm cannot deal properly with important discontinuities in illumination. This algorithm even fails when considering all images of the sequence (25 images per second) as the change in illumination cannot be integrated to the model and generates a similar quantity of false positives. Our method appears to be more robust since the only false positives are situated on strong shadows or saturated areas. This can be explained by the fact that our method is based on keypoints which are linked to the image geometry and that the SURF descriptor is robust to illumination variations.

Figure 5 shows more qualitative results. Whereas the PETS and EPFL sequences are not challenging in terms of illumination variation they show that our algorithm behaves well on weakly textured scenes. The *train* sequence is another example of challenging sequence for which our algorithm is much more stable when sudden changes in illumination occur.

Quantitative results are given in table 1. It appears that our algorithm is clearly more stable than others. These statistics may seem to be low but sequences on which they were computed are very challenging (important changes in illumination) as well as the experi-

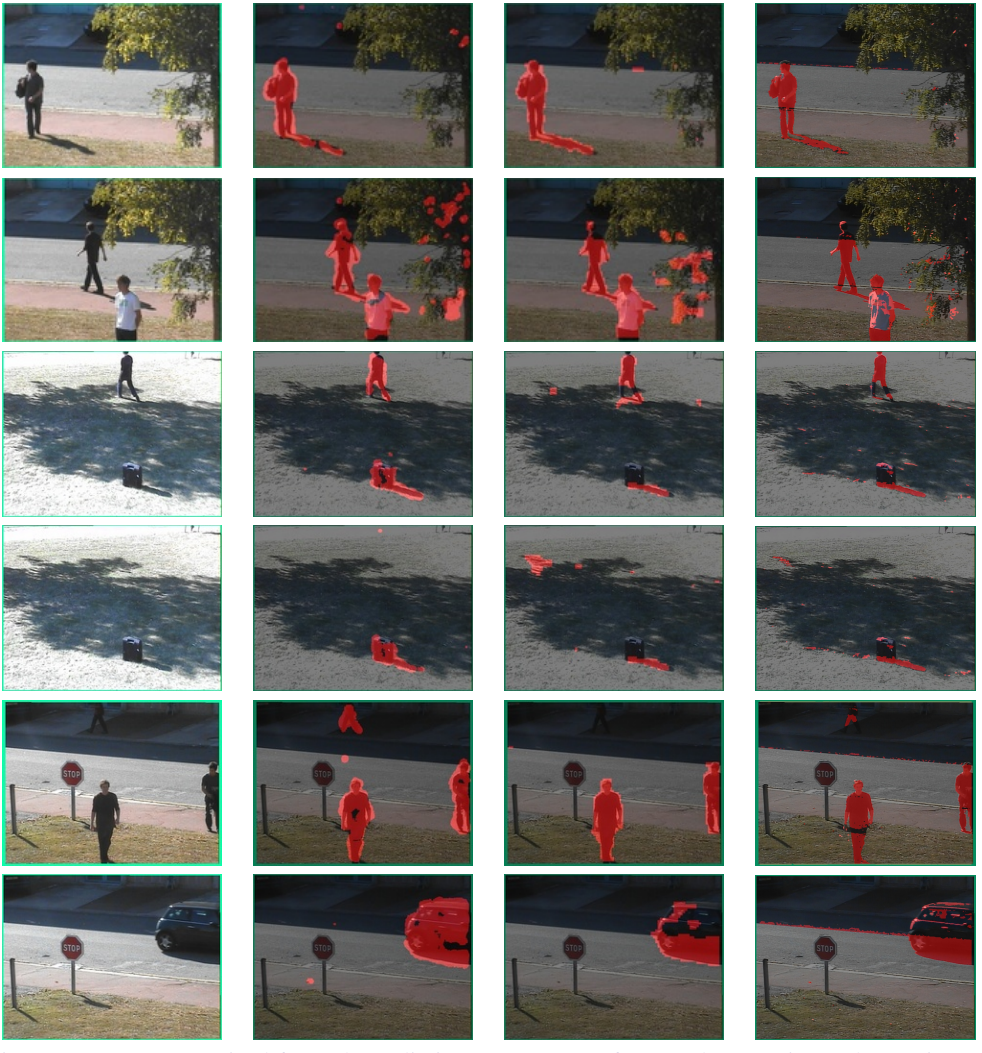

Figure 3: Images acquired from three distinct sequences of a guard tour. First column: images acquired from the PTZ camera. Second column is the background subtraction results with out algorithm. Third column with Chen *et al.*'s [6]. Fourth column Stauffer and Grimson's [13].

