Adv Syst Sci Appl 2018; **3**:1–16
Published online at http://ijassa.ipu.ru.

Person re-identification based on SAD and Histogram in a non-overlapping camera network

Abdullah Salem BAQUHAIZEL *, Belal ALSHAQAQI, Meriem BOUMEHED, 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

Abstract: In this paper, we present the conception and implementation of a system for person re-identification in a camera network, based on the appearance. This system aims to associate an identifier to each detected person, which keeps this identifier in the same camera and in other cameras even if he or she disappears and then reappears again. Our system comprises an improved moving objects detection step that is implemented by combining the Mixture of Gaussians method (MoG) and a proposed difference method, to improve the detection results. Then localization and verification step to eliminate false detection. It also comprises a tracking step that is implemented using the sum of absolute differences algorithm (SAD), with an acceleration strategy to reduce calculation complexity. The re-identification stage is realized using three steps: the tracking for the temporal association, the histogram and the new smart exploitation of the interpolation technique for comparison. Our proposed system build an online database that contains the history of every person that enters the field of view of the cameras, it does not require any previously collected training data, it integrates both the detection and the tracking so it can be used in real applications, and another advantage is the simplicity of the used techniques in term of implementation and calculation rapidity, without affecting the quality of results. The global system was tested on a real data set collected by three cameras. The experimental results show that our approach gives very courageous results.

Keywords: person re-identification, MoG, SAD, histogram, interpolation

1. INTRODUCTION

In recent times, video surveillance has grown more and more. This resulted in an increase of cameras installed in different places (public or private), making their exploitation and monitoring extremely difficult for a human being. Consequently, 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 (people recognition, access control in sensitive locations, control of traffic congestion, ...etc.).

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

*Corresponding author: abdullah.baquhaizel@{univ-usto.dz ; gmail.com}

image details for extracting robust biometrics [31]. Alternatively, 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 person re-identification. The main contributions of this paper are:

- Our proposed system is able to build a fully automated online database that contains the history of every person captured by the cameras, based on their appearance.
- The proposed system does not require any previously collected training data, so no specific reference for a person is stored a priori.
- Compared to most of the other state-of-the-art approaches, our proposed system is a self-contained system that is ready to use in real applications since it integrates both the detection and the tracking phases.
- The simplicity of used techniques. Furthermore, the improvement in each block, object detection by combining the MoG method and the difference method, the tracking and the re-identification.

The rest of this paper is organized as follows: Section 2 presents some related works from the literature. In section 3, we describe the proposed system and give the details for its blocks. The experimental results 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 person re-identification can be grouped into several classes, according to several criteria [20]:

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 [4, 9, 12, 26, 30, 39]. For example, in[4], authors project the gallery and the probe data into the Regularized Canonical Correlation Analysis (RCCA) subspace then the reference descriptors (RDs) of the gallery and probe are constructed by measuring the similarity between them and the reference data. In [12], a metric learning framework is used to obtain a robust metric for large margin nearest neighbor classification with rejection.
   The second family is the family of multi-sample methods, where multiple images are used to calculate the signature of a person as in [10, 11, 13, 19, 22, 29, 34, 36]. For example, authors in [10] propose Custom Pictorial Structure (CPS) for re-identification. In [22] they use the Implicit Shape Model (ISM) and SIFT features for the person re-identification. In [29] spatiotemporal person features are extracted using multi-frame twin-channel descriptor based on a Gabor filter, then Mahalanobis distance metric learning algorithm is used for matching. In [36] the discriminating nature of the sparse representation is exploited in order to perform people re-identification task.
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 [4, 5, 7, 17, 21, 27]. For example, in [21], the Jensen-Shannon kernel is proposed to learn nonlinear distance metrics. In [5] non-articulated 3D body models are exploited to spatially map appearance descriptors (color and gradient histograms) into the vertices of a regularly sampled 3D body surface. In[17], the appearance by a set of region covariance descriptors is modeled, then a discriminative model is learned using boosting for feature selection. In [27], authors extract KDES (Kernel DEScriptor) from human ROI for person classification.
   The second family is that of local approaches, which represent the image by several feature vectors, each vector describes a region or an interest point locally detected, such

as in [11, 14, 15, 42]. For example, in [11] interest points are detected with Fast-Hessian detector, then they are described in Color and SURF description. The approach in [15] uses matching of signatures based on interest points descriptors detected by a variant inspired from SURF called Camellia key-points. In [42], authors use the face to video retrieval, they analyzed four local features for describing face images: Harris corner operators, SIFT descriptors, SURF descriptors, and eigenfaces.

3. The existence of a set of images mapped a priori:
   This class includes supervised approaches like in [7, 9, 22, 25, 28]. For example, two structured learning based approaches are proposed in [25], they explore the adaptive effects of multiple low-level visual features with an optimal ensemble of their metrics. In [28], a data driven approach is proposed for learning color patterns, they model color feature generation by jointly learning a linear transformation and a dictionary to encode pixel values.
   In the other hand, we have the unsupervised approaches as in [18, 23, 38, 40]. For example, a video representation, called Spatio-Temporal Pyramid Sequence (STPS) is developed by [23] to encode space-time information, and they formulate a novel Time Shift Dynamic Time Warping (TS-DTW) model and its Multi-Dimensional extension named MDTS-DTW for selective matching between sequences. In [40] a dynamic graph matching (DGM) method is proposed; a graph for samples in each camera is constructed, and then graph matching scheme is introduced for cross-camera labeling association.

A very nice survey of people re-identification approaches is presented in [37]. They are therein grouped as a multidimensional taxonomy according to camera setting, sample set cardinality, adoption of a body model, signature, machine learning techniques, and application scenario.

