SSD and Histogram for Person Re-identification System

Abdullah Salem Baquhaizel, Safia Kholkhal, Belal Alshaqaqi, Mokhtar Keche

HAL Id: hal-01913904

https://inria.hal.science/hal-01913904

Submitted on 7 Nov 2018

SSD and Histogram for person re-identification system

Abdullah Salem BAQUHAIZEL, Safia Kholkhal, Belal ALSHAQAQI and
Mokhtar KECHE

Laboratoire Signaux et Images, département d'électronique, Université des sciences et
de la Technologie d'Oran Mohamed Boudiaf, USTO-MB. Oran, BP 1505, El
M'naouer. 31000, Algeria
`abdullah.baquhaizel@{univ-usto.dz ; gmail.com}`

Abstract. In this paper, we give the design and implementation of a
system for person re-identification in a camera network, based on the
appearance. This system seeks to construct an online database that con-
tains the history of every person that enters the field of view of the cam-
eras. This system is qualified to associate an identifier to each detected
person, which keeps this identifier in the same camera and in other cam-
eras even if he or she disappears and then appears again. Our system
comprises a moving objects detection step that is implemented using
the Mixture of Gaussians method and a proposed difference method,
to improve the detection results. It also comprises a tracking step that
is implemented using the sum of squared differences algorithm. The re-
identification stage is realized using two steps: the intersection of tracking
and detection for the temporal association, the histogram for compari-
son. The global system was tested on a real data set collected by three
cameras. The experimental results show that our approach gives very
satisfactory results.

Keywords: Person Re-identification, MoG, SSD,Histogram, Intersec-
tion.

1   Introduction

In recent years, video surveillance has grown more and more. This resulted in
an increase of cameras installed in different places (private or public), making
their exploitation and monitoring very difficult for human being. That is why
much research has been done to create intelligent vision systems that can help
the human being, in interpreting scenes and reacting with alarms in case of any
anomaly. Currently there are several types of video surveillance systems (access
control in sensitive locations, people recognition, control of traffic congestion,
...etc.).

In this paper we are interested in the problem of person re-identification in
a camera network. Re-identification in computer vision systems aims to follow a
person, associate an identifier to him, and store it in a database. If the person

leaves the scene then reappears in the field of view of any camera, it will be assigned the same identifier. In a crowded and uncontrolled environment observed by cameras from unknown distances, person re-identification relying upon conventional biometrics, such as face recognition, is neither feasible nor reliable, due to insufficiently constrained conditions and insufficient image details for extracting robust biometrics [17]. Instead, visual features based on the appearance of people, determined by their clothing and objects carried or associated with them, can be exploited more reliably for re-identification.

The remainder of this paper is organized as follows: Section 2 presents some related works from the literature. Section 3 descibes in details each block of the proposed system. The experimental results and and their discussion are presented in section 4. Finally, some conclusions are drawn in section 5.

2   RELATED WORKS

In the literature, the approaches of re-identification can be grouped in several classes, according to several criteria [12]:

1. The number of images per person:
   This class comprises two families. The first family is the family of mono-sample methods, where the signature of a person is extracted from a single image as in [15,3,24,6,16,1] . The second family is the family of multi-sample methods, where multiple images are used to calculate the signature of a person as in [5,14,21,11,4,7,19].
2. The type of representation:
   The first family in this class is the family of global approaches, where the whole information in the image is exploited for calculating the person's signature, as in [13,1,2]. The second family is that of local approaches, which represent the image by several feature vectors, each vector describes a region or a locally detected point, such as in [5,8,9].
3. The existence of a set of images mapped a priori:
   This class includes supervised approaches like in [3,2,14] and unsupervised approaches as in [23,10].

A very nice survey of people re-identification approaches is presented in [22]. They are therein grouped as a multidimensional taxonomy according to camera setting, sample set cardinality, signature, adoption of a body model, machine learning techniques and application scenario.

3   DESCRIPTION OF THE PROPOSED SYSTEM

In this section, we dscribe the different stages of the proposed system, for person re-identification in non-overlapping camera network. These stages are: person detection, their localization and verification, their tracking and their re-identification. The overall flowchart of the proposed system is shown in Figure 1.
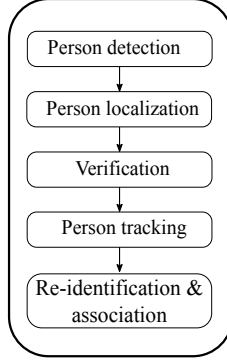
Fig. 1: The flowchart of the proposed system

3.1 Person detection

This initial stage is accomplished by combining the Mixture of Gaussians (MoG) method [20] and the difference method. The MoG is one of the most used and successful methods in surveillance systems, because it is adaptive, and can handle multimodal backgrounds.

