Background Subtraction for PTZ Cameras Performing a Guard Tour and Application to Cameras with Very Low Frame Rate

C. Guillot[1], M. Taron[1], Patrick Sayd[1], Q.C. Pham[1], C. Tilmant[2], and J.M. Lavest[2]

[1] CEA, LIST, Vision and Content Engineering Laboratory
BP 94, Gif-sur-Yvette, F-91191 France
[2] LASMEA UMR 6602, PRES Clermont Université/CNRS
63177 Aubière cedex, France

Abstract. Pan Tilt Zoom cameras have the ability to cover wide areas with an adapted resolution. Since the logical downside of high resolution is a limited field of view, a guard tour can be used to monitor a large scene of interest. However, this greatly increases the duration between frames associated to a specific location. This constraint makes most background algorithms ineffective. In this article we propose a background subtraction algorithm suitable to cameras with very low frame rate. Its main interest consists in the resulting robustness to sudden illumination changes. The background model which describes a wide scene of interest consisting of a collection of images can thus be successfully maintained. This algorithm is compared with the state of the art and a discussion regarding its properties follows.

1 Introduction

While the number of cameras used in public areas constantly increases, a strong effort is made to develop robust algorithm able to automate scene monitoring. Background subtraction is a popular pre-processing task often required to introduce scene understanding in video sequences.

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 lies in its limited field of view.

When dealing with 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, allows the detection of objects of interest by estimating a distance to the current image. As for PTZ camera, to maintain a whole background model is challenging since the necessary information is rarely available.

In this article, a PTZ camera performing a guard tour over a wide area is used to detect objects of interest. The camera follows a predefined set of positions

R. Koch et al. (Eds.): ACCV 2010 Workshops, Part I, LNCS 6468, pp. 33–42, 2011.

(pan, tilt, zoom) covering the area at an adapted resolution. For each of these positions it can be considered that we are in the case of a static camera suffering a very low frame rate (approximately 1 image every 10 to 20 seconds). Such a duration between frames constitutes a major difficulty since the background model will not be continuously updated and show important disparities in terms of illumination between the model and the current image.

This article presents a thorough study of a very low update rate background subtraction algorithm. It briefly reviews the related work (section 2), then presents a previous contribution of the authors (section 3) which has motivated this study. A comparison between local texture descriptors and the introduction of a more robust feature descriptor is then presented (section 4.2). A discussion regarding the background model update strategy follows (section 4.3) and additional experimental results are provided in section 5.

2 Related Work

There exist many background subtraction techniques in the literature, most of which are designed for static cameras with a frame rate above 12fps. Starting with basic frame differencing [1], it was soon necessary to build more evolved frameworks to describe the background. Stauffer and Grimson [2] first introduced a popular statistical approach based on a mixture of Gaussian distributions to model the luminance of each pixel. The model is updated at each frame to account for the variations of the background. An overview of background subtraction methods based on mixtures of Gaussians is given by Bouwmans *et al.* [3]. Elgammal *et al.* [4] have even achieved greater accuracy by substituting the MoG model with kernel density estimator.

Single pixel luminance does not carry sufficient information to address the complexity of outdoor scenes. It was therefore necessary to introduce spatial and temporal coherence in background subtraction algorithms. In [4], classification as background was enforced by considering the distribution model of neighbouring pixels. Background description models were also improved to carry dynamic information based on optical flow estimates [5]. Even when dealing with static backgrounds, accounting for sets of pixels provide better results. This motivated the work of Chen *et al.*'s [6], where texture descriptor is considered based on the tiling of the image with 8×8 blocks. This descriptor encodes a local colour contrast histogram with 48 parameters and increases robustness to illumination variations. This methods was proved to very efficient in [7] and is used in the remainder of this article to compare the performances of our background subtraction algorithm.

Zhu *et al.*[8] proposed a background subtraction algorithm based on the extraction of Harris keypoints and SIFT descriptors but which can only detect moving objects.

In the specific case of PTZ cameras, most approaches are based on the creation of a mosaic of the scene background. New images from the camera are registered on the mosaic as a prerequisite to background subtraction model update

[9,10,11,12]. The drawback of these methods is that there is no global update of the background model. There is no warranty that the model of an area that has not been visited for a while is usable. Therefore it turns out that these methods are more suitable to the tracking or moving object than the complete modelling of a large scene of interest.

