

RESEARCH

Open Access



# Soft-biometrics evaluation for people re-identification in uncontrolled multi-camera environments

Daniela Moctezuma<sup>1\*</sup>, Cristina Conde<sup>2</sup>, Isaac Martín De Diego<sup>2</sup> and Enrique Cabello<sup>2</sup>

## Abstract

A novel method for person identification based on soft-biometrics and oriented to work in real video surveillance environments is proposed in this paper. Thus, an evaluation of relevance's level of several appearance features is carried out with this purpose. First, a bag-of-soft-biometric features related to color, texture, local features, and geometry are extracted from individuals. The relevance of each feature has been deeply analyzed through different proposed methods. Features are ranked and weighted according to their relevance value. Later, each method is evaluated under two different scenarios: mono-camera and multi-camera surveillance images. In order to test the system in a realistic way, it has been evaluated over standard databases in the surveillance community: PETS 2006, PETS 2009, CAVIAR, SAIVT-SoftBio, and CAVIAR4REID. Moreover, a new database was acquired at Adolfo Suarez Madrid-Barajas international airport. This database was acquired under regular conditions and infrastructure of the Barajas airport, no additional camera or special settings were installed for this purpose. An analysis of relevance for each feature acquired in these two scenarios is presented. The results obtained demonstrate the promising potential of the soft-biometric approach. Finally, an optimal system configuration according to each scenario is obtained.

**Keywords:** Soft-biometrics; Video-surveillance; Person re-identification; Feature relevance

## 1 Introduction

For many areas of current society, re-identification of human beings based on their biometric or soft-biometric features is becoming an important task. Person identification is turning to an essential part of the security needs of diverse infrastructures such as airports, shopping centers, government buildings, and train stations, among others. Person re-identification is about identifying a specific individual across non-overlapping distributed cameras at different times and locations. This task is challenging due to the dramatic changes in an individual's appearance, in terms of lighting, occlusion, pose, zoom, and camera quality, among others [1]. The importance of intelligent video surveillance systems, its application in the security sector, and its increasing social interest have turned person identification an important research topic in the

last few years. Generally, video surveillance of wide and critical areas requires a system of multiple cameras to monitor people constantly. Camera surveillance is more than a set of monitors connected to some cameras, in fact, it can actually be seen as a powerful technology for security control. One of the main problems in a multiple camera system is that people's appearance changes as observed from one camera to another [2]. Usually, a person's visual identification can be either based on biometric (face, iris, fingerprint) or soft-biometric features (appearance, height, clothes). Gait is a feature that can be measured at a distance, this can be considered as suitable for surveillance environments. Nevertheless, gait has many limitations that make it unsuitable for wide areas such as current surveillance environments. The issues of pose variations, occlusions, and lighting are still a challenge for gait recognition. Moreover, in most works, the gait is considered as a biometric [3], due to the above and the fact

\*Correspondence: dmocteo@gmail.com

<sup>1</sup> Centro de Investigación en Geografía y Geomática "Ing. Jorge L. Tamayo", A.C., Contoy 137 Esq. Chemax, Col. Lomas de Padierna, Delegación Tlalpan, 14240 Mexico D.F., Mexico

Full list of author information is available at the end of the article

that on this paper only soft-biometrics are considered, gait feature is not used in this evaluation. On the one hand, biometric features are based on the unique characteristics of each person; biometric identifiers cannot be shared or misplaced and they intrinsically represent the person's identity, thus, offering a high performance singling out individuals as long as camera conditions are good (lighting, quality, location, and high resolution). On the other hand, the soft-biometric features are not intrusive during the acquisition process and can be applied directly in most of the existing camera systems. For these reasons, they can be considered as a promising approach. This research uses a soft-biometric appearance model focused on a watch-list approach, where the objective is to recognize and track people who are in a list (like a terrorist or criminals list). That is, a large number of cameras that are installed in many locations are used, while a very large number of people will pass by these surveillance cameras, only a set of individuals must be recognized. That is, the system must reject every subject unless the subject is in the watch-list [4]. This work proposes an identification system that can identify  $N$  pre-specified individuals, while rejecting everybody else. This approach would be particularly useful for intelligent video-surveillance, where the  $N$  individuals would be the suspects in the watch-list. Moreover, for person identification purposes, a bag-of-soft-biometric features is composed of different categories of features: color, texture, local features, and geometry. The importance of each feature is analyzed by studying the contribution of each feature to the person's aspect. In this paper, four different methods to measure the relevance of each feature are considered. Features are sorted into a ranking according to its relevance. Then, features are weighted based on their ranking position. In order to generate an exhaustive evaluation of soft-biometric features, the experiments were conducted using standard databases well-known in the surveillance community (PETS 2006, PETS 2009, CAVIAR, SAIVT-SoftBio, and CAVIAR4REID). Furthermore, a new completely realistic database, Multi-camera Barajas International Airport (MUBA), was acquired at Barajas airport in Madrid. Notice that the proposed methods do not require calibration or special configuration, they work under real conditions of each surveillance environment. Results show a promising response of the system under different conditions and scenarios. Moreover, the results show the huge potential of soft-biometric features.

The paper is organized as follows: a brief overview of related works is shown in Section 2, databases description and its analysis are presented in Section 3, and the bag-of-soft-biometric features and methods to measure feature relevance are presented in Section 4. Section 5 shows the experimental results and analysis, and finally, Section 6 concludes.

## 2 Overview

Since people tracking and identification is a very important research topic, the research community is focused on detecting, tracking and identifying people, and interpreting their behavior. In the literature, there are many works related to intelligent video surveillance systems. Most of these works employ biometric or soft-biometric features to identify people. The soft-biometrics are related to people's appearance, such as color, clothing, and height. The work proposed in [2], an appearance model represented by a hierarchical structure where each node maintains a color Gaussian mixture model (GMM), is proposed. The identification task is performed with a Bayesian decision approach. Results show that this appearance model is robust to rotation and scale variations. Nevertheless, only four different people were used for experiments. In [5], a multi-task distance metric, in order to achieve people re-identification in a camera network, is proposed. This method designs multiple Mahalanobis distance metrics. Moreover, a multi-task maximally collapsing metric is presented. The experiments were performed in the GRID and the VIPeR databases. The results are approximately 80 % of recognition rate.

In [6], representative clothing-colors are extracted by applying an octree-based color quantization technique to the clothing region. The height is extracted from the geometrical information of the images, and the Euclidean distance is used for comparison purposes. Results are limited to know the exact camera position, and the images are acquired in a controlled environment. The work [7] proposes a person identification method that uses three soft-biometric features (clothing, complexion, and height). These soft-biometrics are employed so, a robot recognizes people in real time in a social environment. Here, the soft-biometrics were evaluated as a part of an experiment of person identification task, this experiment was carried out at Fleet Week, NY. In the experiments, the robot must identify an specific person between groups of three people. The experimental results show that the clothing is the most relevant soft-biometric, reaches an 85 % of identification rate. Another method to person re-identification by embedding middle level clothes attributes is presented in [8]. Here, the person re-identification problem is studied specifically in three main points: a part-based human appearance representation approach, a person re-identification method by embedding into the discriminant classifier by a latent SVM framework, and a person re-identification benchmark. This benchmark includes a large-scale database and an evaluation based on open-set experimental settings. The experiments were carried out with their own database (NUS-Canteen Database), and the results show a verification rate of 85 %.

In [1], a system based on soft-biometric features (such as gender, backpack, jeans, and short hair) is presented.

These soft-biometrics are detected, and attribute-based distances are calculated between pairs of images by using a regression model. Experiments are conducted on the ViPER database and the obtained results show that the effectiveness of this method depends on the accuracy of soft-biometric extraction. The lowest classification result was obtained with a short hair feature, reaching 53 % accuracy, and the highest classification result was obtained with a “carrying” feature, reaching 75 % accuracy. A framework that automatically registers soft-biometric features every time that users log-in is proposed in [9]. These soft-biometric features (color of user’s clothing and facial skin) are fused with the conventional authentication with password and face biometric. Experimental results show the effectiveness of the proposed method for continuous user authentication. In [10], a multi-camera tracking through spatial and appearance features approach is proposed. An approximated object position is estimated for the spatial feature, and color is used for the appearance features. The similarity calculation is based on the Earth Mover’s Distance (EMD). Here the experiments show good results.

In [11], several visual low-level features with supervised learning methods are used. Average RGB value, color structure, and histogram on HMMD color space are some of the considered features. The results here, show a reliable performance; nevertheless, the experiments were done with their own database containing few people. An approach to classify people, groups of people, and luggage in the halls of an airport is proposed in [12]. For this, two kinds of features are used: foreground density features and features related to real-size of objects obtained by applying a homographic model. A classification scheme based on  $k$  nearest neighbors ( $k$ -nn) algorithm and voting system is proposed and the obtained results are good. In [13], a multi-modal method for human identification is presented. Here, the gait and other type of movements are considered and results above 78 % of accuracy rate are achieved. Another method for human identification that uses gait feature is proposed in [14]. Here, first the contour in each gait image is extracted, followed by encoding each of the gait contour images in the same gait sequence with a multichannel mapping function. Experiments on gait databases are carried out and results over 75 % of identification rate are obtained.

