Neural Moving Object Detection by Pan-Tilt-Zoom Cameras

Alessio Ferone[1,⋆], Lucia Maddalena[2], and Alfredo Petrosino[1]

[1] Department of Applied Science
University of Naples Parthenope, Italy
{alessio.ferone,alfredo.petrosino}@uniparthenope.it
[2] Institute for High-Performance Computing and Networking
National Research Council, Italy
lucia.maddalena@cnr.it

Abstract. Automated video surveillance using video analysis and understanding technology has become an important research topic in the area of computer vision. Most cameras used in surveillance are fixed, allowing to only look at one specific view of the surveilled area. Recently, the progress in sensor technologies is leading to a growing dissemination of Pan-Tilt-Zoom (PTZ) cameras, that can dynamically modify their field of view. Since PTZ cameras are mainly used for object detection and tracking, it is important to extract moving object regions from images taken with this type of camera. However, this is a challenging task because of the dynamic background caused by camera motion.

After reviewing background subtraction-based approaches to moving object detection in image sequences taken from PTZ cameras, we present a neural-based background subtraction approach where the background model automatically adapts in a self-organizing way to changes in the scene background. Experiments conducted on real image sequences demonstrate the effectiveness of the presented approach.

Keywords: Visual Surveillance, Motion Detection, Background Subtraction, Self Organization, Artificial Neural Network, PTZ Camera.

1 Introduction

Moving object detection is an important research problem in the field of video surveillance because it provides a focus of attention for recognition, classification, and activity analysis, allowing the analysis only of moving pixels [1].

Background subtraction is one of the most common approaches to moving object detection by static cameras (see surveys in [2,3,4,5]). It consists in constructing and updating a model of the fixed background, and detecting moving objects as those that do not belong to the model. Compared to other approaches, such as optical flow, no assumptions about the velocity of the object are made

⋆ Corresponding author.

B. Apolloni et al. (Eds.): *Neural Nets and Surroundings*, SIST 19, pp. 129–138.
DOI: 10.1007/978-3-642-35467-0_14

and it does not suffer of the foreground aperture problem. Moreover, background subtraction approaches are usually computationally affordable for real world applications. Nevertheless, the background subtraction approach is highly sensitive to dynamic scene changes due to lighting and extraneous events, with the consequent need for a suitable adaptation of the background model [6,7].

In order to overcome the limitations of static cameras, in the recent years the progress in sensor technologies has lead to a growing adoption of PTZ cameras, that can change their field of view through the use of panning, tilting, and zooming (i.e., moving left and right, up and down, closer and farther away), thus enabling to focus the attention on automatically selected areas of interest [8,9,10,11]. However this type of camera has introduced new challenges, because even pixels belonging to static objects appear to move in the camera frame. This effect is known as *ego-motion* and its estimation and compensation represents one of the main objectives of the research in the active video area [10].

Extensive research on moving object detection by PTZ cameras relies on background subtraction, where the scene background is obtained either by a mosaic image of the background or by ego-motion compensation of the background (see section 2). In this paper we embed an ego-motion compensation technique into a neural-based background subtraction approach to moving object detection. The background model automatically adapts in a self-organizing way to changes in the scene background. Background variations arising in a usual stationary camera setting are accurately handled by the neural background model originally proposed for this type of setting [12], while handling of variations due to the PTZ camera movement is ensured by having the neural background model to automatically compensate the eventual ego-motion, estimated at each time instant.

The paper is organized as follows. In section 2 we review the literature concerning background subtraction techniques applied to PTZ cameras. In section 3 we sketch the neural self-organizing model for image sequences and describe how to embed the background compensation in order to handle image sequences taken from PTZ cameras. In section 4 we present some preliminary results achieved with the implementation of the presented approach. Section 5 includes concluding remarks.

2 Background Subtraction Approaches for PTZ Cameras

Approaches to moving object detection for PTZ cameras based on background subtraction can be classified into two categories: mosaiced background-based and background compensation-based approaches.

The *mosaiced background*-based approaches [13], also referred to as *frame to background* methods [14], create and maintain a mosaic image of the whole scene background, on which background subtraction techniques are applied to extract moving object regions [15,16,17,18,19,20,21,22,23,24,25,26,27]. A mosaic image, also referred to as panorama, is a compound image built through properly composing multiple images of the same scene taken from different viewpoints and warping them into a common reference coordinate system. The result consists in a single image of the scene, with greater resolution or spatial extent [28]. The basic idea behind mosaiced background-based approaches to moving object detection consists in the use of traditional background subtraction techniques on a mosaic image of the observed scene built by the PTZ camera images. Subsequent video frames are registered into the mosaic coordinate system first, in order to locate the subset of the mosaic background model corresponding to the current frame. This subset is then adopted to perform background subtraction and update.