3. DESCRIPTION OF THE PROPOSED SYSTEM

We present the conception and implementation of a system for person re-identification in a camera network, based on the appearance. In this section, we describe the different blocks of the proposed system and how all these blocks are connected. These blocks are: the detection of moving objects, their localization and verification, their tracking and their re-identification. The detailed flowchart of the proposed system is shown in Figure 3.1.

In the beginning, we initialize the number of found objects to zero. Then for each frame, we start detecting moving objects, next we perform their localization, then a verification and confirmation stage is applied to improve the results by eliminating the false detection.

In parallel to the detection and for each frame also, we have the tracking process of any found objects from the previous frame. After that, the obtained results from the tracking are fused with those of the detection and verification stage, by using the technique of intersecting and overlapping.

Then comes the feature extraction step where the Histogram for each found object is calculated. For the re-identification, we use the tracking result for the temporal association, the histogram and the trajectory interpolation for the comparison.

The re-identification and association block tries to associate the found objects with objects already seen or add them as new objects if they appear for the first time, thus, building an online database which contains the history of every person that appears in front of the camera. In the coming subsections, we will discuss in detail the techniques and the tools used in each block.

3.1. Detection of moving objects

This initial phase is performed by combining the Mixture of Gaussians (MoG) method [35] and a proposed method we called the difference method.

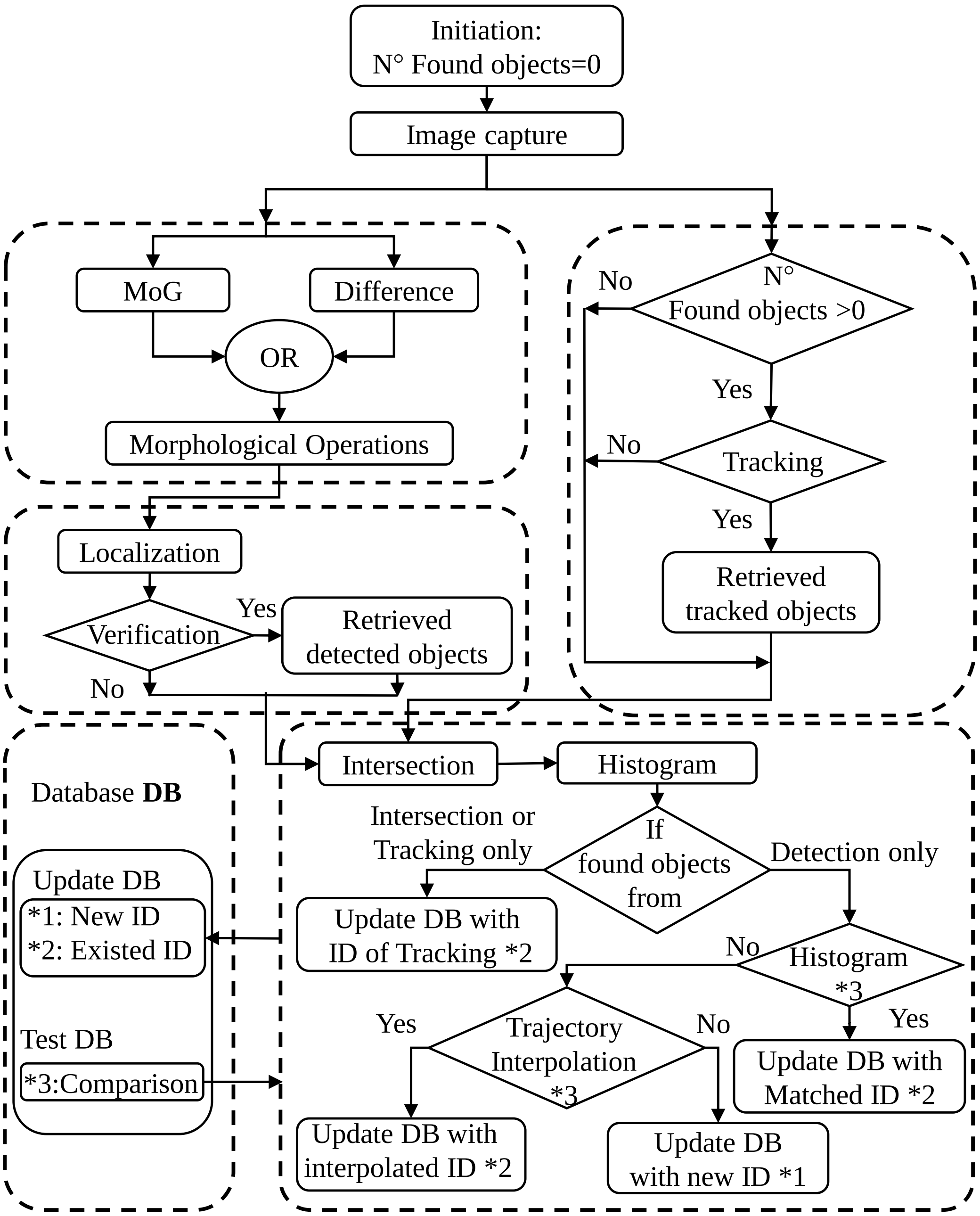


Fig. 3.1. The Detailed flowchart of the proposed system

The MoG is one of the most used and successful methods in surveillance systems, because it is adaptive, and can handle multimodal backgrounds [8]. It is the most common method to build a background[32]. Furthermore, it is robust to slow lighting changes, periodic motions from a cluttered background, slow moving objects, long term scene changes, and camera noises [41]. In this method, the recent history of each pixel, $\{X_1, ..., X_t\}$, is modeled by a mixture of K Gaussian distributions. The probability of observing the current pixel value is given by:

$$P(X_t) = \sum_{i=1}^{K} \omega_{i,t} * \eta(X_t, \mu_{i,t}, \Sigma_{i,t}) \tag{3.1}$$