Finally, an experimental study of the benefits of soft-biometrics to improve person recognition in scenarios at distance is presented in [15]. The available soft-biometric information in scenarios or varying distance between camera and subject is analyzed here. The experiments are conducted in the Southampton multi-biometric tunnel database, and the results show that the use of soft-biometrics is able to improve the performance of recognition. In most of the analyzed works, only a few features, cameras, and individuals are considered, as can

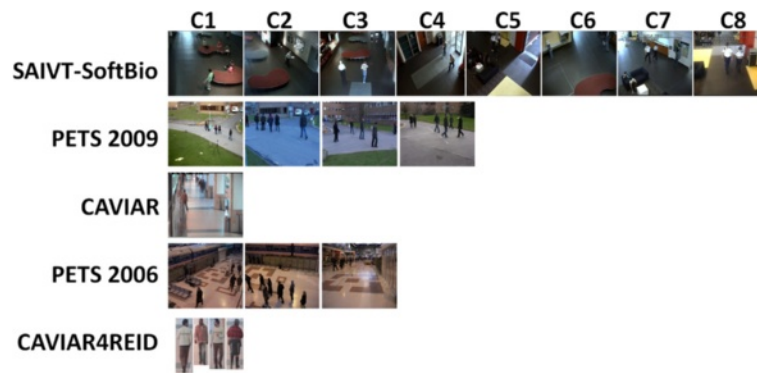
be seen. In contrast, the work presented here evaluates a complete set of different feature categories and each feature’s relevance is calculated. Furthermore, for a deep analysis, six different databases were employed for the evaluation.

### 3 Databases description

Currently, intelligent video surveillance community has made proposals to standardize the performance evaluation of computer vision-based surveillance through proposing several standard databases acquired in realistic video surveillance scenarios. In this work, five different public databases have been considered. Moreover, a more realistic database acquired at Barajas International Airport in Madrid, is presented. Figures 1 and 3 show sample images of standard databases and MUBA database, respectively. In a mono-camera scenario, the databases considered were the following: CAVIAR, PETS 2006, PETS 2009, and MUBA. Since that the multi-camera scenario is more complex, a total of five databases were used: PETS 2006, PETS 2009, CAVIAR4REID, SAIVT-SoftBio, and MUBA. Notice that only the CAVIAR database was considered using a unique camera. As a consequence, CAVIAR was only used in the mono-camera scenario.

#### 3.1 Standard databases

In this work, the well known databases PETS 2006 [16], PETS 2009 [17], CAVIAR [18], SAIVT-SoftBio [19] and CAVIAR4REID [20] have been used. These databases are recognized as standard in the intelligent video surveillance research. These databases were obtained in public spaces: a mall, a train station, a park, etc. That is, both indoor and outdoor environments are considered. The PETS 2006 database has been designed for activity recognition and surveillance of public spaces at Victoria station in London. The sequences used from this database were as follows: S2-T3-C and S4-T5-A-C (for both, camera 1, camera 3, and camera 4). S2-T3-C and S4-T5-A-C are the names given to these sequences by the authors of PETS 2006. The PETS 2009 database comprises multi-sensor sequences containing crowd scenarios with increasing scene complexity in outdoor environment. Here, the following sequences were selected: S2-L1-Timen (camera 1, camera 5, camera 7, and camera 8). S2-L1-Time12-34 is the name given to this sequence by the authors of PETS 2009. The CAVIAR database consists of images showing a corridor at a shopping mall in Portugal. For the experiments, the OneStopMoveEnter1cor and EnterExitCrossingPaths1cor sequences were used in this work. OneStopMoveEnter1cor and EnterExitCrossingPaths1cor are the names given to these sequences by the authors of CAVIAR. The SAIVT-SoftBio database consists of several sequences of subjects walking in a building environment along eight cameras from various angles and



**Fig. 1** Images from public databases, PETS2009 with four cameras, PETS2006 with three cameras, CAVIAR with one camera, SAIVT-SoftBio with eight cameras, and CAVIAR4REID with two cameras (C# represents the camera number)

different illumination conditions. The CAVIAR4REID is a database to evaluate person re-identification algorithms. CAVIAR4REID has been extracted from CAVIAR, one of the most famous database for person tracking and detection. For the generation of CAVIAR4REID, a total of 72 individuals were manually extracted, for this work all individuals were considered. Figure 1 shows some sample images from each standard database. As a general disadvantage from these standard databases, in the mono-camera scenario specially, only a few suspects have been selected, this is because cameras are overlapping and there are not many different suspects to select.

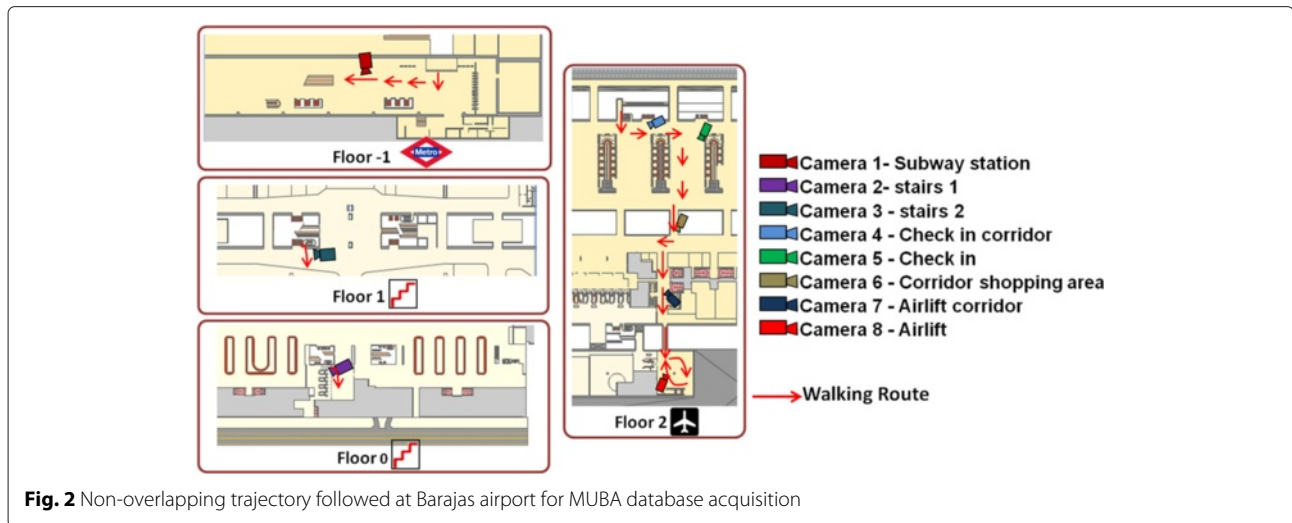
### 3.2 Multi-camera Barajas Airport (MUBA) database

In order to evaluate the system in a complex real-world scenario, several image sequences were acquired at Barajas International Airport (terminal 4) in Madrid. These images were obtained in collaboration with Civil Guard and A.E.N.A company. This database was acquired using several cameras across the airport. In this database acquisition, the airport's infrastructure and airport's cameras were used, specific cameras were not added nor changes in the usual operation of the security staff were required. This way, the video surveillance camera system of Barajas airport was exploited for this work. Eight different non-overlapping cameras were used in this database, these cameras show spaces like a subway station, a check-in room, and airport corridors. The acquisition was carried out in a conventional way where security guards handled the cameras in the control center. A select group of people walked passing through the eight cameras to generate this database. This to ensure that at least a group of people covered all the cameras where we were allowed to record. Nevertheless, not only were considered images from this selected group. Later, images from eight cameras (all recorded at the same time) were provided to us to generate this database. That is, images from each camera were captured simultaneously as in real operation. This

database is conformed by 91,482 images proceeding from eight cameras, from which 141 people were extracted (62 for mono-camera and 79 from multi-camera) in total. Around 60 images were extracted per individual, that is, from 91,482 images, a subset of 41,640 were used, which represent more of 45 % of the whole database. All images of individuals were collected through their complete track across all cameras where they were seen (sometimes two, three, four until eight cameras). For more details on the use of this database, please contact the authors. Figure 2 shows a scheme of places within the airport used in this database acquisition. A large area of this airport has been considered, as it can be seen in this image. The kind of acquired images are very complex, containing real and absolutely non-controlled situations, with sequences of few people and crowded sequences. Figure 3 shows sample images from each different camera of MUBA database.