The *background compensation*-based approaches [13], also named *frame-by-frame* methods [14], estimate the transformation parameters between time related images by using corresponding features extracted from these images, and create a difference image and/or a motion-compensated background in order to detect moving object regions [29,30,31,32,14,33,34,35,36,13,37]. These approaches first estimate the apparent motion of the static background due to the camera movement, and then compensate the previous image for the estimated motion. Moving objects are usually detected by frame differencing, i.e., by subtracting the compensated previous image and the current image.

Both mosaiced background-based and background compensation-based approaches need to register (or align) various pairs (or collections) of images. Therefore, assuming a motion model (e.g., affine or projective) relating pixel coordinates in one image to pixel coordinates in another, image matching techniques should be envisaged to detect and match salient features among images, with the final aim of estimating motion parameters [38].

Most methods deal with PT cameras, and some of them are applicable to a PTZ camera. Indeed, in the case of zooming, the motion parallax problem arises, where the apparent motion of objects closer to the image plane is higher than that of objects that are further away. Instead, when there is no motion parallax, the apparent motion of all objects in the scene does not depend on their distance from the camera. This can be guaranteed by rotating the camera around its optical center (approach taken by many commercial systems) and holds also for most cameras when objects are far from it [20].

According to the adopted image matching techniques, the algorithms can be classified into two main families: intensity-based (or direct) methods and feature-based methods. *Intensity-based methods* usually attempt at iteratively estimating the transformation parameters by minimizing an error function based on pixel-by-pixel brightness differences in overlapping areas. Exploiting the information associated with every single pixel, these methods achieve highly accurate registration; however, they are computationally demanding. *Feature-based methods* first extract distinctive features (i.e, regions, lines, histograms, intensity projection profiles, and keypoints) from each image, then match these features to establish a global correspondence, and finally estimate the geometric transformation between the images. Outlier detection helps to filter out bad tracked features or mismatched points.

Moreover, in the case of mosaiced background-based approaches, techniques must be developed to compute a globally consistent set of alignments among multiple images existing in the panorama, and to efficiently discover which images overlap one another [38]. Two classes of mosaicing algorithms may be distinguished, with regard to the number of frames that are simultaneously combined: global and sequential registration methods. *Global registration methods* compute the best alignment among several images by simultaneously minimizing the misregistration between all the overlapping pairs of images. They achieve accurate geometric reconstruction, but are computationally intensive and require all the images to be known in advance. *Sequential algorithms* allow the construction of a mosaic by continuously combining new images as soon as they become available. Every new image is aligned with the previous one (frame-to-frame registration) or with the mosaic built thus far (frame-to-mosaic registration). These methods allow faster computation and do not need all the images in advance; however, the achieved registration is only locally optimal, and may lead to error accumulation.

From the above analysis we can conclude that both the approaches to moving object detection by PTZ cameras have their pros and cons. Background compensation-based approaches usually require less computational cost and memory storage, while the main advantage of creating a mosaic model is to save the information of the scene background that is readily available whenever the camera moves to a new position or/and returns to a previous captured location.

It should be mentioned that other approaches to moving object detection by PTZ cameras exist that are not based on background subtraction. This is the case of the so-called *optical flow clustering*-based approaches, which compute clusters of dense or sparse optical flows in order to identify regions of movement [39,40,41]. Other methods exploit further information of the scene settings. This is the case of the works in [42,43], where the authors consider PTZ cameras performing a guard tour, following a predefined set of positions covering the area under surveillance; and the works in [44,45], where motion parameters are measured by using specialized hardware. Finally, a comprehensive introduction of PTZ camera networks, highlighting how cameras cooperation and reconfiguration can be exploited for active surveillance, has been recently provided [10].