3 Background Subtraction by Keypoint Density Estimation

In [13] we presented a background subtraction method based on the estimation of the density of non matching keypoints. The motivation for this method came from the fact that edge descriptor reveal themselves more robust to illumination changes than texture descriptors. This algorithm has been proved to be very effective in the experimental context of a PTZ camera.

We assume that keypoints which cannot be matched from the current image to the background image belong to objects of interest. Harris keypoints are extracted from both images and SURF [14] descriptor are computed on both images. Due to the mechanical error of the PTZ camera images are first registered using keypoints with the highest Harris score (strong edges).

Because Harris keypoints are not stable, corresponding points may not be present on both images for matching. We have used the union of keypoint location on both images prior to the matching and classification of points based on the Euclidean distance in the space of SURF descriptors.

Once we have a set of non matching keypoints, we use kernel smoothing techniques to estimate a continuous density $\hat{d_h}$:

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

with $(p_1, \ldots, p_N)$ the set of N non matching keypoints, K a Gaussian kernel function and h a smoothing parameter which specifies the influence of each observation on its neighbourhood. Pixels are classified as foreground if $N\hat{d_h} > s$ and as background otherwise.

Background model is updated according to the following equation:

$$bg_n(x) = \begin{cases} bg_{t-1}(x) & \text{if } N\hat{d_h} > s \\ bg_{t-1}(x)\frac{N\hat{d_h}(x)}{s} + img_t(x)\left(1 - \frac{N\hat{d_h}(x)}{s}\right) & \text{otherwise} \end{cases} \tag{2}$$

At this point, it is important to note that this approach presents some limitations in terms of implementation:

- SURF Gradient is based on image gradient and normalised to achieve better robustness to changes in illumination. In poorly textured areas the normalisation step amplifies the influence of camera noise and the artefacts due to image compression. This prevents some keypoints from matching in homogeneous areas, which leads to false positives.

– Harris threshold is chosen especially to prevent keypoints candidates to be located in homogeneous areas. Setting this value is still empirical and sequence dependant.
– Considering a variable set of points for background subtraction has lead to the use of a plain image as a background model. This is not plainly satisfactory because the update (eq. 2) actually blurs the model, and might lead to the creation of ghosts and false detection.

The remainder of the article presents an update to the algorithm in order to address these limitations.

4 Texture Descriptors for Background Subtraction

We propose a background subtraction algorithm similar to the keypoint which relies on a regular grid of modified SURF descriptors as a background model instead of a variable set of keypoints. These descriptors can be computed on weakly textured areas (sec. 4.1). This algorithm is as effective as the keypoint density algorithm but Harris feature point extraction and its associated manual threshold is no longer necessary.

Background subtraction is performed by computing the distance of a descriptor from a point of the grid to the corresponding one in the background model. Once we have a set of matching and not matching points we use equations 1 as a post processing to smooth out the classification results. The threshold is set to avoid false alarms in case of isolated detection. Meanwhile isolated mis-detection are automatically filled in by the neighbouring detections.

4.1 Weighting the SURF Descriptor

The SURF descriptor has been thought to be discriminative when computed on textured zones, but the normalisation process renders this inefficient, when in low textured areas, noise overcomes gradient information.

The intuition is that if there is no gradient information we can't decide at a local level whether a pixel is foreground or background. Thus, we arbitrarily decide that two low textured areas (those where the SURF norm is low) should match.

To do so, we consider the distribution of the norm of the SURF descriptors on a set of sequences displaying texture and homogeneous areas (Fig. 1). The resulting distribution presents two modes. The lowest corresponding to homogeneous, it is removed with an appropriate weighting function applied to the SURF descriptor values.

$$D'_{SURF} = D_{SURF} * f(\|D_{SURF}\|) \quad \text{with} \quad f(x) = \begin{cases} 0 & \text{if } x < 120 \\ \dfrac{x - 120}{480 - 120} & \text{if } 120 \leq x \leq 480 \\ 1 & \text{if } 480 < x \end{cases} \tag{3}$$