### 3.3 Complexity of the databases

The difficulty level of each database could be analyzed in two ways. First, a quantitative analysis of quality of images. Second, a qualitative analysis is made by security staff members according to subjective parameters that could affect to the identification task. In order to analyze the quality of images from each camera in each database, peak signal-to-noise ratio (also known as PSNR), has been calculated (for more details see [21] and [22]). The peak signal-to-noise ratio is a video quality metric. This metric analyzes the ratio between the maximum possible power of a signal (in this case, an image) and the power of corrupting noise that affects the fidelity of its representation. Notice that when PSNR is higher, the image has higher quality. For subjective and qualitative analysis of complexity of images, two measures have been observed: a background segmentation and people detection complexity. First, people detection difficulty, defined as the complexity to detect people in several situations like people moving, people temporarily stationary, the

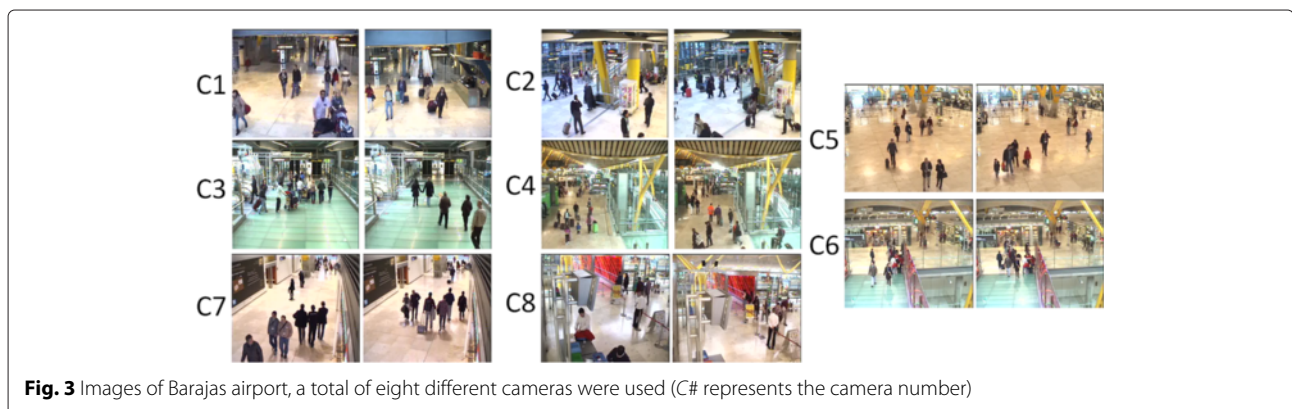


number of persons, partial occlusions, and pose variations are considered. Second, background segmentation difficulty, defined as the complexity to extract the background from the scene, as well as noise presence, lighting changes, shadows, and objects belonging to the background, was proposed. Table 1 shows the quality and complexity values for each database. It can be seen that PETS 2006 is the “easiest” database because it has a high quality of images and the detection and background subtraction is easier to carry out with PETS 2006 than with the other databases considered. CAVIAR and CAVIAR4REID have similar complexity due to the fact that CAVIAR4REID was generated from CAVIAR, that is, CAVIAR4REID is comprised by the cropped images from the people shown in CAVIAR images. On the other hand, MUBA database is the worst quality database. The SAIVT-SoftBio has high complexity too, but MUBA database has a lower quality value (PSNR). This is comprehensible due to the realistic and non-controlled (with poor quality cameras) environment of MUBA’s acquisition. Therefore, MUBA is the most complex database using both quantitative and qualitative measures. As another measure of complexity, it would be

interesting to analyze the biometric/soft-biometric extraction complexity, but in practical terms, with the combination of detection and background segmentation quality measures, a biometric/soft-biometric extraction complexity can be determined. To summarize; Table 2 shows how many people and images were used from each scenario and each database.

#### 4 System description

The description of the proposed system is presented in this section. Figure 4 shows a general scheme of the proposed system. A set of soft-biometric features (called bag-of-soft-biometric features) are extracted from the person’s images on the first step. That is, a total of 23 features are extracted from an input image creating the bag-of-soft-biometric features. The bag-of-soft-biometrics is evaluated in order to calculate the relevance level of each feature on the second step during the training phase. Here, the ranking features are generated according to several proposed methods, and then each feature is weighted according to its ranking position. The 30 % of images from each database was considered on the training phase



**Table 1** Difficulty classification of the images from each database

Database	Quality (PSNR)	Detection	Background segmentation	Average difficult
PETS 2006	13.31	Low	Medium	Easy
PETS 2009	12.27	High	Medium	Hard
CAVIAR	9.16	Medium	High	Medium
SAIVT-SoftBio	12.10	High	High	Hard
CAVIAR4REID	9.10	Medium	High	Medium
MUBA	7.86	High	High	Very hard

and the remaining 70 % was used throughout the testing phase. This analysis could demonstrate that the relevance of each feature is related to the scenario, for example, mono- and multi-camera and indoor and outdoor scenarios. Four methods to measure this relevance and two ways to weighting are proposed for this evaluation. The best ranking and weighting methods are selected during the course of the third step. These best methods are used to acquire the final results in the fourth step. That is, once the results are obtained from the four proposed methods, the method that reached the best results in each database is selected for the testing phase. Finally, the final results per database are obtained based on the Euclidean distance classification.

#### 4.1 Bag of soft-biometric features

The first step of the proposed method is to extract the people presented in each image using the HoGG algorithm presented in [23]. Then, a set of soft-biometric features is extracted from each person, constituting the so-called bag-of-soft-biometrics. These features have

**Table 2** Summarized description from all databases considered

Database	People	Images
Mono-camera		
PETS 2006	7	470
PETS 2009	7	460
CAVIAR	7	480
MUBA	62	3720
Total	83	5130 (1539 for training and 3591 for test)
Multi-camera		
PETS 2006	17	6120
PETS 2009	8	2880
MUBA	79	37,920
SAIVT-SoftBio	80	29,336
CAVIAR4REID	72	720
Total	256	76,976 (23,093 for training and 53,883 for test)

been selected considering the restrictions that bad acquisition conditions impose on resolution and lighting in a real surveillance multi-camera environment. In most of the analyzed works, only a few features are considered, whereas our work presents a complete set of different feature categories. In order to obtain a high identification rate, several feature categories have been considered. These categories are the RGB color space, gray-scale statistics, geometry, gray-scale histogram, HSV color space, co-occurrence matrix, and LBP (see Table 4). Since results in soft-biometrics improve with higher amount of features, a total of 23 different features have been used in this work.

Given that the appearance of a person is dominated by its clothes, color and texture features are suitable for people description. Features of color and texture are two low-level features widely used for image classification, indexing, and retrieval [24]. Color is usually represented as a histogram, which is a first-order statistical measure that captures global color distribution from an image. In this work, two color spaces have been used, RGB and HSV. The RGB is the most popular and natural color model, because it can compose any color adequately. The RGB color model is defined by three channels (red, green, and blue). Features based on RGB color space used in this work are: channel mean, the mean of the three channels, the standard deviation of the three channels, and the brightness. Standard deviation and mean are two well-known statistic measures. Brightness is the result of sum of three channels (red, green, and blue). The HSV color model divides the luminance component (intensity value) of a pixel color from its chrominance components (hue and saturation). Hue component represents pure color, and saturation gives a measure of the degree in which a pure color is diluted by white light [24]. The considered features for the HSV color model are the mean and the standard deviation. Mean and the standard deviation in gray-scale images have been considered as features too. Since a histogram can give an approximate idea of the distribution of gray levels in the image and it provides different useful features like mean value, dispersion, and contrast, several global features based on gray-scale histogram have been used: mean, standard deviation, entropy, dispersion, energy, and kurtosis. The last global extracted feature is the eccentricity, which represents the relationship between height and width of people. One of the main drawbacks of the histogram-based features is that spatial distribution is not considered. In order to avoid this, local spatial information from a gray-scale co-occurrence matrix has been used. The gray-scale co-occurrence matrix (GCM) stores the number of pixel neighborhood in an image that has a particular gray-scale combination [24]. Considering the image spatial domain, the configuration of the GCM used is distance  $d = 1$

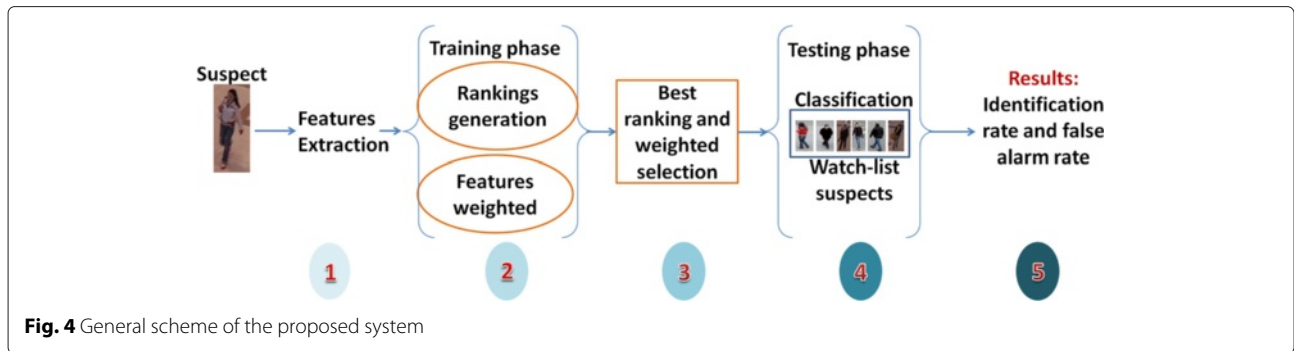


Fig. 4 General scheme of the proposed system

and four orientations (0°, 45°, 90° and 135°). From this, a unique GCM matrix is created by the sum of the four orientations mentioned before. From GCM, several features are extracted: energy, maximum probability, entropy, inertia, and homogeneity. The formula from each feature extracted from GCM can be seen in Table 3. Here,  $i, j$  are the image coordinates,  $GCM$  is the co-occurrence matrix, and  $N$  is the dimension of the  $GCM$  ( $GCM$  is a quadratic matrix).