mental conditions (very low frame rate). Moreover, as it can be seen on figures 3, 4 and 5, the loss of precision of our algorithm is mainly due to the fact that it tends to over segment foreground blobs and is not due actual false alarms in terms of keypoint matching.

Figure 6 shows the precision and recall curves computed on two subsequences, the first one is during an important change in illumination while the second is during a phase when the global illumination is stable. It clearly points out that our algorithm is more robust than Chen *et al*. and confirms the qualitative results of figure 4.

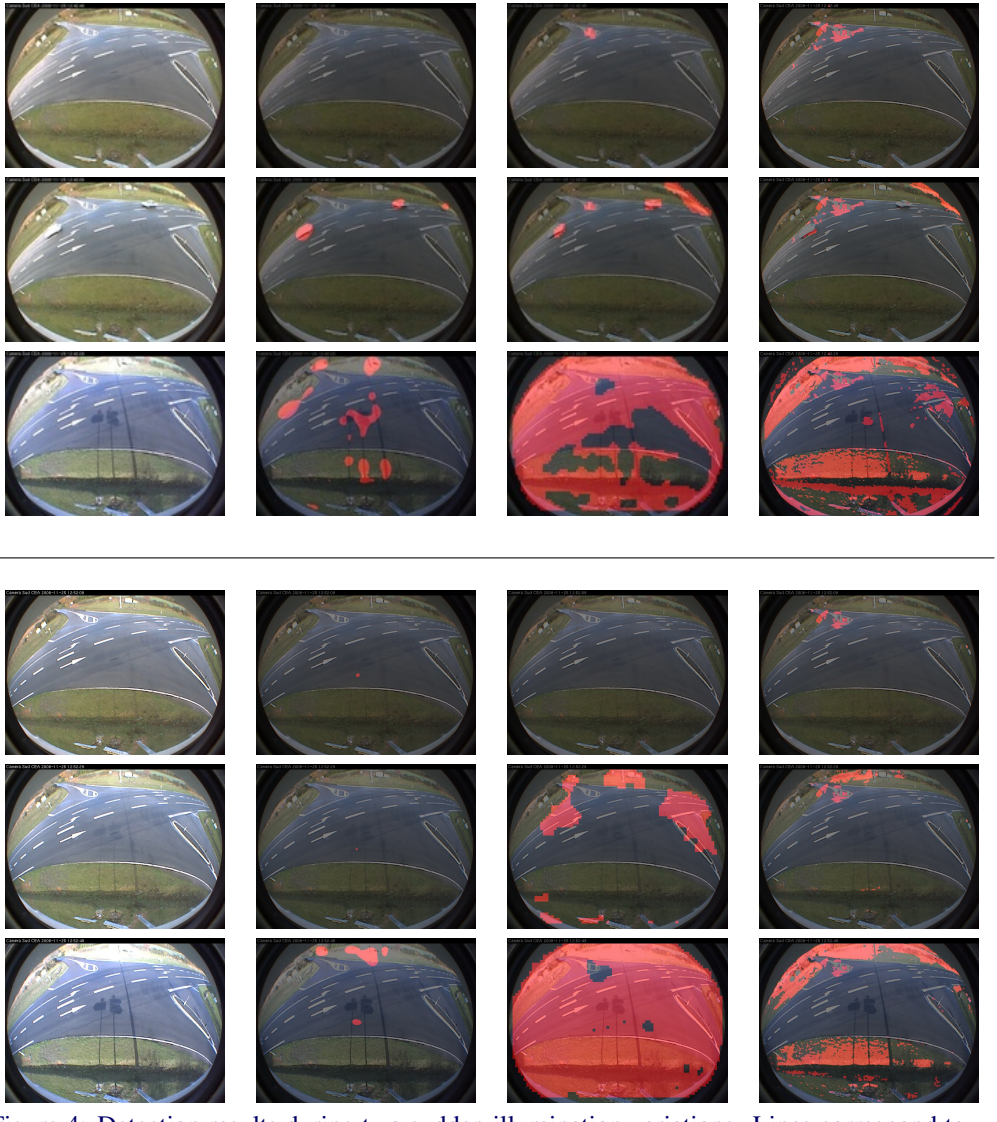

Figure 4: Detection results during two sudden illumination variations. Lines correspond to two groups of three consecutive frames. First column is the image acquired by the camera. Second column is our segmentation results. Third column is Chen *et al.*'s results. Fourth column is Stauffer and Grimson's.

5 Conclusion