Where $\omega_{i,t}$ is the portion of the data accounted by the *i*th Gaussian in the mixture at time *t* and $\eta(X_t, \mu_{i,t}, \Sigma_{i,t})$ is a Gaussian probability density function with mean value $\mu_{i,t}$ and covariance matrix $\Sigma_{i,t}$ of the *i*th Gaussian in the mixture at time *t* , where η is given by:

$$\eta(X_t, \mu, \Sigma) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(X_t-\mu_t)^T \Sigma^{-1} (X_t-\mu_t)} \tag{3.2}$$

K is the maximum number of Gaussian components, depending on the available memory. Currently, this parameter ranges from 3 to 5. At each pixel X_t in a new frame, the following update equations for the K distributions are performed:

$$\omega_{K,t} = (1-\beta)\,\omega_{K,t} + \beta \tag{3.3}$$

β is the learning rate of adaptation. This value ranges from 0.0 to 1.0, determines the importance of the previous observation. The first **B** distributions are chosen as the background model, where:

$$\mathbf{B}_t = argmin_b \left(\sum_{K=1}^{b} \omega_k > T \right) \tag{3.4}$$

The value of T determines the minimum portion of the data that should be accounted by the background. When higher values for T are chosen, a multi-model background model is formed which can handle repetitive background motion [35].

In the other hand, we have the difference method, we first take the difference between two successive images in gray-scale $I_{g(t)}$ and $I_{g(t-1)}$, as in Eq. (3.5), and then we compare the resulting difference image I_{diff} to a threshold to detect pixels in movement.

$$I_{diff} = I_{g(t)} - I_{g(t-1)} \tag{3.5}$$

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

After that, we apply morphological operations as erosion, dilatation, and fill the holes. The holes of a binary image corresponding to the set of its regional minima, which are not connected to the image border [33].

3.2. Localization and Verification

To localize the detected objects, we use the labeling technique [16]. It 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 propose a verification phase to eliminate false detection. We first calculate some proprieties for each localized area, e.g. x and y coordinates, height, width and sum of foreground pixels. Then to be validated, each object has to verify the following three conditions:

- The height to width ratio has to lie between min and max thresholds.
- The surface of the rectangle containing the object (surface = height x width) has to lie between min and max thresholds. This is to reject 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 bounded.

3.3. Tracking

The tracking process is done by template matching using the Sum of Absolute Differences Algorithm (SAD), like in our previous proposed driver drowsiness detection systems[2, 3]. This algorithm is widely used for image compressing and object tracking in real-time application [1], it can give high localization rate when the image is with high illumination variation [24].

In digital image processing, the SAD is a measure of the similarity between image blocks. It is calculated by taking the absolute 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 the previous detection).

These differences are summed to create a simple metric of block similarity as in Eq. (3.6), 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 SAD value for two blocks X and Y is calculated by:

$$SAD = \sum_{i=1}^{M} \sum_{j=1}^{N} |X(i,j) - Y(i,j)| \tag{3.6}$$

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

3.4. Re-identification

After 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. The lower part of Figure 3.1 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 objects'.

First, we proposed to perform the intersection between the found objects resulting from the detection and tracking, then we calculate the percentage of the intersection if it exceeded a predefined threshold we merge the objects from detection and tracking into one object, otherwise, they stay separated objects. 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.

Now, we test each found object, if it resulted from the detection only, the tracking only or from both (intersection). If the found object comes from the intersection or tracking only, we update the database with the identifier of tracked object.

On the other hand, if that found object 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 object is compared to the histograms of identified objects stored in the database. If there is a match, then we update the database by associating that object with the matched identifier, otherwise, we make a trajectory interpolation. The interpolation is a method of constructing new data points within the range of a discrete set of known data points. If there is an interpolation, we update the database with the interpolated identifier, otherwise, we consider this object 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.

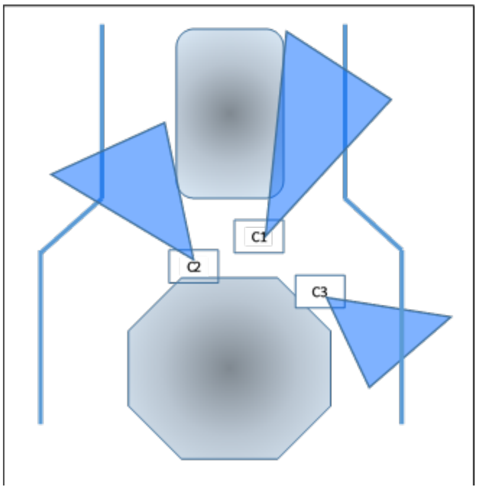


Fig. 4.2. Layout of the cameras

(a) (b) (c)

Fig. 4.3. Fields of view of the 3 cameras: (a) Camera 1, (b) Camera 2, (c) Camera 3

4.1. System development environment

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

1. A laptop with:
   - Processor: Intel ®core^{TM} 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
   - Adjustable Focal between (3mm and 8 mm).