Another local approach based on a simple local binary pattern (LBP) has been considered as a feature too [25]. In summary, a total of 23 features are obtained, and later they are normalized between 0 and 1, these features are summarized in Table 4.

#### 4.2 Feature relevance measurement: ranking and weighting

Taking into account the fact that the relevance level for each extracted feature is different, four methods to measuring it are proposed. These methods are based on principal component analysis (PCA), dissimilarity measure, and kernel alignment approaches.

Moreover, a last method has been generated by combining the previous ones.

The two first methods proposed are based, as already stated, on the well-known PCA. The key idea behind PCA is to reduce the dimensionality of a data set consisting of a large number of interrelated variables [26]. The PCA method looks for variability in the data and sorts the information of the data in order of importance, hence, PCA has been considered appropriate to measure the feature's relevance. According to the 23 features extracted, 23

principal components are generated. Firstly, an approach that considers the significant presence of each feature in each component of the PCA has been developed. This method will be called PCA-feature-presence method (PCA-FP). Secondly, looking for a quantitative value of a method's output, another approach that considers the feature's score from each eigenvector weighting according to its corresponding proportion of variance has been proposed too. This method will be called PCA-feature-value method (PCA-FV). In the case of PCA-FP, the fact that variance of components decreases progressively has been taken into account. The PCA-FP is a sequential method that ranks each feature according to the importance of the eigenvector (highest proportion's variance greater importance) in which is significantly present or not, i.e., if its absolute score value is higher than a threshold equals to 0.1000 is statistically significant and non-significant in a contrary case. For example, the variables that are nonsignificantly present in the first component (first component equals to eigenvector with the highest proportion's variance) are located in the last ranking positions, and those that are significantly present in all components (until component equals to eigenvector with the lowest proportion's variance) are located in the first ranking positions. Finally, a ranking between 1 to 23 positions is generated.

PCA-FV uses the absolute score value for each coordinate of vector in each feature and the proportion of variance of the component. Each coordinate of eigenvector is weighted by its proportion of variance. Let  $\lambda_c$  be the proportion's variance of eigenvector  $V_c$ ,  $c$  be the number of component ( $c = 1..23$ ), and  $V_c(x)$  be a coordinate of  $V_c$ , where  $x = 1.23$ . The rank is calculated as follows (Eq. 1):

$$\text{Rank}_x = \sum_{c=1}^n \sum_{x=1}^n \lambda_c * V_c(x) \tag{1}$$

That is, the PCA-FV ranking (from the highest to the lowest) the features according to this value ( $\text{Rank}_x$ ).

A different approach has been considered for the third method. This method uses a dissimilarity measure (it will

Table 3 Features extracted from the GCM

Energy	$\sum_{i=1}^N \sum_{j=1}^N [GCM(i, j)]^2$
Max probability	$\max_{i,j} GCM(i, j)$
Entropy	$\sum_{i=1}^N \sum_{j=1}^N GCM(i, j) \log_2[GCM(i, j)]$
Inertia	$\sum_{i=1}^N \sum_{j=1}^N (i - j)^2 GCM(i, j)$
Homogeneity	$\sum_{i=1}^N \sum_{j=1}^N \frac{p(i,j)}{1+(i,j)^2}$

**Table 4** General description of the extracted features. Each color represents a different feature category

Number	Category	Description
1	RGB Color	Mean of Red color channel
2	RGB Color	Mean of Green color channel
3	RGB Color	Mean of Blue color channel
4	RGB Color	Mean of three RGB channels
5	RGB Color	Standard Deviation of three RGB channels
22	RGB Color	Brightness
6	Grayscale	Mean
7	Grayscale	Standard Deviation
8	Geometry	Eccentricity
9	Grayscale Histogram	Entropy
10	Grayscale Histogram	Dispersion
11	Grayscale Histogram	Mean
12	Grayscale Histogram	Standard Deviation
15	Grayscale Histogram	Energy
16	Grayscale Histogram	Kurtosis
13	HSV Color	Mean
14	HSV Color	Standard Deviation
17	Statistics of Co-occurrence Matrix	Energy
18	Statistics of Co-occurrence Matrix	Maximum Probability
19	Statistics of Co-occurrence Matrix	Entropy
20	Statistics of Co-occurrence Matrix	Inertia
23	Statistics of Co-occurrence Matrix	Homogeneity
21	Local Binary Pattern (LBP)	Simple LBP

be called DM for future references), where each feature is ranked by comparing the averages and standard deviations in each subject. This ranking is calculated as follows:

$$\text{Rank}_i = \frac{(|Xn_i - Xm_i|)}{(S(n_i) + S(m_i))} \tag{2}$$

where  $Xn_i$  is the average value of feature  $i$  in  $n$  (subject  $n$ ) and  $S(n_i)$  is the standard deviation of feature  $i$  in the same subject  $n$ .  $Xm_i$  is the average value of feature  $i$  in another subject  $m$  and  $S(m_i)$  is the standard deviation from this same subject  $m$ . This rank calculation tries to identify which features have the highest difference between all subjects and the lowest variation from this mean. Thus, the variables that represent the highest difference between all subjects have a better position in the ranking.

For the fourth method, a kernel-based approach has been considered. Kernel-based methods are increasingly being used for data modeling because of their conceptual simplicity and outstanding performance in many tasks [27]. Specifically, a kernel target alignment approach has been considered as the fourth ranking method (it will be called KA for future references) [28]. With the kernel alignment method, the alignment between each feature and the ideal kernel is calculated. The higher the alignment is, the kernel highly “fits” the class represented by the data. For this, a kernel has been created for each feature and the alignment is calculated by the following expression:

$$A_v = \frac{\sum k_v(i, j) \cdot yy^t(i, j)}{\sqrt{\sum k_v^2(i, j) \cdot \sum_{m=1}^i (ns_m^2)}} \tag{3}$$

where  $k_v$  is a matrix that on its diagonal block contains the values of all samples (from all subjects) in variable  $v$ . That is, the  $k_v$  matrix is constructed concatenating each of the feature values in all samples for all classes. Therefore, the  $k_v$  is a square matrix with dimension equal to  $n$  samples (in all classes), and  $(i, j)$  represents the row and column position. Notice that with this construction procedure, the matrix is made up of several concatenated blocks. Each block corresponds to the samples in each class (each class represents each subject). The ideal kernel is represented by  $yy^t$ , which is a square matrix made up of 1s on its diagonal block and of 0s in the rest of cases. That is,  $\sum k_v(i, j) \cdot yy^t(i, j)$  is the summatory of diagonal block of the  $k_v$  matrix and  $ns$  is the total number of samples in each class (subject)  $i$ . Notice that, for each class there are different number of samples. Therefore,  $A_v$  gives the alignment value of each feature, and looking for the best alignment value for each feature, these values are sorted from the highest to the lowest.

Finally, the last method is made by the average of the results of the four methods previously defined. That is, it works by averaging the feature’s position generated by all methods (it will be called the CM for future references).

Once the features are sorted in a ranking, two different approximations to weight each feature according to its ranking position have been carried out. In the first one, a pre-established score has been assigned to each position. Position 1 has 23 points, position 2 has 22 points, and so on until position 23, which has 1 point (this technique will be named non-parametric weighting). These weights are calculated as Eq. 4 shows, where  $i$  represents a feature number and  $points_i$  represents the points obtained by the feature  $i$ .

$$\text{Weight}_i = \frac{points_i}{\sum_{p=1}^{23} p} \tag{4}$$



In the second one, the weight is obtained from the different output values in the feature ranking methods presented before (this will be named parametric weighting). The parametric weighting can be applied only if the ranking method returns a numeric value as output, not just a position. According to this requirement, this weighting has been only applied to the PCA-FV, DM, and KA methods. As a result, a total of eight different methods to measure the relevance of each feature are proposed: non-parametric PCA-FP, non-parametric and parametric PCA-FV, non-parametric and parametric DM, non-parametric and parametric KA, and finally, the non-parametric combination method (CM). Also, for comparison purposes, the unweighted features have been considered as a baseline approximation.