In the difference method, we first take the difference between two successive images in grayscale $I_{g(t)}$ and $I_{g(t-1)}$, as in Eq. (1), and then we compare the resulting difference image I_{diff} to a threshold to detect pixels in movement.

$$I_{diff} = I_{g(t)} - I_{g(t-1)} \tag{1}$$

The hybrid of detections resulting from the MoG and difference methods is performed using the logical OR operation.

After detecting moving objects we fill the holes [18]. The holes of a binary image correspond to the set of its regional minima, which are not connected to the image border.

3.2 Person localization

The localization of the detected person is done using the labeling technique. This technique consists in separating the areas in the mask obtained from the detection step. We associate with each area an integer value (label) by using an 8-connected neighborhood, then we calculate some proprieties for each area, e.g. x and y coordinates, height, width and sum of foreground pixels.

3.3 Verification

To eliminate false detections, we propose a verification phase. To be validated each detected person has to verify the following three conditions:

- The ratio of width to height: this ratio has to lie between min and max thresholds.
- The surface of the rectangle containing the person (surface = height x width) has to lie between min and max thresholds. This is to eliminate very small and very big objects due to false detection.
- The ratio of the sum of foreground pixels to the surface also has to be limited.

3.4   Person Tracking

The person tracking process is done by template matching using the Sum of Squared Differences Algorithm (SSD). In digital image processing, the SSD is a measure of the similarity between image blocks. It is calculated by taking the square of the difference between each pixel in the original block X (a portion from the current frame) and the corresponding pixel in the Y block being used for comparison (Model from previous detection).

These differences are summed to create a simple metric of block similarity as in Eq. (2), zero means that the two blocks are identical. We sweep all the positions in the frame, then the block with the smallest metric is the tracked block.

The SSD value for two blocks X and Y calculated by:

$$SSD = \sum_{i=1}^{M} \sum_{j=1}^{N} (X(i,j) - Y(i,j))^2 \qquad (2)$$

For a given Y model, the most similar block X is the one that minimizes the SSD.

3.5   Re-identification and association

Following the stages of detection, localization, verification, and tracking, we have the stage of re-identification and online construction of database DB containing the history of each person that appeared in the view field of the cameras.Figure 2 presents a detailed flowchart of this stage.

This stage deals with the moving objects obtained from the detection and tracking stages, which are called 'found person'.

First, we calculate the intersection between the found persons resulting from the detection and tracking, the intersection $(A \cap B)$ of two rectangles A and B is the rectangle that contains all elements of A that also belong to B.

Then we test if the found persons resulted from detection only, tracking only or from both. If found person comes from intersection or tracking only, we update the database with the identifier of tracked person.

On the other hand, if that found person comes from detection only, then we calculate its histogram. An image histogram is a type of histogram that acts as a graphical representation of the tonal distribution in a digital image. It plots the number of pixels for each tonal value. The histogram of the found person is
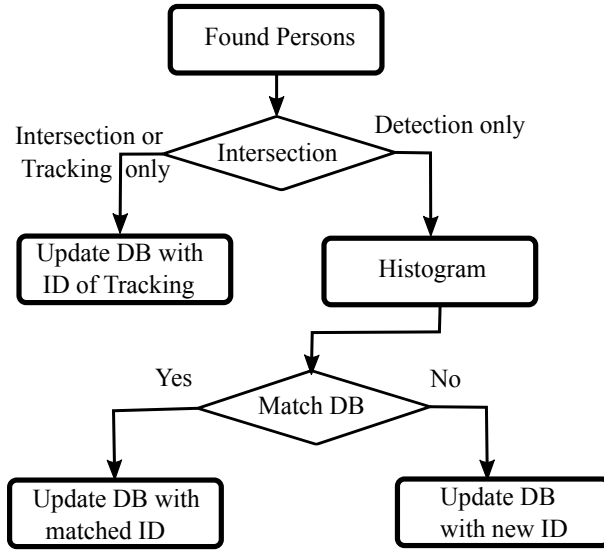
Fig. 2: Detailed flowchart of re-identification and online construction of database

compared to the histograms of identified persons stored in the database. If there is a match, then we update the database by associating that person with the matched identifier, otherwise, we consider this person as a new one and assign to it a new identifier that is added to the database.