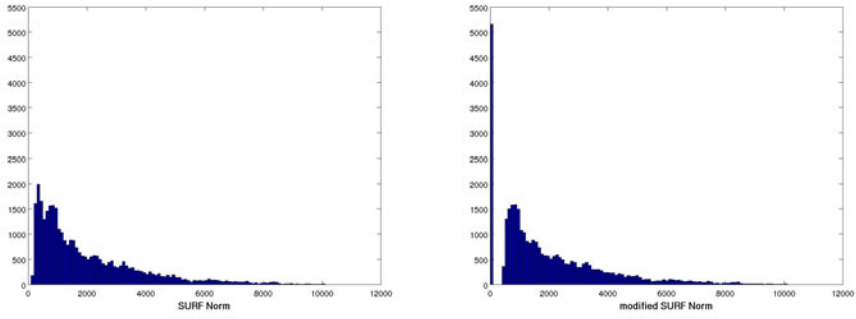

Fig. 1. Left: Histogram of the norm of SURF descriptor before normalisation. Right: Histogram of the norm of SURF descriptor before normalisation and after applying equation 3.

As a consequence, if the SURF descriptor is computed on a textured zone its norm remains 1. If it is computed on a zone which is not textured, it is set to 0, with a continuous transition between these two cases.

4.2 Evaluating the Quality of the Texture Descriptor

Chen *et al.*[6] have designed their own texture descriptor to perform background subtraction. It encodes a histogram of contrast between the different colour components. However we can question the choice of such a descriptor since there exists well known other descriptors used in other fields of computer vision.

We have compared the Chen descriptor to the SURF descriptor [14] on a sequence presenting challenging changes in illumination (Fig. 3a). To assess the quality of the descriptors only, we performed background subtraction on this sequence with no post processing of any kind. Descriptors are computed on the same regular grid and the classification as foreground or background is done only according to the distance toward the corresponding descriptor on a reference image (no statistical modelling in the space of descriptors). We Have computed ROC curves on this sequence with a variation of the classification threshold (Fig. 2).

Results are very poor if one consider a single frame as a reference (very strong disparities between images). However the obtained precision with SURF is always twice as better than the Chen Descriptor. If one considers consecutive images, there is a global increase in robustness with the use of SURF descriptors and the modified SURF descriptors.

4.3 Background Update

As descriptors are computed on a regular grid rather than a set of keypoints. It is now possible to handle the background model update in the space of descriptors rather than the image space. If $D_{\mathrm{Bkg},t}$ is a background descriptor and $D_{\mathrm{Img},t}$

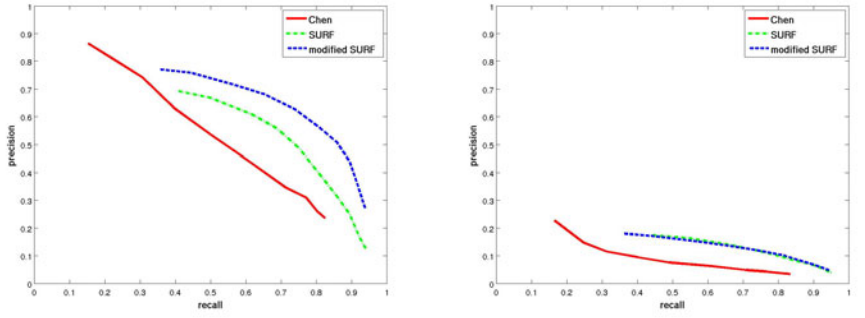

Fig. 2. Precision and recall curves of the Chen, SURF and modified SURF (sec. 4.1) descriptors computed on the sequence from Fig. 3a. Left: comparing two consecutive images. Right: comparing one specific image to all images of the sequence.