## 5 Results and discussion

In this section, the evaluation system and the main results are presented. In order to cover the wide operational modes of a surveillance system, two different surveillance scenarios have been considered: mono-camera and multi-camera images. In the mono-camera scenario, the subjects are identified in frames acquired from the same camera. In this case, images are very homogeneous, that is, images from a same person have slight changes according to a single camera. This is a usual and important scenario considered in some works in the state-of-the-art, but is far away from the current situation of the most surveillance infrastructures where a huge amount of cameras are installed. In order to consider this realistic situation, the multi-camera scenario was considered too. In this case, the subjects are extracted from images of different cameras. In this scenario, images from the same person have big changes according to different views from the cameras. As a consequence, images show huge variability in different conditions such as lighting, zoom, and perspective (see for example, Fig. 3).

For both scenarios and for all considered databases, the process of database construction have been done in the same way. First, several samples of each subject have been acquired for each database. These samples were split in two disjoint sets, 30 % of the samples conform the training set and 70 % the testing set. It is important to mention that the training images are from one subset of individuals and testing images are from a second subset of individuals. The experiments have been designed to compare each subject with all people contained in the whole database, not taking into account the time of appearance, that is, the people present before and after of the target persons are considered. This is equivalent to look for subjects in the whole video surveillance database: the recorded images and the current images. Only the number of cameras have been considered to conform to the evaluation scenarios:

subject's images from one camera for the mono-camera scenario, and subject's images from several cameras for the multi-camera scenario. There are several ways to represent a biometric system performance [29]. In this work, two of them oriented to measure the system performance in two different operational modes have been selected: watch-list and Cumulative Match Curve (CMC) tests. In the watch-list evaluation, the purpose is to identify and track a limited number of people (target person) who are on a watch-list. That is, the system has the task to recognize a set of people while rejecting everyone else (e.g., the legal travelers) according to predefined threshold [4]. In this work, the watch-list is defined by all individuals collected in each database (of the testing set), that is, each individual is considered as a suspect and he/she is compared with the rest of people, and so on, until all individuals have been considered as suspects. This approach is the same for all tests (CMC and Wilcoxon). This is because this research intended to create a system with high capacity of generalization, and in this way, each individual is compared to all the individuals in the database. The results obtained in this approach are presented as identification and false positive rate. Furthermore, the CMC test has been done as another way of evaluation [30]. The CMC curve is used as a measure for the  $1 : K$  identification system performance, it judges the margin capabilities of an identification system. In the CMC curve, also known as the "one to many" matching, the identification result is obtained from the first  $K$  best scores [29]. Notice that, in this case, no predefined threshold is applied to these scores. These two evaluation approaches have been selected because they analyze the system in a more restrictive way (watch-list approach) and a more tolerant way (CMC test). In addition, these evaluation measures are widely used in the biometric community. Moreover, because of the high number of situations considered in this work and in order to compare the global performance of all methods, a Wilcoxon signed-rank test was made in the mono-camera and multi-camera scenarios [31, 32]. In summary, the results are obtained from two different scenarios (mono- and multi-camera), two operational evaluations (the watch-list approach and CMC test) have been used, and a general analysis of performance of all methods is shown in a graph of Wilcoxon signed-rank test.

### 5.1 Mono-camera surveillance scenario

In this scenario, a total of 21 suspects from standard databases and 62 suspects from MUBA database were considered. That is, a total of 83 suspects under the mono-camera scenario from 15 different cameras are employed. From each suspect, around of 60 samples were extracted, that is, a total of 5130 images were used in the mono-camera scenario. From these images, 30 % were used in

the training phase and the remaining 70 % were used in the testing phase. It is important to mention that the training images are from one subset of individuals and testing images are from a second subset of individuals. The identification and false alarm rate were calculated for each proposed method in each database.

In order to simplify the presentation of results, only the result of the method that achieves the best performance in each database is shown in Table 5; nevertheless, full results in the Table 9 of the Appendix are presented. The methods based on PCA achieve best results in PETS 2006 and CAVIAR databases. The DM method obtains the best result in PETS 2009, and with MUBA database KA reaches the best result. The best overall result, was achieved in PETS 2006 database, and the worst overall result was obtained with MUBA. This is due to the difference in terms of complexity and quality (explained in Section 3.3) from each database. It can be observed that, in all databases, the identification rate is higher than 92 %. This is a great result given the realistic and difficult conditions of the images.

Figure 5 shows a watch-list ROC curve from the best method in each database. It can be observed that the performance of the proposed methods under the mono-camera scenario is really good, reaching in the best case a 99 % of identification rate with a low false alarm rate of 0.1 %. Under most un-controlled conditions and higher number of suspects encountered in MUBA database, a 92 % of identification rate and 6.5 % of false alarm rate have been achieved. In order to have a better visualization, each point, specified in Table 5, is represented by a red point in Fig. 5.

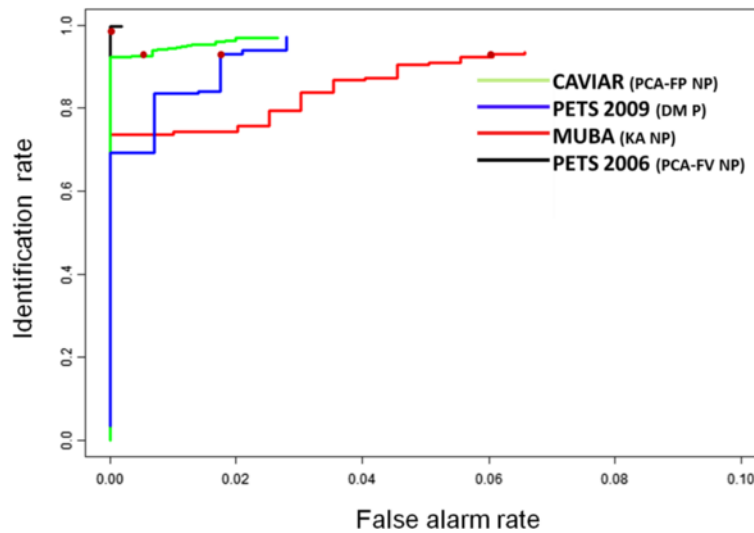
As another way for evaluation, the CMC test was calculated with a  $K$  value of 3. The  $K$  value was established in 3 because with this margin the limitations of the system can be observed taking into account an operation in real environments. Figure 6 shows CMC curves for the best method in each database. For standard databases, only PETS 2009 presents an improvement of the results, and in the other standard databases, there is not a significant improvement when a higher margin is considered ( $K > 2$ ). In the case of MUBA database, the improvement is very relevant: identification rate reaches an 95 % and false alarm rate decreases to 4.84 % when  $K$  increases. That is, when the system returns the three best scores,

the identification rate is above 95 %. The difference in the results between the standard (PETS 2006, PETS 2009, and CAVIAR) and MUBA databases are easily detected. The worst results have been obtained always in the MUBA database due to its complexity level. The main reason that the best method is different in each database is because each database is different from the others, in terms of lighting, image quality, zoom, and camera perspective. Nevertheless, the difference between the results for each database is relatively low. Nevertheless, in the MUBA database, the improvement is higher than in the other databases when a value of  $K$  increases. Moreover, in order to provide a global overview of all methods in all databases, a Wilcoxon signed-rank test was done. For each database, this rank value has been assigned (from 1 for the best method, to 9 for the worst) to all methods used. The standard deviation of these ranks has been calculated. Figure 7 shows the average ranks and the corresponding 95 % confidence interval for each method. It can be observed that PCA-FV with parametric weighting is lightly better than other methods. The main contribution here is that all rankings are always better than the unweighted features. Notice that unweighted features are considered as baseline approximation. In Figure 7, the high value of the standard deviation in most of the methods can be observed; this indicates that each method presents good results in some cases and bad results in others. This means that under the mono-camera scenario, all ranking methods work similarly, only the unweighted features always reach the worst results.