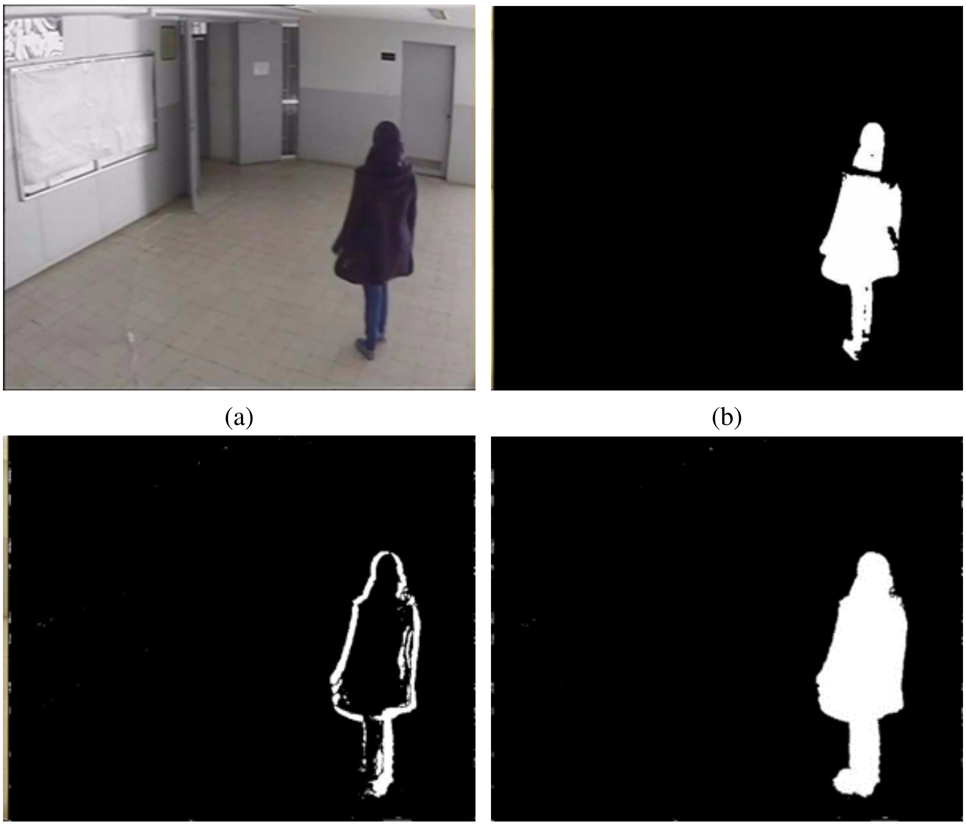
To test our system we build our own database [6], composed of sequences of images recorded on the third floor of the Department of Electronics at USTO university. Three cameras, set to a height of (2.30m) and with an angle of ($-30°$), were used to take these images. Each sequence contains from one to 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 4.2.

To fulfill 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 4.3 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.

4.2.1. Results of the detection The Mixture of Gaussian gives us raw results of detection from each camera, after having defined suitable settings according to some criteria, like: indoor or outdoor environment, people movement speed and lighting changes. Figure 4.4 (b) presents an example of these results. To improve these raw results, we combine them with the results of the difference method (Figure 4.4 (c)), which allows for the detection of the edges of moving objects, then we apply morphological operations as erosion, dilatation and proceed to a holes filling of the resulting image to minimize false detection and obtain better results as illustrated in Figure 4.4 (d).



(a) (b)

(c) (d)

Fig. 4.4. Example of moving objects 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

4.2.2. Results of the verification To evaluate the verification phase, we compare the results before and after the verification. We consider the true detection from the detection before verification as positive and the false detection as negative. Table 4.1 presents for videos from Camera 1, Camera 2 and Camera 3, the obtained evaluation measures, namely the TP (True Positive), the FN(False Negative), the FP (False Positive), and the TN (True Negative). To understand these measures we present Figure 4.5, let us assume that we have four objects;

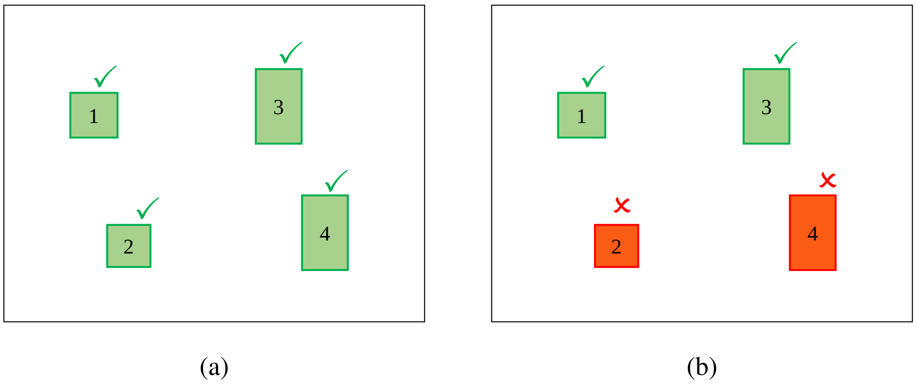


Fig. 4.5. Before and after verification evaluation, (a) Detection results (before the verification), (b) Results after the verification

Table 4.1. Evaluation results of the proposed verification technique

	Camera 1	Camera 2	Camera 3
TP	411	399	341
FN	23	8	7
FP	2	2	3
TN	19	30	73

objects 1 and 2 are not people, however, objects 3 and 4 are people. Figure 4.5 (a) is the results before the verification, so objects 1 and 2 are false detection, however objects 3 and 4 are true detection. In Figure4.5 (b), we have the results after the verification, objects 1 and 3 are kept, but objects 2 and 4 are removed. Now we can assign each measure of evaluation to the corresponding object:

- Object 1 is a False Positive (FP).
- Object 2 is a True Negative (TN).
- Object 3 is a True Positive (TP).
- Object 4 is a False Negative (FN).

From these measures, we can define the following rates:

- TP_TPFN % = TP/ (TP + FN): % of true person kept.
- FN_TPFN % = FN/ (TP + FN): % of true person removed.
- FP_TNFP % = FP/ (TN + FP): % of false person kept.
- TN_TNFP % = TN/ (TN + FP): % of false person removed.