4   EXPERIMENTAL RESULTS

In this section, we present the material and the database used, the experimental results, and their discussion.

4.1   System development environment

The material that was used for the development of our application is:

1. A laptop with:
   – Processor: Intel core i7 4702MQ CPU @ 2.20 GHz 2.20 GHz.
   – RAM memory: 8.00 Go.
   – Operating system: Windows 8.1, 64-bits
   – Hard Drive: 1 TB.
2. Digital video recorder DVR.
3. Camera with characteristics:
   – 1 /3 Sony HR CCD
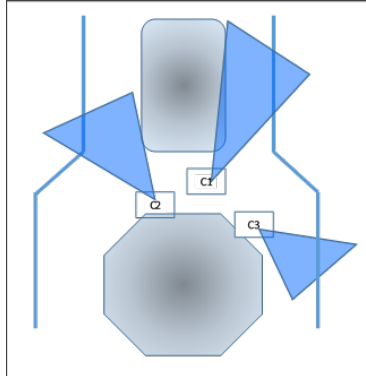   – 420 TV lines
   – 0.2 Lux

Fig. 3: Layout of the cameras

| (a) | (b) | (c) |

Fig. 4: Fields of view of the 3 cameras: (a) Camera 1, (b) Camera 2, (c) Camera 3

- Adjustable Focal between (3mm and 8 mm).

To test our system we build our own database, composed of sequences of images recorded on the third floor of the Department of Electronics at USTO university. Three cameras, set to a height of (2.30m) and with an angle of (-30), were used to take these images. Each sequence contains from one to three people who walk in the fields of view of the three cameras. The cameras were placed as shown in the layout presented in Figure 3.

To fulfil the condition of a non-overlapping camera network, the database was realized so that a person lies in the field of view of only one camera, at a given instant. Figure 4 shows the fields of view of the three cameras.

4.2   Experimental results, and discussion

In this section, we will present and discuss the results of each step of the proposed system. The Mixture of Gaussian gives us raw results of detection from each camera, after having defined suitable settings according to some criteria, like:
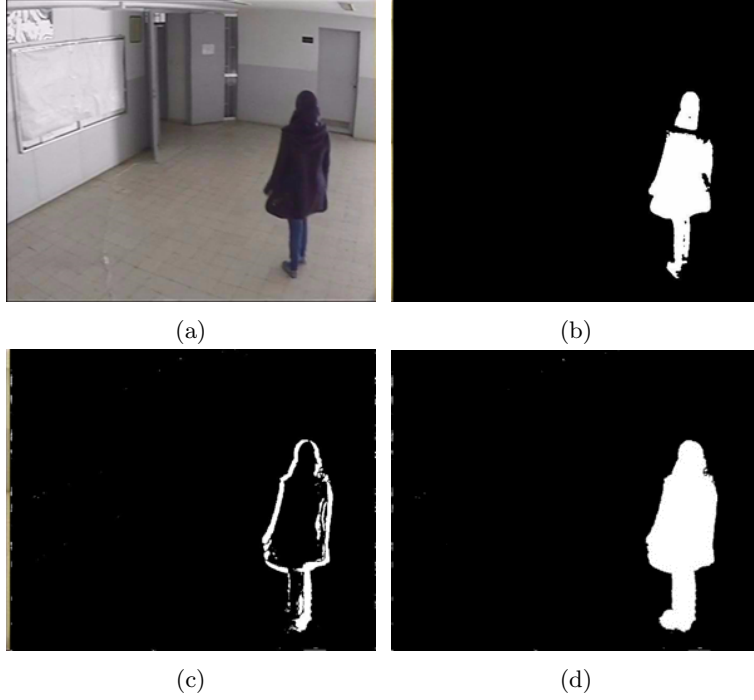
Fig. 5: Example of person detection. (a) Original image, (b) Results of detection by MoG, (c) Results of detection by difference, (d) Results of the holes filling of the OR between b and c

indoor or outdoor environment, people movement speed and lighting changes. Figure 5 (b) presents an example of these results. To improve these raw results, we combine them with the results of the difference method (Figure 5 (c)), which allows for the detection of the edges of moving objects, then we proceed to a holes filling of the resulting image to obtain better results as illustrated in Figure 5 (d).