The rankings of the features that achieve the best results in each database are shown in Table 6. Here, it presented the feature ranking obtained by the best method from each database in the mono-camera scenario, each color represents the group of features to which it belongs and the number from 1 to 23 is the position in the ranking, from best (1) to worst (23). That is, for each ranking, the best result in each database was obtained. Here, the different relevance given to each feature or group of features can be observed. In general, in a mono-camera scenario, features based on color are located in the highest position, being very relevant in this scenario. The standard deviation from a grayscale image (feature 7) achieves the best position in the ranking from three databases (top position in PETS 2006 and PETS 2009, and third position

**Table 5** The best result from each database under a mono-camera scenario

Database	Method	Weighted	Identification rate (%)	False alarm rate (%)
PETS 2006	PCA-FV	Non-parametric	99.82 (0.9982 in Fig. 5)	0.18 (0.0018 in Fig. 5)
PETS 2009	DM	Parametric	93.01 (0.9301 in Fig. 5)	2.10 (0.0210 in Fig. 5)
CAVIAR	PCA-FP	Non-parametric	98.21 (0.9821 in Fig. 5)	0.51 (0.0051 in Fig. 5)
MUBA	KA	Non-parametric	92.30 (0.9230 in Fig. 5)	6.59 (0.0659 in Fig. 5)



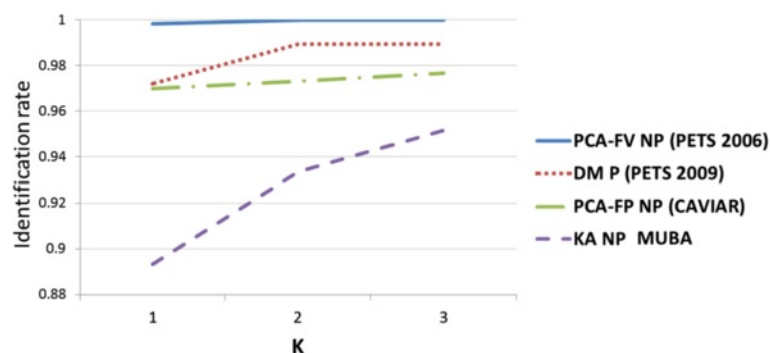
**Fig. 5** Watch-list ROC curve of the best method in each database in a mono-camera scenario

in MUBA). Features of co-occurrence matrix statistics and grayscale histogram have a medium position. It can be concluded that, in the mono-camera scenario, features based on color (both RGB and HSV color spaces) are definitively most important for a people identification task. This is because in this scenario, all information used in the identification process comes from the same camera, in this situation lighting changes are few and the color information is maintained. For a clearer visualization, Fig. 8 shows a graph with the average rank and confidence interval (95 %) of all features weighted with all ranking methods from mono-camera databases. That is, this figure shows each feature average position and its variation in regard to each weighting method. Here, it can be observed that in most cases, color features have low average and low variance values. Feature 8 has a huge

variation in all cases and texture features have a high variance value.

**5.2 Multi-camera surveillance scenario**

In this scenario, a total of 256 subjects from standard databases and from MUBA database were considered, that is, features from a total of 256 different individuals from 24 different cameras were extracted. From each suspect, features were extracted in several frames (approximately 90 per camera). With this, a total of 76,976 images in the multi-camera scenario were used. From these images, 30 % were used in the training phase and the remaining 70 % were used in the testing phase. In order to simplify the presentation of results, only the result of the method that achieves the best performance in each database is shown in Table 7;



**Fig. 6** CMC curve of best methods from each database in mono-camera

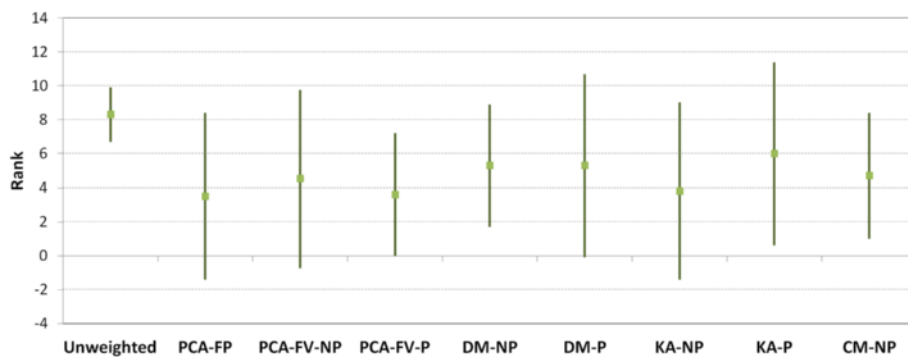


Fig. 7 Average rank and confidence interval from a mono-camera scenario

nevertheless, full results in Table 10 of the Appendix are presented.

In PETS 2006 database, the best result was obtained by DM method; nevertheless, methods based on PCA achieve the best result in three of five databases. Parametric weighting becomes more relevant in multi-camera images, that is, the weighting that uses method's output value works better than a simple weighting. The identification rate, in most of databases, is higher than 89 %. Nevertheless, in CAVIAR4REID database, poor results were achieved with a 75.23 % of identification rate. This is because the images of CAVIAR4REID are cropped to a small size and this affects their quality for recognition purposes (see Fig. 9). But, the results obtained in most of databases are relevant results given the un-controlled and difficult conditions, specially in the MUBA and SAIVT-SoftBio images. The watch-list ROC curve of the best method in each database is shown in Fig. 10. It can be concluded that the performance of the proposed methods under the multi-camera scenario is promising, reaching a 96.16 % of identification rate with a low false alarm rate (lower than 6.90 %). Only in CAVIAR4REID database a high false alarm rate of 24.76 % is obtained. In this curve, it can be seen that the best result was obtained in PETS 2006 database and the worst result in CAVIAR4REID database.

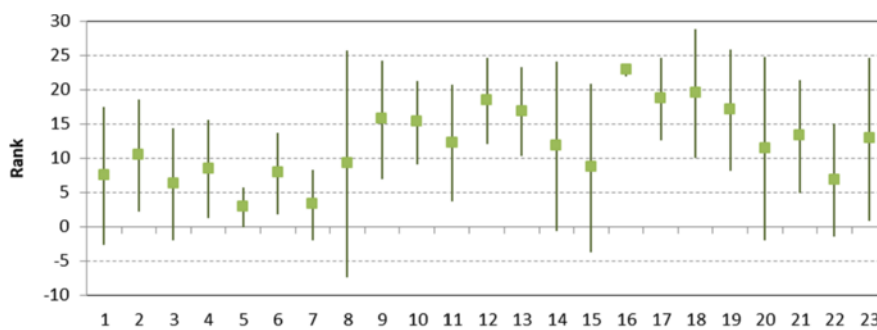
This is easily understood because of the complexity and quality of each database (see Table 1). Furthermore, the cropped images of CAVIAR4REID contributed to obtain bad results in this database. In addition, the CMC test was calculated in the multi-camera scenario. Figure 11 shows CMC curve of the best method in each database in this scenario. It can be observed that when  $K$  equals 3, the improvement of performance is higher than in the previous mono-camera scenario. In a multi-camera situation when the system fails in its first score (or output), most of the times, the system succeeds in the second or third output. That this, as the worst result, there is a 10.35 % of false alarm rate and 89.65 % of identification rate (in CAVIAR4REID database) when the three best scores have been considered. Superior results have been reached in the case of PETS 2006, PETS 2009, SAIVT-SoftBio, and MUBA databases. The highest results were always obtained with PETS 2006 database. This is because of the good quality of this database. In this Figure, the huge difference of performance of CAVIAR4REID database in comparison with the rest of the databases is clear to see.

In Fig. 11, it can be seen that in all cases the increment of the success rate when  $K = 2$  is higher than the success obtained when  $K = 3$ . That is, with a margin of tolerance of  $K = 2$ , the methods obtain promising

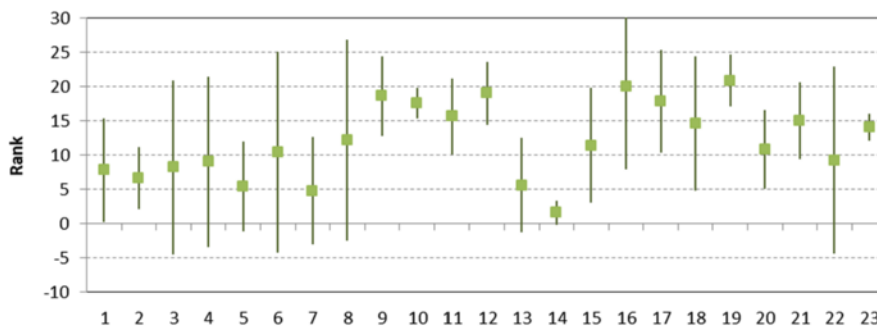
Table 6 Best feature ranking from each database in a mono-camera scenario

Method /Database	Ranking position																						
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
PCA-FV (PETS 2006)	7	15	5	8	1	6	22	4	3	20	14	23	2	21	11	9	13	10	19	17	12	18	16
DM (PETS 2009)	7	5	14	13	2	20	1	18	16	21	11	17	23	9	12	10	15	19	3	4	22	8	6
PCA-FP (CAVIAR)	8	3	2	1	6	22	4	13	5	14	7	20	15	10	12	17	18	23	9	21	11	19	16
KA (MUBA)	8	13	7	1	5	6	22	3	4	2	14	15	23	20	21	11	9	10	12	19	17	18	16