The obtained evaluation rates are given in Table. 4.2, for videos from Camera 1, Camera 2 and Camera 3. From this table one can observe, that the proposed verification technique performs well, since it manages to remove 90.48%, 93.75% and 96.05% of false detection in Camera 1, Camera 2 and Camera 3, respectively, while removing only 5.30%, 1.97% and 2.01% of true detection, in Camera 1, Camera 2 and Camera 3, respectively.

Table 4.2. Evaluation rates of the proposed verification technique

	Camera 1	**Camera 2**	**Camera 3**
TP_TPFN%	94.70	98.03	97,99
FN_TPFN%	5.30	1.97	2.01
FP_TNFP%	9.52	6.25	3.95
TN_TNFP%	90.48	93.75	96.05

In Figure 4.6, we give the localization and verification results. After the localization by the labeling technique, we apply the verification procedure to each object. In Figure 4.6 (a), only the objects that pass the conditions of verification are kept (the person in green rectangle), the others in red are ignored. Figure 4.6 (b) shows the results of detection .

(a) (b)

Fig. 4.6. Example of localization and verification, (a) Localization and verification results on the original image, (b) Results of detection

4.2.3. Results of the tracking The tracking step is run in parallel with the detection step and it is realized by the SAD. To accelerate 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 4.7 shows the tracking results, Figure 4.7 (a) is a detection in frame 137 and Figure 4.7 (b) is its tracking in frame 194, it means that we still maintain this object even after about 57 frames.

(a) (b)

Fig. 4.7. Tracking results, (a) Detection in frame 137, (b) The tracking of detection of frame 137 in frame 194

4.2.4. Results of the re-identification The re-identification stage is realized using three techniques, the tracking for the temporal association, the histogram and the interpolation technique for comparison. In Figure 4.8, we present the multiplication of the detected object with its mask to extract the silhouette only. Then, we calculate histograms of Red, Green, Blue channels and gray-scales as shown in Figure 4.9. 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 object, otherwise, we use the trajectory interpolation technique, as shown in Figure 4.10 (a), to determine if this object is in the trajectory of an existing object. If so, we associate the actual object to this object, otherwise, we consider that the actual object is new and add it to the database with a new identifier. From Figure 4.10 (b), we can deduce that in the period between the frames 220 and 270, mentioned by two verticals lines in red, the two person appear together and Figure 4.10 (c) is a sample from that period. Note that the appearance of person 1 starts from the frame 120, and that of person 2 from the frame 180. The miss in the appearance of the person are due to the elimination of certain items by the verification stage.

The performance of our re-identification system is evaluated using two criteria:

1. Number of created IDs,
2. Number of appearances and associations of each ID,

- Evaluation based on the number of created IDs: we can evaluate our proposed re-identification system by comparing the number of assigned IDs that created by the system to the number of real IDs. The ideal is when these two numbers are equal.

In Table. 4.3, these numbers are compared, using videos from the different cameras. The number of real IDs is obtained by the ground truth of used dataset, and their appearance is how many time they appear. However, the number of created IDs is obtained by the system, and their appearance is how many time they appear. It can be observed that the number of created IDs is larger than the number of real IDs; this is due to the false detection. It can also be observed that by eliminating most of the false detection using the verification stage, we reduce the number of created IDs to a level close to the number of real IDs.

Table 4.3. Comparison of the number of created IDs (Nb_C_IDs) to the number of real IDs (Nb_R_IDs), and their number of appearance (Nb_A)

	Real		**Created**	
	Nb_R_IDs	**Nb_A**	**Nb_C_IDs**	**Nb_A**
Camera 1	3	434	7	515
Camera 2	3	407	5	541
Camera 3	3	348	8	494

Table 4.4. Number of appearances (Nb_A) and Person-IDs associations of each ID(P) in camera 1, 2 and 3

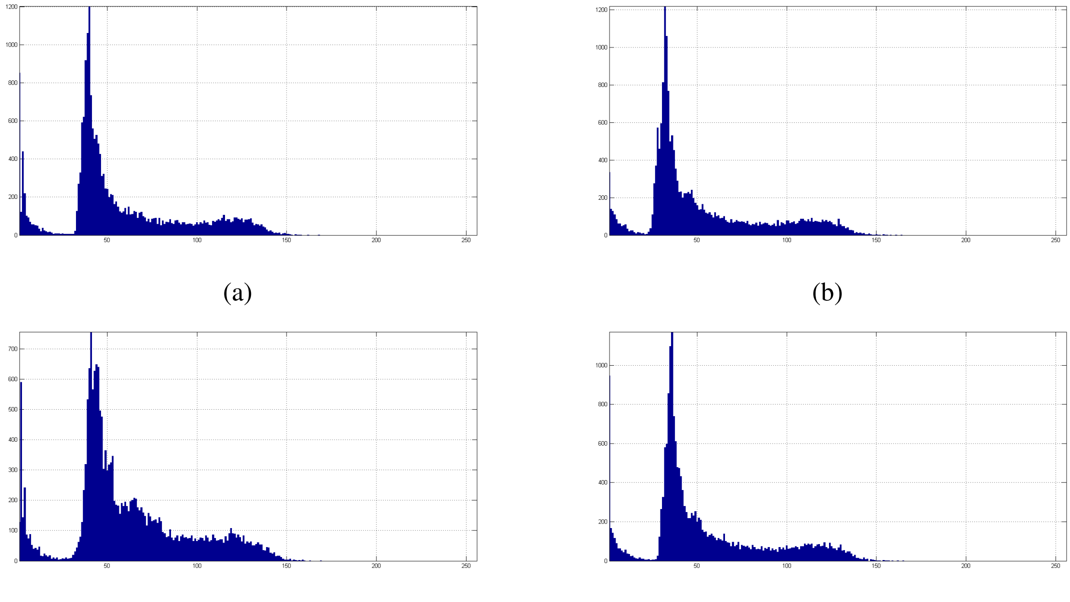
ID	**Camera 1**		**Camera 2**		**Camera 3**	
	Nb_A	**P**	**Nb_A**	**P**	**Nb_A**	**P**
1	175	P1	181	P1	98	P1
2	36	P2	27	-	131	P2
3	31	-	172	P2	29	-
4	41	P3	57	-	40	P1
5	117	P1	104	P3	30	-
6	27	-	-	-	33	-
7	88	P3	-	-	102	P3
8	-	-	-	-	31	P3