3 Self-Organizing Background Subtraction for PTZ Cameras

Relying on the recent SOBS algorithm [12], we build the sequence background model by learning in a self-organizing manner image sequence variations, seen as trajectories of pixels in time. A neural network mapping method is proposed to use a whole trajectory incrementally in time fed as an input to the network. Each neuron computes a function of the weighted linear combination of incoming inputs, and therefore can be represented by a weight vector, obtained collecting the weights related to incoming links. An incoming pattern is mapped to the neuron whose set of weight vectors is most similar to the pattern, and weight vectors in a neighborhood of such node are updated. Differently from [12], at each time instant the neural background model automatically compensates the eventual ego-motion due to the PTZ camera, leading to what will be called the PTZ-SOBS algorithm.

3.1 Neural Model Representation

Given an image sequence $\{I_t\}$, for each pixel $\mathbf{p}$ in the image domain D, we build a neuronal map consisting of $n \times n$ weight vectors $m_t^{i,j}(\mathbf{p}), i, j = 0, \ldots, n-1$, which will be called a *model* for pixel $\mathbf{p}$ and will be indicated as $M_t(\mathbf{p})$:

$$M_t(\mathbf{p}) = \left\{ m_t^{i,j}(\mathbf{p}),\ i, j = 0, \ldots, n-1 \right\}. \tag{1}$$

If every sequence frame has N rows and P columns, the complete set of models $M_t(\mathbf{p})$ for all pixels $\mathbf{p}$ of the t-th sequence frame I_t is organized as a 2D neuronal map B_t with $n \times N$ rows and $n \times P$ columns, where the weight vectors $m_t^{i,j}(\mathbf{p})$ for the generic pixel $\mathbf{p} = (x, y)$ are at neuronal map position $(n \times x + i, n \times y + j)$, $i, j = 0, \ldots, n-1$:

$$B_t(n \times x + i, n \times y + j) = m_t^{i,j}(\mathbf{p}),\ i, j = 0, \ldots, n-1. \tag{2}$$

This configuration of the whole neuronal map B_t allows to easily take into account the spatial relationship among pixels and corresponding weight vectors.

3.2 Background Subtraction and Neural Model Update

At each subsequent time step t, background subtraction is achieved by comparing each pixel of the t-th sequence frame I_t with the model for that pixel. In the general case of image sequences taken from PTZ cameras, the incoming pixel $\mathbf{p}$ of the t-th sequence frame I_t could have moved as compared to the previous time $t-1$. Therefore, the current model $M_{t-1}(\mathbf{p})$, whose weight vectors are stored in B_{t-1} as described in Eq. (2), could be an improper model for actual pixel $\mathbf{p}$. In order to keep track of such spatial movements, we compute the homography H between sequence frames I_{t-1} and I_t, that allows to obtain, for each pixel $\mathbf{p}'$ of I_{t-1}, the corresponding pixel $\mathbf{p} = H\mathbf{p}'$ of I_t. This information is exploited in order to address the proper model for current pixel $\mathbf{p}$, to perform background subtraction, and to update the proper model through a selective weighted running average analogous to [12].

4 Experimental Results

Several experimental tests have been conducted to validate our approach to moving object detection in image sequences taken from PTZ cameras. In the following, qualitative and quantitative results will be described for the *Lab1* sequence, that represents a typical indoor situation critical for video surveillance systems. The *Lab1* sequence, publicly available in the download section