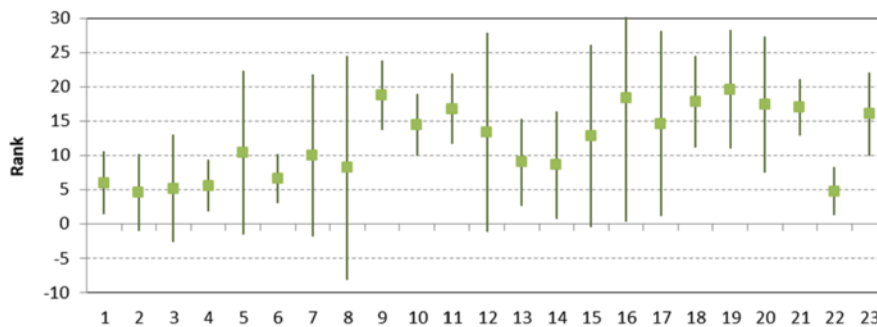
Categories: gray color is RGB category, turquoise color is grayscale category, green color is geometry, orange color is grayscale histogram, yellow color is HSV color, pink color is statistics of co-occurrence matrix and blue color is LBP



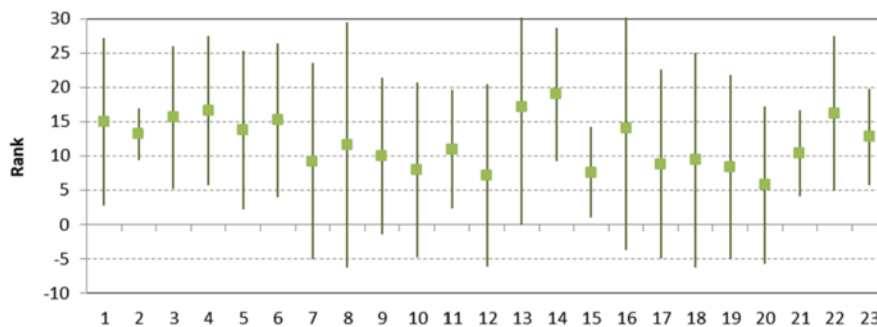
PETS 2006



PETS 2009



CAVIAR



MUBA

Fig. 8 Average rank and confidence interval of all ranking methods from mono-camera databases

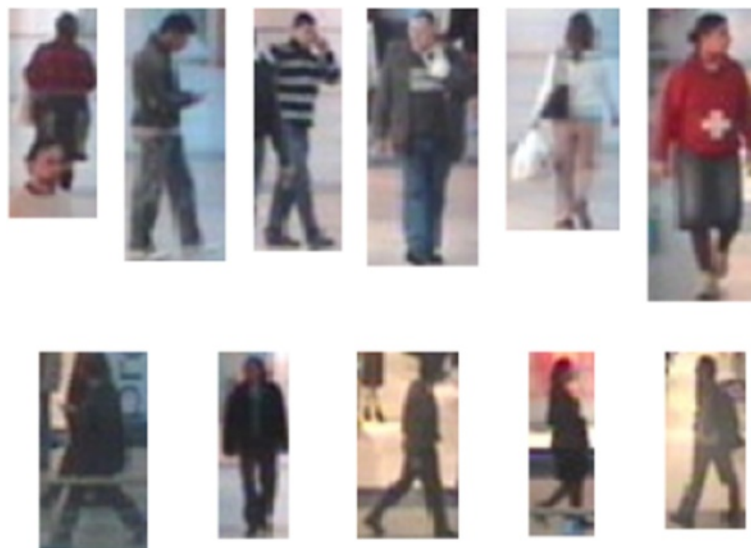
**Table 7** Best result in each database in a multi-camera scenario

Database	Method	Weighted	Identification rate (%)	False alarm rate (%)
PETS 2006	DM	Parametric	96.16 (0.9616 in Fig. 10)	3.78 (0.0378 in Fig. 10)
PETS 2009	PCA-FP	Non-parametric	94.24 (0.9424 in Fig. 10)	5.44 (0.0544 in Fig. 10)
CAVIAR4REID	PCA-FV	Parametric	75.23 (0.7523 in Fig. 10)	24.76 (0.2476 in Fig. 10)
SAIT-SoftBio	KA	Parametric	89.33 (0.8933 in Fig. 10)	10.66 (0.1066 in Fig. 10)
MUBA	PCA-FV	Parametric	93.10 (0.9310 in Fig. 10)	6.90 (0.0690 in Fig. 10)

results. In contrast to the mono-camera scenario where each database has a different best method, in the multi-camera scenario, due to the increase of number of cameras, the best methods in multi-camera are based on PCA. In order to observe a global performance of all methods, a graph of Wilcoxon signed-rank test under the multi-camera scenario is shown in Fig. 12. It can be seen that, although PCA-FP and PCA-FV obtained the best result in PETS 2009, CAVIAR4REID, and MUBA databases, they reached the worst results in the other databases, so they have a huge variability. On the other hand, the DM with parametric weighted is the best method (in global results) with a low standard deviation, that is, the good performance of DM method is independent from environmental situations from each database. The worst results were obtained with CM with non-parametric weighting, PCA-FV with non-parametric, and unweighted features. This is because the CM method averages all results from all ranking methods and, in the case of the multi-camera scenario, the rankings have huge differences (in feature's position) damaging the performance of this combination

method. Unweighted features work usually bad, so the weighted features always work better than them. In spite of PCA-based methods obtained good results, in the multi-camera scenario, the PCA-FV NP method did not have a good performance due to NP weighting affecting its results. From Fig. 12, it can be observed that the difference between performances for each method is high, which is not the case for the mono-camera scenario. This can be interpreted as follows: with easy images, the performance of the methods is homogeneous, and with complex images, the performance of the methods has more variability.

Table 8 shows the feature rankings of the best result in each database. Here, the feature ranking obtained by the best method from each database in the multi-camera scenario is presented, each color represents the group of features to which it belongs and the number from 1 to 23 is the position in the rank, from best (1) to worst (23). It can be observed, in the rankings in the multi-camera scenario, that color features are located in a lower position than in the mono-camera scenario. This is because with multiple



**Fig. 9** Images from CAVIAR4REID

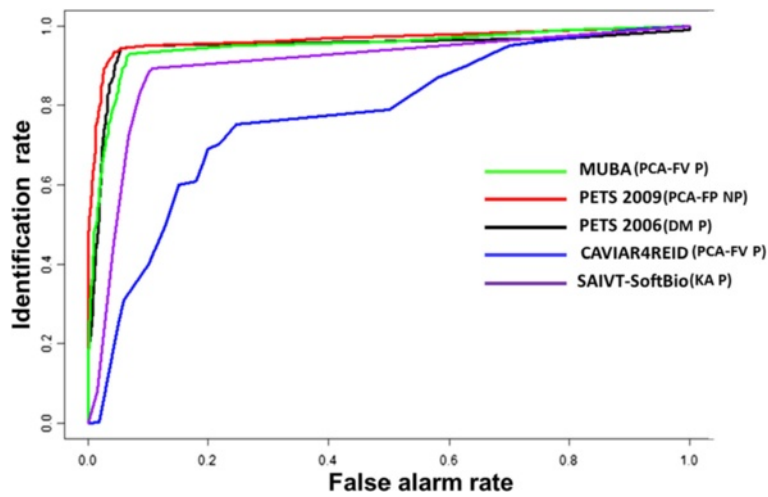


Fig. 10 Watch-list ROC curve of the best method in each database in a multi-camera scenario

cameras, color features are not maintained for different cameras' views. Nevertheless, mean and standard deviation on HSV space color maintain high positions. This is because HSV color space is more robust to lighting changes than RGB color space. Features of co-occurrence matrix statistics and grayscale histogram achieve high positions under the multi-camera scenario. The kurtosis feature achieves the best result in PETS 2006 and PETS 2009. That is, kurtosis helps to reach the best results under the multi-camera scenario. This is because kurtosis is a measure relative to flattening of histogram, and under the mono-camera scenario, using a single camera, the individual representation does not have significant peaks on histogram; in contrast, under the multi-camera scenario, multiple views of people give more different frequencies

(more peaks) into histogram. It can be concluded that, when different views (from different cameras) from the same person are acquired, the appearance related with texture and grayscale histogram statistics are more relevant. For a clearer visualization, Fig. 13 shows a graph with the average rank and confidence interval (95 %) of all features weighted with all ranking methods from multi-camera databases. That is, this figure shows each feature's average position and its variation in regard to each weighting method. In this figure, it can be observed that variance of all features is higher than in mono-camera databases, this confirms that the multi-camera scenario is more complex than the mono-camera. Feature 8 still has huge variance and features related to RGB color are not in first ositions.