In Figure 6, we present the localization and verification results. After the localization by the labeling technique, we apply the verification procedure to each person. In Figure 6 (a), only the persons that verify the validation conditions are kept (the person in green rectangle), the others in red are ignored. Figure 6 (b) presents the detection results.

The tracking step is run in parallel with the detection step and it is realized by the SSD. To acelerate its execution we decided to apply it only to a limited region of interest, instead of searching in the whole frame. This region is determined by the coordinates of the model to track. The obtained results of tracking are satisfactory. Figure 7 shows the tracking results, Figure 7 (a) is a detection in frame 159 and Figure 7 (b) is its tracking in frame 222.
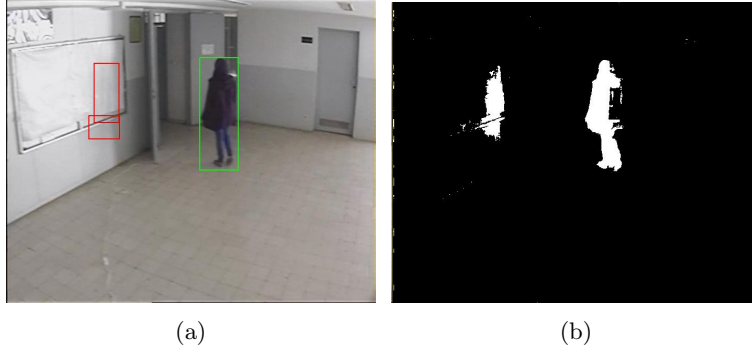
Fig. 6: Example of localization and verification, (a) Localization and verification results on the original image, (b) Results of detection
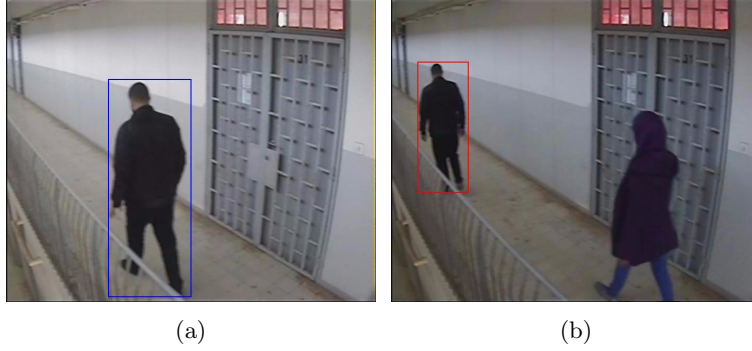

Fig. 7: Tracking results, (a) Detection in frame 159, (b) The tracking of detection of frame 159 in frame 222

The re-identification stage is realized using two techniques, the intersection of detection and tracking for the temporal association, and the histogram for comparison. In Figure 8, we present the multiplication of the detected person with its mask to extract the silhouette only. Then, we calculate histograms of Red, Green, Blue channels and grayscales as shown in Figure 9. We use the histogram of the silhouette to avoid the effect of the background. These different histograms are used for comparison with the models stored in the database. If there is a match, we associate the matched identifier to the actual person, otherwise, we consider that the actual person is new and add it to the database with a new identifier.

A sample of the constructed database is presented in Figure 10. This database contains the history of every person that enters the field of view of the cameras.

(a)   (b)   (c)

Fig. 8: The multiplication of detected person with its mask, (a) Detected person, (b) Mask of detected person and (c) Results of multiplication

5   CONCLUSION

In this paper, we presented the conception and implementation of a system for person re-identification in a camera network, based on the appearance. This system aims to build an online database that contains the history of every person captured by the cameras.

This system is able to assign an identifier to each detected person, that it keeps everywhere in the fields of view of the cameras and even if he or she disappears and then appears again.

Our system implements an improved detection technique that combines the Mixture of Gaussians method and the difference method. The SSD algorithm with an acceleration strategy is used for the tracking step, whereas the re-identification stage is realized using two techniques: the intersection for temporal association and the histogram comparison.

The global system was tested on a real data set collected by three cameras. The experimental results show that our approach leads to very satisfactory results with an opportunity for improvement in the re-identification stage, by using a local histograms instead of using the global one. Also as a future work, we plan to evaluate our method quantitatively and compare it with other methods.