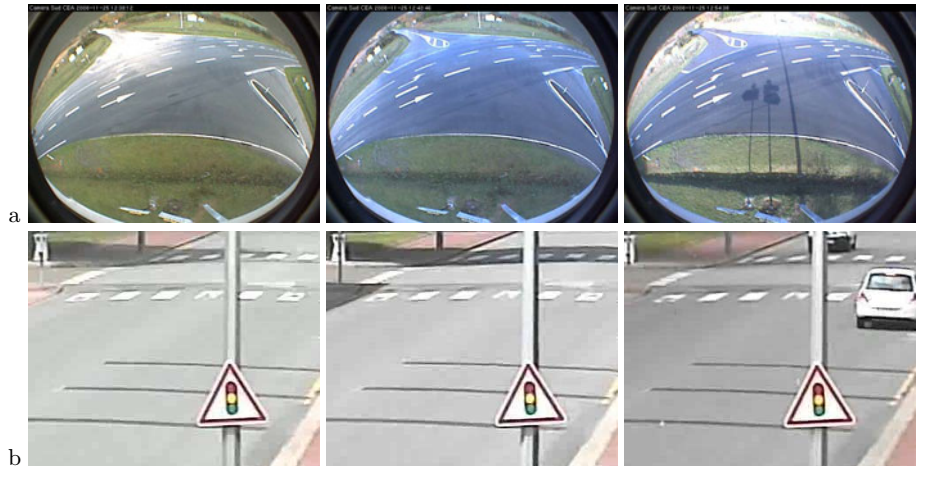

Fig. 3. Test Sequences. These sequences present important illumination changes, shadows and reflections on a rain-soaked road.

is a descriptor computed from the current image at time t and classified as background, then we use the following updating rule:

$$D_{\mathrm{Bkg},t+1} = \alpha D_{\mathrm{Bkg},t} + (1-\alpha) D_{\mathrm{Img},t} \tag{4}$$

For our application the learning rate α is chosen rather high ($\alpha > 0.25$). As we consider sequences with very low frame rate, it is necessary to update the model quickly to follow the global illumination changes.

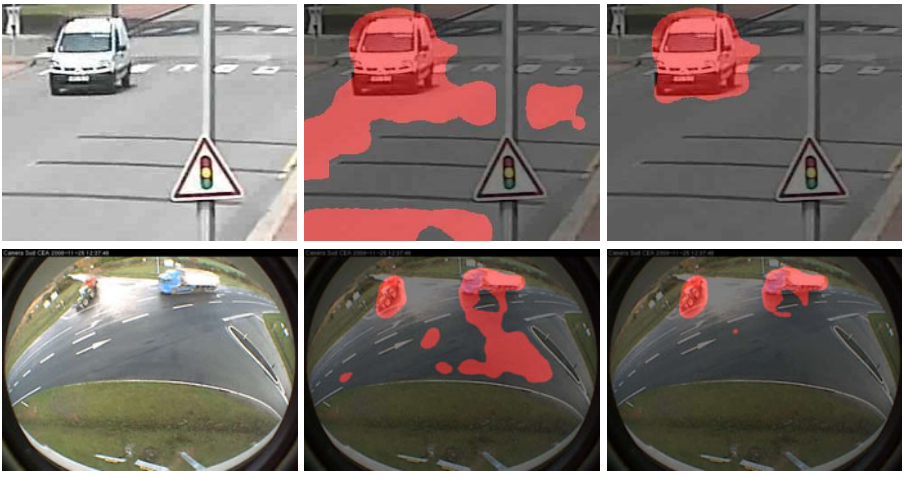

Fig. 4. Detection results on a sequence with very low textured zones and compression artefacts. First column: original image. Second column: segmentation result using keypoints density with a low threshold on the Harris score (Harris points can be located in homogeneous areas). Third column: grid of modified SURF descriptors.

Table 1. Comparison of detection results on various sequences

Method	Statistic	PTZ	Train	Outdoor1	PETS	Outdoor2
Modified SURF grid	Recall	0.74	0.77	0.69	0.75	0.63
	Precision	0.74	0.79	0.67	0.75	0.70
Keypoint density	Recall	0.61	0.83	0.53	0.8	0.64
	Precision	0.61	0.73	0.58	0.65	0.67
Chen *et al.*[6]	Recall	0.47	0.63	0.51	0.73	0.55
	Precision	0.24	0.61	0.69	0.84	0.60
Stauffer and Grimson	Recall	0.56	0.5	0.22	0.63	0.46
	Precision	0.12	0.44	0.46	0.6	0.55

5 Experimental Results

The first part of the experiments is devoted to the comparison of the SURF and modified SURF descriptor. Figure 4 shows the kind of issue which may arise on poorly textured areas and how the modified SURF descriptor deals with it. On these areas, the compression artefact create unstructured gradients which make the original SURF descriptor ineffective. Figure 4 shows that the modified SURF descriptor can be computed on uniform areas while not generating false mismatches. Figure 5 displays ROC curves which confirms quantitatively what can be observed on figure 4.