We propose in this paper a simple yet effective background subtraction algorithm. A probability density function is computed by examining matching failures between the background model and the query image. We successfully apply this algorithm to the challenging case of PTZ cameras performing a guard tour and for which illumination issues are critical. Experiments show that the proposed approach is more robust to this phenomenon than other methods. Our algorithm successfully detects blobs with a precision sufficient as a first step

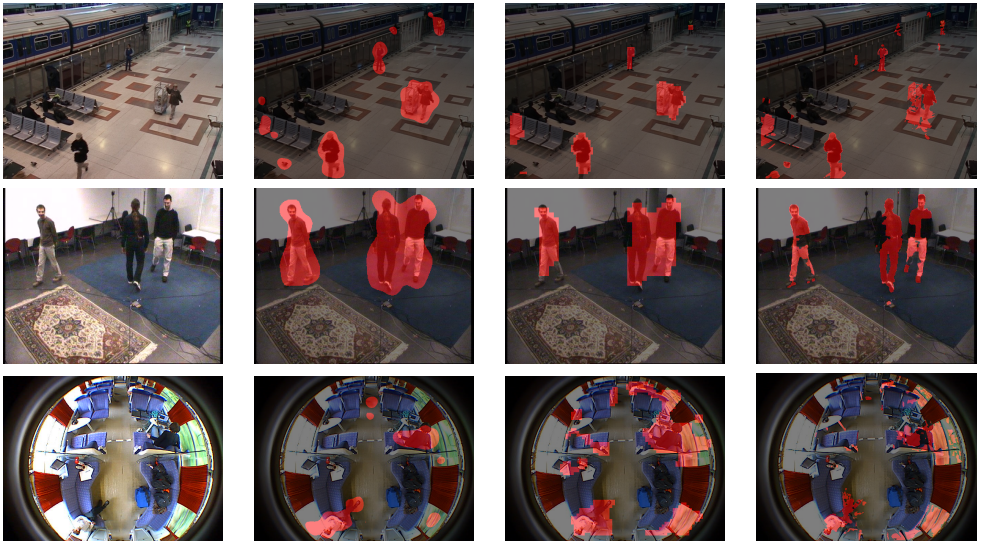

Figure 5: Detection results on PETS, EPFL and *train* sequences. First column is the image acquired by the camera. Second column is our segmentation results. Third column is Chen *et al.*'s results. Fourth column is Stauffer and Grimson's.

Method	Statistic	seq1	seq2	Train	Light change	PETS
Our method	Recall	0.68	0.61	0.83	0.53	0.8
	Precision	0.77	0.61	0.73	0.58	0.65
Chen *et al.* [6]	Recall	0.51	0.47	0.63	0.51	0.73
	Precision	0.69	0.24	0.61	0.69	0.84
Stauffer and Grimson	Recall	0.44	0.56	0.5	0.22	0.63
	Precision	0.55	0.12	0.44	0.46	0.6