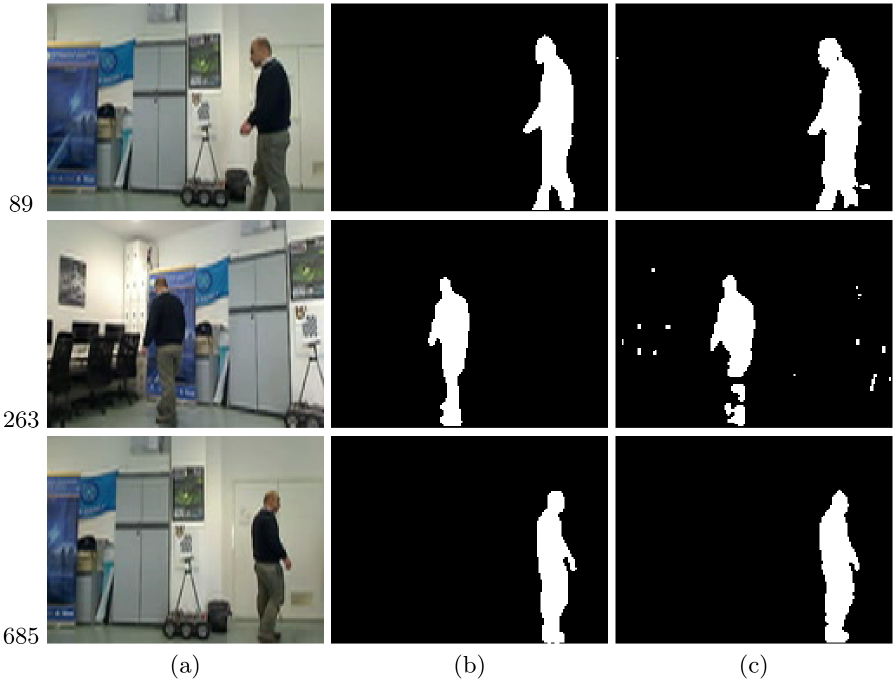

Fig. 1. Results of moving object detection for the frames 89, 263, and 685 of the *Lab1* sequence: (a) original frames; (b) ground truth masks; (c) moving object detection masks computed by the PTZ-SOBS algorithm

of `http://cvprlab.uniparthenope.it`, is an indoor sequence consisting of 886 frames of 160 × 120 spatial resolution. The scene consists of an office, where a person comes in and walks, and the PTZ camera follows the movement of the person. Representative frames together with obtained results are reported in Fig. 1, where we report the original sequence frames no. 89, 263, and 685 (first column), the corresponding ground truth masks (second column), and the moving object detection mask computed by the PTZ-SOBS algorithm (third column). In frame 89 a person has entered the office, and the camera is panning from right to left in order to follow him; the compensated background provides a good representation of the real background, and the corresponding detection mask is quite accurate. In frame 263 a person has moved further left, while the camera is still following him. The compensated background still provides a good representation of the real background. The corresponding detection mask is quite accurate, although few white pixels of the moving person and of the background have been misdetected. In frame 685 the person is going back from left to right, still followed by analogous movement of the camera. The compensated background still provides a good representation of the real background, as shown by the quite accurate corresponding detection mask.

The average segmentation accuracy results achieved by the proposed approach on the considered sequence, in terms of Precision, Recall and F_1 measure, are quite encouraging. Indeed, the achieved *Recall*=0.94 value ensures that most of the moving pixels are indeed detected as moving; at the same time, the high *Precision*=0.88 value indicates that only few of the pixels detected as moving are instead background pixels. The consequent high value of F_1=0.91, that is the weighted harmonic mean of *Precision* and *Recall*, allows us to deduce the overall high segmentation accuracy of the proposed approach.

5 Conclusions

We present an approach to the problem of moving object detection in image sequences taken from PTZ cameras based on the idea of exploiting the available knowledge concerning the self-organized learning behavior of the brain, which is the foundation of human visual perception, and traducing it into models and algorithms that can accurately solve the problem. A neural self-organizing background model is presented, that automatically adapts to variations in the scene background, both arising in a usual stationary camera setting and due to the PTZ camera movement. The compensated background model is adopted for accurate moving object detection, as demonstrated by experimental results on real image sequences.

References

1. Collins, R.T., Lipton, A.J., Kanade, T., Fujiyoshi, H., Duggins, D., Tsin, Y., Tolliver, D., Enomoto, N., Hasegawa, O., Burt, P., Wixson, L.: A system for video surveillance and monitoring. Technical Report CMU-RI-TR-00-12, Carnegie Mellon University, Pittsburgh, PA (2000)
2. Cheung, S.C.S., Kamath, C.: Robust techniques for background subtraction in urban traffic video. In: Panchanathan, S., Vasudev, B. (eds.) Proc. Visual Communications and Image Processing, SPIE, vol. 5308, pp. 881–892 (2004)
3. Elhabian, S., El Sayed, K., Ahmed, S.: Moving object detection in spatial domain using background removal techniques: State-of-art. Recent Patents on Computer Science 1(1), 32–54 (2008)
4. Piccardi, M.: Background subtraction techniques: a review. In: Proc. IEEE SMC, vol. 4, pp. 3099–3104 (October 2004)
5. Radke, R.J., Andra, S., Al-Kofahi, O., Roysam, B.: Image change detection algorithms: A systematic survey. IEEE Trans. Image Process. 14, 294–307 (2005)
6. Toyama, K., Krumm, J., Brumitt, B., Meyers, B.: Wallflower: principles and practice of background maintenance. In: Proc. ICCV, vol. 1, pp. 255–261 (1999)
7. Huang, K., Wang, L., Tan, T., Maybank, S.: A real-time object detecting and tracking system for outdoor night surveillance. Pattern Recognition 41(1), 432–444 (2008)
8. Cristani, M., Farenzena, M., Bloisi, D., Murino, V.: Background subtraction for automated multisensor surveillance: a comprehensive review. EURASIP J. Adv. Signal Process, 43:1–43:24 (February 2010)