Figure 6 presents qualitative results for the case of a PTZ camera performing a guard tour. The time elapsed between consecutive frames is 24 seconds. Notice

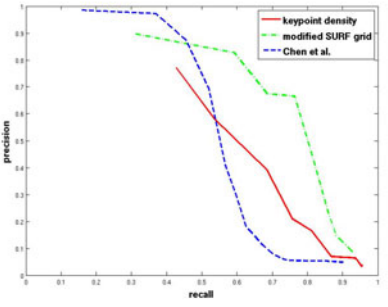

Fig. 5. Precision and recall curves computed on the *light change* sequence. The threshold for the keypoint density algorithm is the same as the one used in figure 4.

Fig. 6. PTZ Sequence

the green borders on the image due to the registration between images acquired during the tour.

The second part of the experiments compares the modified SURF grid background subtraction algorithm to Chen *et al.*'s [6] algorithm and our previous algorithm based on keypoints density estimation [13]. The application to PTZ cameras performing a Guard tour is equivalent to a fixed cameras with a very low frame rate. Therefore we have applied the algorithms to fixed cameras presenting challenging sequences and artificially lowered the frame rate to one image every 20 seconds.

Figure 7 shows qualitative results. Whereas the PETS sequence is not challenging in terms of illumination variation they show that our algorithm behaves well on weakly textured scenes. The *train* sequence is another example sequence for which our algorithm is stable even when sudden changes in illumination occur. On these sequences modified SURF grid behaves as well as the keypoint density algorithm.

Figure 8 shows quantitative results on two sequences where sudden changes in illumination occur. Table 1 sums up the results from various sequences and shows that our algorithm is more stable than others. The presented statistics may seem low at first sight, but these were computed in the most challenging experimental conditions. Moreover, as can be seen on figure 7, the loss of precision of our

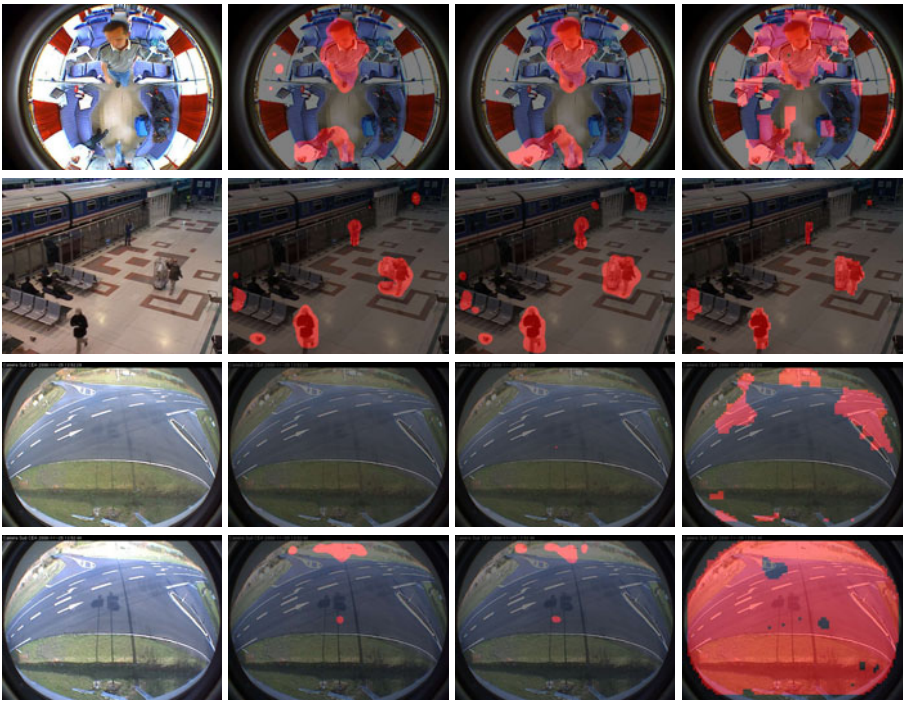

Fig. 7. Qualitative results obtained in various situations. First row is captured on board of a train. Second row is a sequence extracted from the PETS 2006 challenge (http://www.cvg.rdg.ac.uk/PETS2006/). Rows 3 and 4 are consecutive images extracted from the sequence in Fig. 3a. First column: original image. Second column: our algorithm. Third column: keypoint density algorithm [13]. Fourth column: Chen *et al.*'s algorithm.