Table 1: Comparison of detection results on various sequences

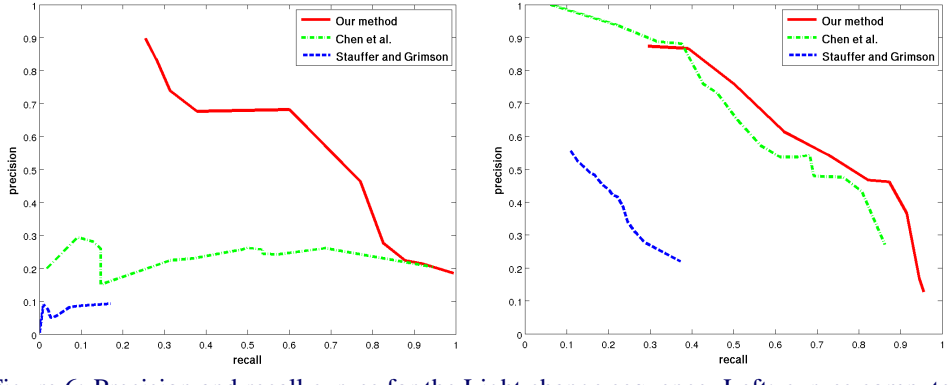

Figure 6: Precision and recall curves for the Light change sequence. Left: curves computed on a subsequence reduced to the period of an important change in illumination. Right: curves computed on a subsequence where there is no sudden change in illumination.

toward an object detection application.

References

[1] P. Azzari, L. Di Stefano, and A. Bevilacqua. An effective real-time mosaicing algorithm apt to detect motion through background subtraction using a ptz camera. In *AVSS*, 2005.

[2] Herbert Bay, Andreas Ess, Tinne Tuytelaars, and Luc Van Gool. Surf: Speeded up robust features. In *CVIU*, 2008.

[3] A. Bevilacqua and P. Azzari. High-quality real time motion detection using ptz cameras. In *AVSS*, 2006.

[4] Kiran Bhat, Mahesh Saptharishi, and Pradeep K. Khosla. Motion detection and segmentation using image mosaics. In *ICME*, 2000.

[5] Thierry Bouwmans, Fida El Baf, and Bertrand Vachon. Background Modeling using Mixture of Gaussians for Foreground Detection - A Survey. *Recent Patents on Computer Science*, 2008.

[6] Yu-Ting Chen, Chu-Song Chen, Chun-Rong Huang, and Yi-Ping Hung. Efficient hierarchical method for background subtraction. In *Pattern Recognition*, 2007.

[7] Rita Cucchiara, Andrea Prati, and Roberto Vezzani. Advanced video surveillance with pan tilt zoom cameras. In *Proc. of Workshop on Visual Surveillance (VS) at ECCV*, 2006.

[8] Alberto Del Bimbo, Fabrizio Dini, Andrea Grifoni, and Federico Pernici. Exploiting single view geometry in pan-tilt-zoom camera networks. In *Proc. of ECCV Int.'l Workshop on Multi-camera and Multi-modal Sensor Fusion (M2SFA2)*, 2008.

[9] Yoann Dhome, Nicolas Tronson, Antoine Vacavant, Thierry Chateau, Christophe Gabard, Yann Goyat, and Dominique Gruyer. A benchmark for background subtraction algorithms in monocular vision: a comparative study. In *IPTA*, 2010.

[10] Richard O. Duda and Peter E. Hart. *Pattern Classification and Scene Analysis*. John Wiley & Sons, 1973.

[11] Shireen Y. Elhabian, Khaled M. El-Sayed, and Sumaya H. Ahmed. Moving object detection in spatial domain using background removal techniques - state-of-art. In *Recent Patents on Computer Science*, 2008.

[12] Lionel Robinault, Stéphane Bres, and Serge Miguet. Real time foreground object detection using ptz camera. In *VISAPP*, 2009.

[13] C. Stauffer and W. E. L. Grimson. Adaptive background mixture models for real-time tracking. In *CVPR*, 1999.

[14] Rémi Trichet and Bernard Mérialdo. Keypoints labeling for backgroung substraction in tracking applications. In *ICME, IEEE International Conference on Multimedia Expo, Hannover, Germany*, 06 2008.