9. Elgammal, A.: Figure-ground segmentation-pixel-based. In: Moeslund, T.B., Hilton, A., Krüger, V., Sigal, L. (eds.) Visual Analysis of Humans, pp. 31–51. Springer, London (2011)
10. Micheloni, C., Rinner, B., Foresti, G.: Video analysis in pan-tilt-zoom camera networks. EEE Signal Processing Magazine 27(5), 78–90 (2010)
11. Sheikh, Y., Javed, O., Kanade, T.: Background subtraction for freely moving cameras. In: 12th IEEE International Conference on Computer Vision, pp. 1219–1225 (October 2009)
12. Maddalena, L., Petrosino, A.: A self-organizing approach to background subtraction for visual surveillance applications. IEEE Trans. Image Process 17(7), 1168–1177 (2008)
13. Suhr, J.K., Jung, H.G., Li, G., Noh, S.I., Kim, J.: Background compensation for pan-tilt-zoom cameras using 1-d feature matching and outlier rejection, pp. 371–377. IEEE Computer Society Press, Los Alamitos (March 2011)
14. Micheloni, C., Foresti, G.L.: Real-time image processing for active monitoring of wide areas. Journal of Visual Communication and Image Representation 17(3), 589–604 (2006)
15. Bartoli, A., Dalal, N., Horaud, R.: Motion panoramas. Computer Animation and Virtual Worlds 15(5), 501–517 (2004)
16. Bevilacqua, A., Azzari, P.: A fast and reliable image mosaicing technique with application to wide area motion detection. In: Kamel, M.S., Campilho, A. (eds.) ICIAR 2007. LNCS, vol. 4633, pp. 501–512. Springer, Heidelberg (2007)
17. Del Bimbo, A., Lisanti, G., Masi, I., Pernici, F.: Continuous recovery for real time pan tilt zoom localization and mapping. In: 8th IEEE International Conference on Advanced Video and Signal-Based Surveillance (AVSS), pp. 160–165 (September 2011)
18. Hayman, E., Eklundh, J.O.: Statistical background subtraction for a mobile observer. In: Proceedings Ninth IEEE International Conference on Computer Vision 2003, vol. 1, pp. 67–74 (October 2003)
19. Jin, Y., Tao, L., Di, H., Rao, N., Xu, G.: Background modeling from a free-moving camera by multi-layer homography algorithm. In: 15th IEEE International Conference on Image Processing (ICIP), pp. 1572–1575 (October 2008)
20. Mittal, A., Huttenlocher, D.: Scene modeling for wide area surveillance and image synthesis. In: Proceedings IEEE Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 160–167 (2000)
21. Monari, E., Pollok, T.: A real-time image-to-panorama registration approach for background subtraction using pan-tilt-cameras. In: 8th IEEE International Conference on Advanced Video and Signal-Based Surveillance (AVSS), pp. 237–242 (September 2011)
22. Prati, A., Seghedoni, F., Cucchiara, R.: Fast dynamic mosaicing and person following. In: Proceedings of the 18th International Conference on Pattern Recognition, ICPR 2006, pp. 920–923. IEEE Computer Society, Washington, DC (2006)
23. Ren, Y., Chua, C.S., Ho, Y.K.: Statistical background modeling for non-stationary camera. Pattern Recognition Letters 24(1-3), 183–196 (2003)
24. Sankaranarayanan, K., Davis, J.W.: PTZ Camera Modeling and Panoramic View Generation via Focal Plane Mapping. In: Kimmel, R., Klette, R., Sugimoto, A. (eds.) ACCV 2010, Part II. LNCS, vol. 6493, pp. 580–593. Springer, Heidelberg (2011)
25. Sugaya, Y., Kanatani, K.: Extracting moving objects from a moving camera video sequence. Mem. Fac. Eng. Okayama Univ. 39, 56–62 (2005)