- Evaluation based on the number of appearances and associations of each ID: Another way to evaluate our proposed system is the number of appearances of each ID,(see Table. 4.4). From this table, we can determine the number of IDs for each camera and the Number of Appearances of each ID. For example, we can see that for Camera 1, seven

IDs were created, and ID 5 was associated 117 times. In addition, from this table, we can determine the IDs associated with each person. For example, it can be seen from Table. 4.4, that with Camera 3 the IDs 1 and 4 were associated with the person 1, named P1.

(a) (b) (c)

Fig. 4.8. The multiplication of detected object with its mask, (a) Detected object, (b) Mask of detected object and (c) Results of multiplication



(a) (b) (c) (d)

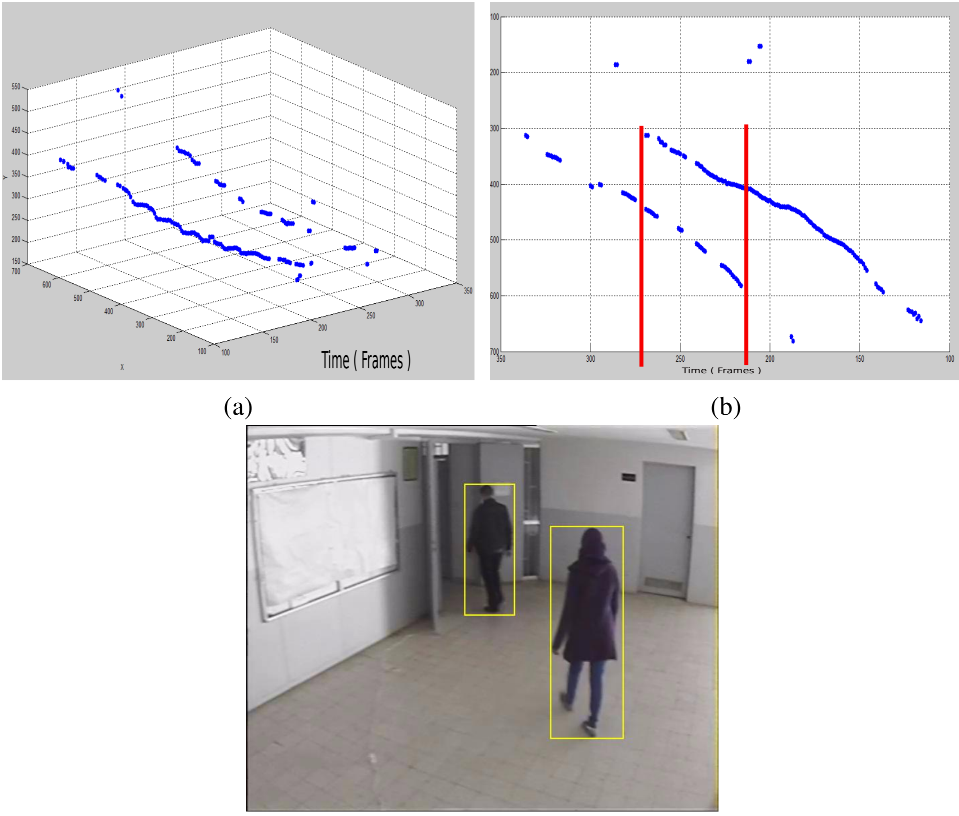
Fig. 4.9. Different histograms of the silhouette, (a), (b), (c) and (d) are the histograms of Red, Green, Blue channels and gray-scales respectively

5. CONCLUSION

In this paper, we proposed the conception and implementation of a system for person re-identification in a camera network, based on the appearance.

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 Gaussian method with the difference method. The SAD algorithm with an acceleration



(a) (b)

(c)

Fig. 4.10. Trajectory Interpolation, (a) The trajectories of the two objects through time, (b) The trajectories of the two objects along the Y-axis (c) The two objects that appear in this sequence

strategy is used for the tracking step, whereas the re-identification stage is realized using three techniques: the tracking for the temporal association, the histogram and the trajectory interpolation technique for comparison.

Our proposed system does not require any previously collected training data to build an online database that contains the history of every person that enters the field of view of the cameras. It can be used in real applications because it integrates both the detection and the tracking. Another advantage is the simplicity of the used techniques in term of implementation and calculation rapidity, without affecting the quality of results.

A real data set is collected by three cameras to test the global system. The experimental results show that our approach leads to very courageous results with an opportunity for improvement in the re-identification stage, by using 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] Alshaqaqi, B., Baquhaizel, A. S., Ouis, M., & Boumehed, M. (2016). Real-Time Implementation Of Face Detection And Tracking. *Algerian Journal of Research and Technology*, *1*(1), 25–30.