References

1. Le An, Mehran Kafai, Songfan Yang, and Bir Bhanu. Reference-based person re-identification. In *2013 10th IEEE International Conference on Advanced Video and Signal Based Surveillance*, pages 244–249. IEEE, aug 2013.
2. Martin Bauml and Rainer Stiefelhagen. Evaluation of local features for person re-identification in image sequences. In *2011 8th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pages 291–296. IEEE, aug 2011.
3. Yinghao Cai, Kaiqi Huang, and Tieniu Tan. Human appearance matching across multiple non-overlapping cameras. In *19th International Conference on Pattern Recognition*, pages 1–4. IEEE, dec 2008.
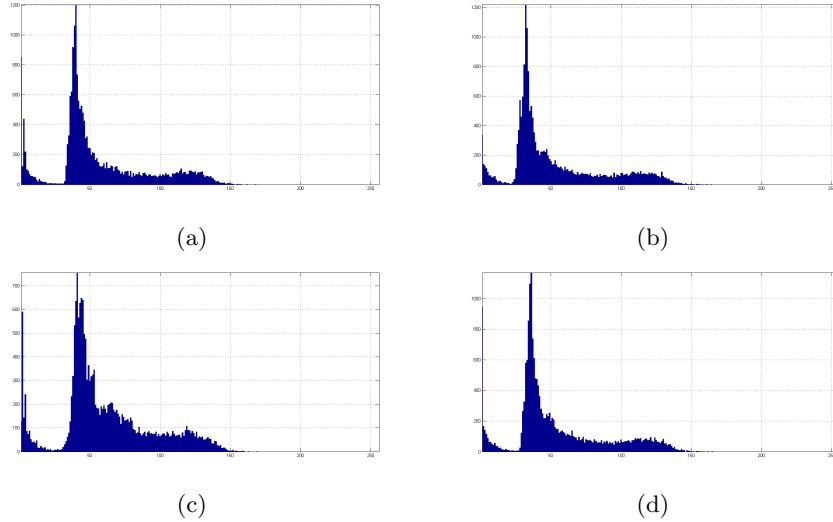
Fig. 9: Different histograms of the silhouette, (a), (b), (c) and (d) are the histograms of Red, Green, Blue channels and grayscales respectively

4. Dong Seon Cheng, Marco Cristani, Michele Stoppa, Loris Bazzani, and Vittorio Murino. Custom Pictorial Structures for Re-identification. In *Proceedings of the British Machine Vision Conference*, pages 68.1–68.11. British Machine Vision Association, 2011.

5. Icaro Oliveira de Oliveira and José Luiz De Souza Pio. People Reidentification in a Camera Network. In *2009 Eighth IEEE International Conference on Dependable, Autonomic and Secure Computing*, pages 461–466. IEEE, dec 2009.

6. Mert Dikmen, Emre Akbas, Thomas S. Huang, and Narendra Ahuja. Pedestrian Recognition with a Learned Metric. In Kimmel Ron, Klette Reinhard, and Sugimoto Akihiro, editors, *Computer Vision - ACCV 2010. Lecture Notes in Computer Science LNCS*, volume 6495, pages 501–512. Springer, Berlin, Heidelberg, 2011.

7. M. Farenzena, L. Bazzani, A. Perina, V. Murino, and M. Cristani. Person re-identification by symmetry-driven accumulation of local features. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 2360–2367. IEEE, jun 2010.

8. Niloofar Gheissari, T.B. Sebastian, and R. Hartley. Person Reidentification Using Spatiotemporal Appearance. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 2 (CVPR'06)*, volume 2, pages 1528–1535. IEEE, 2006.

9. Omar Hamdoun, Fabien Moutarde, Bogdan Stanciulescu, and Bruno Steux. Person re-identification in multi-camera system by signature based on interest point descriptors collected on short video sequences. In *2008 Second ACM/IEEE International Conference on Distributed Smart Cameras*, pages 1–6. IEEE, sep 2008.