26. Xue, K., Liu, Y., Chen, J., Li, Q.: Panoramic background model for PTZ camera. In: 3rd International Congress on Image and Signal Processing (CISP), vol. 1, pp. 409–413 (October 2010)
27. Zhang, J., Wang, Y., Chen, J., Xue, K.: A framework of surveillance system using a ptz camera. In: 3rd IEEE International Conference on Computer Science and Information Technology (ICCSIT), vol. 1, pp. 658–662 (July 2010)
28. Bevilacqua, A., Azzari, P.: High-quality real time motion detection using PTZ cameras. In: IEEE International Conference on Video and Signal Based Surveillance, vol. 23 (November 2006)
29. Araki, S., Matsuoka, T., Takemura, H., Yokoya, N.: Real-time tracking of multiple moving objects in moving camera image sequences using robust statistics. In: Proceedings Fourteenth International Conference on Pattern Recognition, vol. 2, pp. 1433–1435 (August 1998)
30. Bin, L., Qiang, Z., Huanxia, L.: Research on background motion estimation and compensation in image sequences. In: 2011 International Conference on Mechatronic Science, Electric Engineering and Computer (MEC), pp. 1370–1373 (August 2011)
31. Cai, Q., Mitiche, A., Aggarwal, J.: Tracking human motion in an indoor environment. In: International Conference on Image Processing, vol. 1, pp. 215–218 (October 1995)
32. Lee, K.W., Ryu, S.W., Lee, S.J., Park, K.T.: Motion based object tracking with mobile camera. Electronics Letters 34(3), 256–258 (1998)
33. Nguyen, T.T., Jeon, J.W.: Real-Time Background Compensation for PTZ Cameras Using GPU Accelerated and Range-Limited Genetic Algorithm Search. In: Ho, Y.-S. (ed.) PSIVT 2011, Part I. LNCS, vol. 7087, pp. 85–96. Springer, Heidelberg (2011)
34. Pham, X.D., Cho, J.U., Jeon, J.W.: Background compensation using hough transformation. In: IEEE International Conference on Robotics and Automation (ICRA), pp. 2392–2397 (May 2008)
35. Rao, N., Di, H., Xu, G.: Joint correspondence and background modeling based on tree dynamic programming. In: 18th International Conference on Pattern Recognition (ICPR), vol. 2, pp. 425–428 (2006)
36. Shun, Z., Xiuqin, S., Liyin, X.: Global motion compensation for image sequences and motion object detection. In: 2010 International Conference on Computer Application and System Modeling (ICCASM), vol. 1, pp. 406–409 (October 2010)
37. Tordoff, B., Murray, D.: Reactive control of zoom while fixating using perspective and affine cameras. IEEE Trans. Pattern Anal. Mach. Intell. 26(1), 98–112 (2004)
38. Szeliski, R.: Image alignment and stitching: a tutorial. Found. Trends. Comput. Graph. Vis. 2(1), 1–104 (2006)
39. Jung, Y.K., Lee, K.W., Ho, Y.S.: Feature-Based Object Tracking with an Active Camera. In: Chen, Y.-C., Chang, L.-W., Hsu, C.-T. (eds.) PCM 2002. LNCS, vol. 2532, pp. 1137–1144. Springer, Heidelberg (2002)
40. Kim, J., Ye, G., Kim, D.: Moving object detection under free-moving camera. In: 17th IEEE International Conference on Image Processing (ICIP), pp. 4669–4672 (September 2010)
41. Varcheie, P., Bilodeau, G.A.: Adaptive fuzzy particle filter tracker for a PTZ camera in an IP surveillance system. IEEE Transactions on Instrumentation and Measurement 60(2), 354–371 (2011)

42. Guillot, C., Taron, M., Sayd, P., Pham, Q.C., Tilmant, C., Lavest, J.M.: Background Subtraction for PTZ Cameras Performing a Guard Tour and Application to Cameras with Very Low Frame Rate. In: Koch, R., Huang, F. (eds.) ACCV Workshops 2010, Part I. LNCS, vol. 6468, pp. 33–42. Springer, Heidelberg (2011)
43. 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: Proc. BMVC., vol. 34, pp. 1–10 (2010)
44. Murray, D., Basu, A.: Motion tracking with an active camera. IEEE Trans. Pattern Anal. Mach. Intell. 16(5), 449–459 (1994)
45. Xu, C., Liu, J., Kuipers, B.: Motion segmentation by learning homography matrices from motor signals. In: Proceedings of the 2011 Canadian Conference on Computer and Robot Vision, CRV 2011, pp. 316–323. IEEE Computer Society, Washington, DC (2011)