[2] Alshaqaqi, B. [Belal], Baquhaizel, A. S., Amine Ouis, M. E., Boumehed, M., Ouamri, A., & Keche, M. (2013). Driver drowsiness detection system. In *2013 8th International*

Workshop on Systems, Signal Processing and their Applications (WoSSPA) (pp. 151–155). doi:10.1109/WoSSPA.2013.6602353

[3] Alshaqaqi, B. [Belal], Baquhaizel, A. S., Ouis, M. E. A., Boumehed, M., Ouamri, A., & Keche, M. (2013). Vision based system for driver drowsiness detection. In *2013 11th International Symposium on Programming and Systems (ISPS)* (pp. 103–108). doi:10.1109/ISPS.2013.6581501

[4] An, L., Kafai, M., Yang, S., & Bhanu, B. (2013). Reference-based person re-identification. In *2013 10th ieee international conference on advanced video and signal based surveillance* (pp. 244–249). doi:10.1109/AVSS.2013.6636647

[5] Baltieri, D., Vezzani, R., & Cucchiara, R. (2015). Mapping Appearance Descriptors on 3D Body Models for People Re-identification. *International Journal of Computer Vision*, *111*(3), 345–364. doi:10.1007/s11263-014-0747-z

[6] Baquhaizel, A. S., Kholkhal, S., Alshaqaqi, B., & Keche, M. (2018). SSD and Histogram for Person Re-identification System. In *Ifip advances in information and communication technology* (Vol. 522, pp. 585–596). doi:10.1007/978-3-319-89743-1_50

[7] Bauml, M., & Stiefelhagen, R. (2011). 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)* (pp. 291–296). doi:10.1109/AVSS.2011.6027339

[8] Boumehed, M., Alshaqaqi, B., Ouamri, A., & Keche, M. (2013). Moving objects localization by local regions based level set: Application on urban traffic. *Journal of Mathematical Imaging and Vision*, *46*(2), 258–274. doi:10.1007/s10851-012-0400-9

[9] Cai, Y., Huang, K., & Tan, T. (2008). Human appearance matching across multiple non-overlapping cameras. In *19th international conference on pattern recognition* (pp. 1–4). doi:10.1109/ICPR.2008.4761704

[10] Cheng, D. S., Cristani, M., Stoppa, M., Bazzani, L., & Murino, V. (2011). Custom Pictorial Structures for Re-identification. In *Proceedings of the British Machine Vision Conference* (pp. 68.1–68.11). doi:10.5244/C.25.68

[11] de Oliveira, I. O., & Pio, J. L. D. S. (2009). People Reidentification in a Camera Network. In *2009 Eighth IEEE International Conference on Dependable, Autonomic and Secure Computing* (pp. 461–466). doi:10.1109/DASC.2009.33

[12] Dikmen, M., Akbas, E., Huang, T. S., & Ahuja, N. (2011). Pedestrian Recognition with a Learned Metric. In K. Ron, K. Reinhard, & S. Akihiro (Eds.), *Computer vision - accv 2010. lecture notes in computer science lncs* (Vol. 6495, *PART 4*, pp. 501–512). doi:10.1007/978-3-642-19282-1_40

[13] Farenzena, M., Bazzani, L., Perina, A., Murino, V., & Cristani, M. (2010). Person re-identification by symmetry-driven accumulation of local features. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 2360–2367). doi:10.1109/CVPR.2010.5539926

[14] Gheissari, N., Sebastian, T., & Hartley, R. (2006). Person Reidentification Using Spatiotemporal Appearance. In *2006 ieee computer society conference on computer vision and pattern recognition - volume 2 (cvpr'06)* (Vol. 2, pp. 1528–1535). doi:10.1109/CVPR.2006.223

[15] Hamdoun, O., Moutarde, F., Stanciulescu, B., & Steux, B. (2008). 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* (pp. 1–6). doi:10.1109/ICDSC.2008.4635689

[16] Haralick, R. M., & Shapiro, L. G. (1992). Computer and Robot Vision, Vol 1. (pp. 28–48). Addison-Wesley.

[17] Hirzer, M., Beleznai, C., Roth, P. M., & Bischof, H. (2011). Person Re-identification by Descriptive and Discriminative Classification. In H. Kalviainen, J. Parkkinen, &

A. Kaarna (Eds.), *Image analysis: 17th scandinavian conference, scia 2011, ystad, sweden, may 2011. proceedings* (Vol. 3540, *March 2016*, pp. 91–102). Lecture Notes in Computer Science. doi:10.1007/978-3-642-21227-7_9

[18] Hirzer, M., Roth, P. M., & Bischof, H. (2012). Person Re-identification by Efficient Impostor-Based Metric Learning. In *2012 IEEE Ninth International Conference on Advanced Video and Signal-Based Surveillance* (pp. 203–208). doi:10.1109/AVSS.2012.55

[19] Huang, C.-H., Wu, Y.-T., & Shih, M.-Y. (2009). Unsupervised Pedestrian Re-identification for Loitering Detection. In *Advances in image and video technology* (pp. 771–783). doi:10.1007/978-3-540-92957-4_67

[20] Ibn Khedher, M. (2014). *Ré-identification de personnes à partir des séquences vidéo* (Dissertations, Institut National des Télécommunications, Paris).