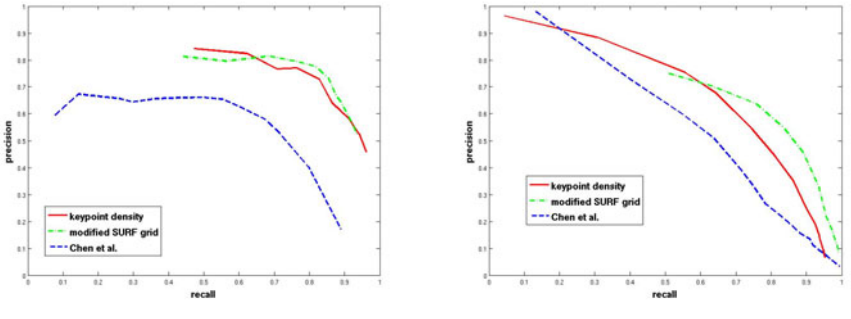

Fig. 8. Precision and recall curves. Left: *train* sequence. Right: *light change 2* sequence.

algorithm is inherent to the method and mainly due to the fact that it always over segment foreground blobs. In no case does it generate actual false alarms.

6 Conclusion

We have propose a simple yet efficient background subtraction algorithm. We use a modified version of the SURF descriptor which can be computed on weakly textured areas. We successfully apply our algorithm in the challenging context of PTZ cameras performing a guard tour and for which illumination issues are critical. Our algorithm successfully detects blobs with a sufficient accuracy used as a first step toward object detection application.

References

1. Jain, R., Nagel, H.: On the analysis of accumulative difference pictures from image sequences of real world scenes, vol. 1, pp. 206–213 (1979)
2. Stauffer, C., Grimson, W.E.L.: Adaptive background mixture models for real-time tracking. In: CVPR (1999)
3. Bouwmans, T., El Baf, F., Vachon, B.: Background Modeling using Mixture of Gaussians for Foreground Detection - A Survey. Recent Patents on Computer Science (2008)
4. Elgammal, A.M., Harwood, D., Davis, L.S.: Non-parametric model for background subtraction. In: Vernon, D. (ed.) ECCV 2000. LNCS, vol. 1843, pp. 751–767. Springer, Heidelberg (2000)
5. Mittal, A., Paragios, N.: Motion-based background subtraction using adaptive kernel density estimation. In: CVPR, vol. 2, pp. 302–309 (2004)
6. Chen, Y.T., Chen, C.S., Huang, C.R., Hung, Y.P.: Efficient hierarchical method for background subtraction. Pattern Recognition (2007)
7. Dhome, Y., Tronson, N., Vacavant, A., Chateau, T., Gabard, C., Goyat, Y., Gruyer, D.: A benchmark for background subtraction algorithms in monocular vision: a comparative study. In: IPTA (2010)
8. Zhu, Q., Avidan, S., Cheng, K.T.: Learning a sparse, corner-based representation for time-varying background modelling. In: ICCV, vol. 1, pp. 678–685 (2005)
9. Bhat, K., Saptharishi, M., Khosla, P.K.: Motion detection and segmentation using image mosaics. In: ICME (2000)
10. Cucchiara, R., Prati, A., Vezzani, R.: Advanced video surveillance with pan tilt zoom cameras. In: Proc. of Workshop on Visual Surveillance (VS) at ECCV (2006)
11. Azzari, P., Di Stefano, L., Bevilacqua, A.: An effective real-time mosaicing algorithm apt to detect motion through background subtraction using a ptz camera. In: AVSS (2005)
12. Robinault, L., Bres, S., Miguet, S.: Real time foreground object detection using ptz camera. In: VISAPP (2009)
13. Guillot, C., Taron, M., Sayd, P., Pham, Q.C., Tilmant, C., Lavest, J.M.: Background subtraction adapted to ptz cameras by keypoint density estimation. In: BMVC (2010) (to appear in bmvc 2010: supplied as supplementary material)
14. Bay, H., Ess, A., Tuytelaars, T., Gool, L.V.: Surf: Speeded up robust features. In: CVIU (2008)