10. Martin Hirzer, Peter M. Roth, and Horst Bischof. Person Re-identification by Efficient Impostor-Based Metric Learning. In *2012 IEEE Ninth International Conference on Advanced Video and Signal-Based Surveillance*, pages 203–208. IEEE, sep 2012.
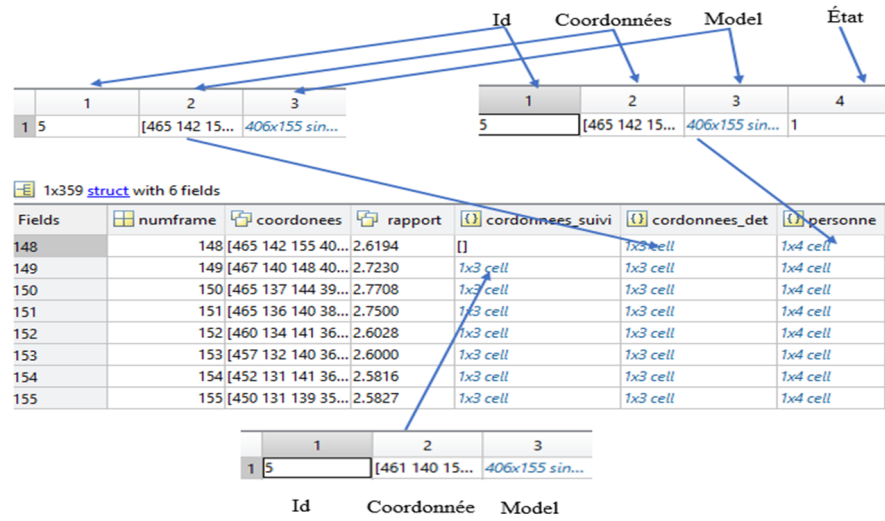
Fig. 10: A sample of the constructed database

11. Chung-Hsien Huang, Yi-Ta Wu, and Ming-Yu Shih. Unsupervised Pedestrian Re-identification for Loitering Detection. In *Advances in Image and Video Technology*, pages 771–783. 2009.

12. Mohamed Ibn Khedher. *Ré-identification de personnes à partir des séquences vidéo*. Dissertations, Institut National des Télécommunications, Paris, 2014.

13. Yoshihisa Ijiri and Shihong Lao. Human Re-Identification Through Distance Metric Learning Based on Jensen-Shannon Kernel. In *Visapp: - International Conference on Computer Vision Theory and Applications*, pages 603–612, 2012.

14. Kai Jungling and Michael Arens. View-invariant person re-identification with an Implicit Shape Model. In *2011 8th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pages 197–202. IEEE, aug 2011.

15. U. Park, A.K. Jain, I. Kitahara, K. Kogure, and N. Hagita. ViSE: Visual Search Engine Using Multiple Networked Cameras. In *18th International Conference on Pattern Recognition (ICPR'06)*, pages 1204–1207. IEEE, 2006.

16. William Robson Schwartz and Larry S. Davis. Learning Discriminative Appearance-Based Models Using Partial Least Squares. In *2009 XXII Brazilian Symposium on Computer Graphics and Image Processing*, pages 322–329. IEEE, oct 2009.

17. Chen Change Loy Shaogang Gong, Marco Cristani and TimothyM. Hospedales. The Re-identification Challenge. In Shaogang Gong, Marco Cristani, Shuicheng Yan, and Chen Change Loy, editors, *Person Re-Identification*, chapter 1. Springer London, London, 2014.

18. Pierre Soille. *Morphological Image Analysis*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2004.

19. Malik Souded. *People Detection, Tracking and Re-identification Through a Video Camera Network*. PhD thesis, Signal and Image processing, Institut National de Recherche en Informatique et en Automatique, UNIVERSITE DE NICE - SOPHIA ANTIPOLIS ECOLE, Nice, 2013.

20. Chris Stauffer and W.E.L. Grimson. Adaptive background mixture models for real-time tracking. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, pages 246–252. IEEE Comput. Soc, 1999.

21. Dung-Nghi Truong Cong, Catherine Achard, and Louahdi Khoudour. People re-identification by classification of silhouettes based on sparse representation. In *2010 2nd International Conference on Image Processing Theory, Tools and Applications*, pages 60–65. IEEE, jul 2010.

22. Roberto Vezzani, Davide Baltieri, and Rita Cucchiara. People reidentification in surveillance and forensics: A Survey. *ACM Computing Surveys*, 46(2):1–37, nov 2013.

23. Taiqing Wang, Shaogang Gong, Xiatian Zhu, and Shengjin Wang. Person Re-identification by Video Ranking. In David Fleet, Tomas Pajdla, Bernt Schiele, and Tinne Tuytelaars, editors, *Computer Vision – ECCV 2014. Lecture Notes in Computer Science LNCS*, volume 8692, pages 688–703. Springer, Cham, 2014.

24. Xiaogang Wang, Gianfranco Doretto, Thomas Sebastian, Jens Rittscher, and Peter Tu. Shape and Appearance Context Modeling. In *2007 IEEE 11th International Conference on Computer Vision*, pages 1–8. IEEE, 2007.