[21] Ijiri, Y., & Lao, S. (2012). Human Re-Identification Through Distance Metric Learning Based on Jensen-Shannon Kernel. In *Visapp: - international conference on computer vision theory and applications* (pp. 603–612). Retrieved from http://www.murase.nuie.nagoya-u.ac.jp/$%5Csim$murase/pdf/951-pdf.pdf

[22] Jungling, K., & Arens, M. (2011). View-invariant person re-identification with an Implicit Shape Model. In *2011 8th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)* (pp. 197–202). doi:10.1109/AVSS.2011.6027319

[23] Ma, X., Zhu, X., Gong, S., Xie, X., Hu, J., Lam, K.-M., & Zhong, Y. (2017). Person re-identification by unsupervised video matching. *Pattern Recognition*, *65*, 197–210. doi:10.1016/j.patcog.2016.11.018. arXiv: 1611.08512

[24] Nourain Dawoud, N., Belhaouari Samir, B., & Janier, J. (2011). Fast Template Matching Method based Optimized Sum of Absolute Difference Algorithm for Face Localization. *International Journal of Computer Applications*, *18*(8), 30–34. doi:10.5120/2302-2912

[25] Paisitkriangkrai, S., Wu, L., Shen, C., & van den Hengel, A. (2017). Structured learning of metric ensembles with application to person re-identification. *Computer Vision and Image Understanding*, *156*, 51–65. doi:10.1016/j.cviu.2016.10.015. arXiv: 1511.08531

[26] Park, U., Jain, A., Kitahara, I., Kogure, K., & Hagita, N. (2006). ViSE: Visual Search Engine Using Multiple Networked Cameras. In *18th International Conference on Pattern Recognition (ICPR'06)* (pp. 1204–1207). doi:10.1109/ICPR.2006.1176

[27] Pham, T. T. T., Le, T. L., Vu, H., Dao, T. K., & Nguyen, V. T. (2017). Fully-automated person re-identification in multi-camera surveillance system with a robust kernel descriptor and effective shadow removal method. *Image and Vision Computing*, *59*, 44–62. doi:10.1016/j.imavis.2016.10.010

[28] Rama Varior, R., Wang, G., Lu, J., & Liu, T. (2016). Learning Invariant Color Features for Person Reidentification. *IEEE Transactions on Image Processing*, *25*(7), 3395–3410. doi:10.1109/TIP.2016.2531280. arXiv: 1410.1035

[29] Sathish, P. K., & Balaji, S. (2017). Multi-frame twin-channel descriptor for person re-identification in real-time surveillance videos. *International Journal of Multimedia Information Retrieval*, *6*(4), 289–294. doi:10.1007/s13735-017-0136-9

[30] Schwartz, W. R., & Davis, L. S. (2009). Learning Discriminative Appearance-Based Models Using Partial Least Squares. In *2009 xxii brazilian symposium on computer graphics and image processing* (pp. 322–329). doi:10.1109/SIBGRAPI.2009.42

[31] Shaogang Gong, Marco Cristani, C. C. L., & Hospedales, T. (2014). The Re-identification Challenge. In S. Gong, M. Cristani, S. Yan, & C. C. Loy (Eds.), *Person re-identification* (Chap. 1). doi:10.1007/978-1-4471-6296-4_1

[32] Shi, Y., Cheng, S., Quan, S., Chen, J., & Chen, D. (2011). Moving objects detection by Gaussian Mixture Model: A comparative analysis. In *2011 International Conference on Electrical and Control Engineering* (20090143110004, pp. 1121–1124). doi:10.1109/ICECENG.2011.6058008

[33] Soille, P. (2004). Morphological Image Analysis. In *Morphological image analysis principles and applications* (p. 208). doi:10.1007/978-3-662-05088-0. arXiv: 1011.1669

[34] Souded, M. (2013). *People Detection, Tracking and Re-identification Through a Video Camera Network* (Doctoral dissertation, Signal and Image processing, Institut National de Recherche en Informatique et en Automatique, UNIVERSITE DE NICE - SOPHIA ANTIPOLIS ECOLE, Nice).

[35] Stauffer, C., & Grimson, W. (1999). Adaptive background mixture models for real-time tracking. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (Vol. 2, *100*, pp. 246–252). doi:10.1109/CVPR.1999.784637. arXiv: CVPR.1999.784637 [10.1109]

[36] Truong Cong, D.-N., Achard, C., & Khoudour, L. (2010). People re-identification by classification of silhouettes based on sparse representation. In *2010 2nd International Conference on Image Processing Theory, Tools and Applications* (pp. 60–65). doi:10.1109/IPTA.2010.5586809

[37] Vezzani, R., Baltieri, D., & Cucchiara, R. (2013). People reidentification in surveillance and forensics: A Survey. *ACM Computing Surveys*, *46*(2), 1–37. doi:10.1145/2543581.2543596

[38] Wang, T., Gong, S., Zhu, X., & Wang, S. (2014). Person Re-identification by Video Ranking. In D. Fleet, T. Pajdla, B. Schiele, & T. Tuytelaars (Eds.), *Computer vision – eccv 2014. lecture notes in computer science lncs* (Vol. 8692, *PART 4*, pp. 688–703). doi:10.1007/978-3-319-10593-2_45

[39] Wang, X., Doretto, G., Sebastian, T., Rittscher, J., & Tu, P. (2007). Shape and Appearance Context Modeling. In *2007 ieee 11th international conference on computer vision* (pp. 1–8). doi:10.1109/ICCV.2007.4409019

[40] Ye, M., Ma, A. J., Zheng, L., Li, J., & Yuen, P. C. (2017). Dynamic Label Graph Matching for Unsupervised Video Re-identification. *Proceedings of the IEEE International Conference on Computer Vision*, *2017-Octob*, 5152–5160. doi:10.1109/ICCV.2017.550. arXiv: 1709.09297

[41] Ying-Li Tian, Lu, M., & Hampapur, A. (2005). Robust and Efficient Foreground Analysis for Real-Time Video Surveillance. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)* (Vol. 1, pp. 1182–1187). doi:10.1109/CVPR.2005.304

[42] Zhou, J., & Huang, L. (2017). A Comparative Study of Local Features in Face-based Video Retrieval. *Journal of Computing Science and Engineering*, *11*(1), 24–31. doi:10.5626/JCSE.2017.11.1.24