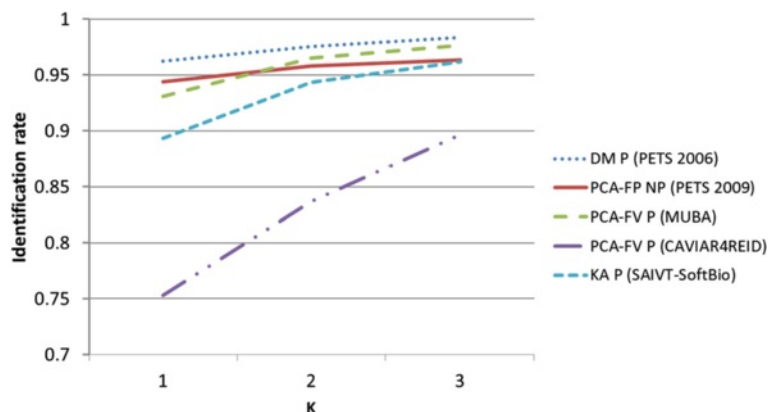


Fig. 11 CMC curve of best method in each database in a multi-camera scenario (with different K values of 1 to 3)

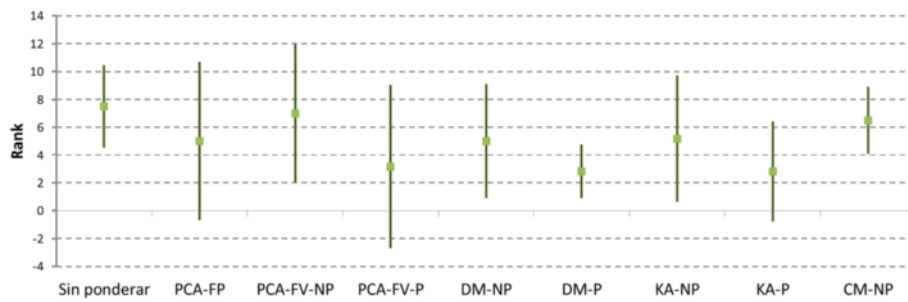


Fig. 12 Average rank and confidence interval from a multi-camera scenario for the proposed methods

### 6 Conclusions

In this paper, a novel approach for human identification in mono and multi-camera surveillance environment scenarios is presented. A bag-of-soft-biometric features is made up of different categories of features: color, texture, local features, and geometry. To measure the relevance of each extracted feature, four methods to rank and two ways of weighting are proposed. Each one of these methods has been deeply analyzed and an optimal configuration to each scenario has been obtained. In order to test the system in a realistic way, several standard databases in the surveillance community have been used: PETS 2006, PETS 2009, CAVIAR, SAIVT-SoftBio, and CAVIAR4REID. Moreover, a new database acquired at Barajas International Airport in Madrid was presented (MUBA database). A total of 83 different suspects and 5130 images have been used in the mono-camera scenario and a total of 256 different suspects and 76,976 images were used in the multi-camera scenario, that is, approximately 4.5 h of video record. Please notice that the proposed method neither needs calibration nor special configuration. That means, the proposed method is flexible and adaptable to the existing conditions of each infrastructure. In the mono-camera scenario, the identification rate was higher than 92 % in all the databases. This is a promising result given the

realistic and difficult conditions of the images, specially of MUBA database. In the CMC test, high results above 95.16 % have been obtained when the system returns the three best scores. Analyzing the categories of features, it can be concluded that, in the mono-camera scenario, features based on color (both RGB and HSV color spaces) are definitively the most relevant for the people identification task. In the multi-camera scenario, the identification rate was higher than 89.33 % with a false alarm rate lower than 10.66 %, in most of databases. Nevertheless, in CAVIAR4REID database, lower results were achieved (75.23 % of identification rate). The parametric weighting, that is, the weighting that uses method's output, becomes more relevant in multi-camera images than mono-camera images. Regarding CMC test, the best result was a 98.34 % of error and a 1.6 % of identification rate when the three best scores were returned, which is an excellent performance given the complex multi-camera environment. Analyzing the ranking of features obtained in the multi-camera scenario, it can be concluded that the appearance related with texture and grayscale histogram statistics are more relevant from multiple cameras. This indicates that when features from different views (from different cameras) are acquired the features related to RGB color are affected by the change in the acquisition camera, in contrast, the features in HSV color space

Table 8 Best feature ranking from each database in a multi-camera scenario

Method	Ranking position																						
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
DM (PETS 2006)	16	14	13	20	12	17	18	5	7	1	10	19	8	9	23	6	22	4	3	21	11	2	15
PCA-FP (PETS 2009)	16	20	21	23	19	9	15	10	5	17	13	14	7	4	22	3	18	6	2	1	12	11	8
PCA-FV (MUBA)	17	12	18	14	19	15	10	2	20	16	9	11	3	6	22	4	23	21	7	5	13	1	8
PCA-FP CAVIAR4REID	15	20	2	23	3	4	22	6	21	18	5	11	1	16	7	9	13	12	17	10	19	14	8
KA SAIVT-SoftBio	8	14	23	2	21	13	3	22	4	6	5	11	10	7	1	15	9	19	12	20	17	18	16

Categories: gray color is RGB category, turquoise color is grayscale category, green color is geometry, orange color is grayscale histogram, yellow color is HSV color, pink color is statistics of co-occurrence matrix and blue color is LBP



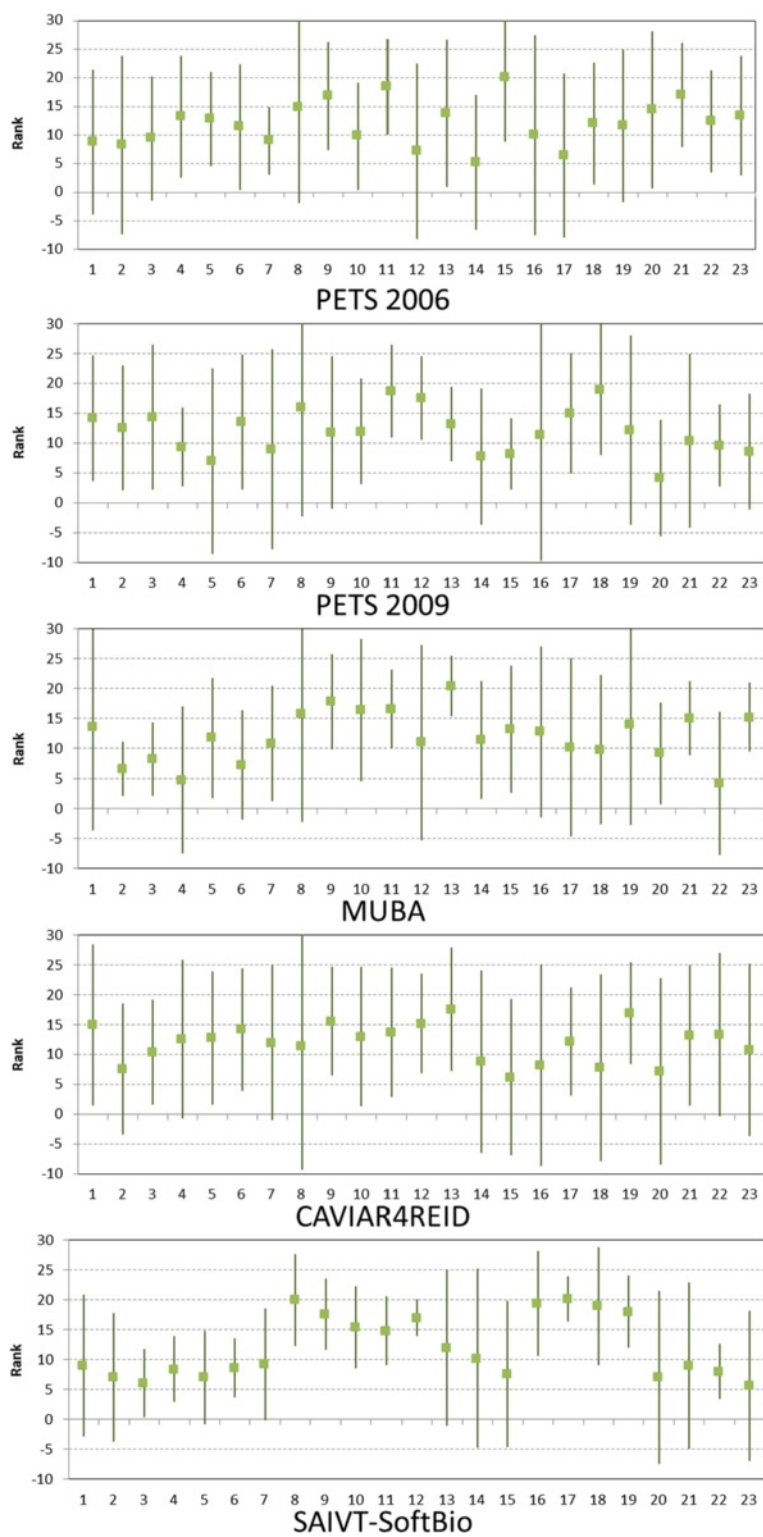


Fig. 13 Average rank and confidence interval of all ranking methods from multi-camera databases

are more robust to lighting changes. In both scenarios, using the unweighted features, the worst results were reached.

Due to the distinct conditions and restrictions that are imposed to each of them, it is very important to rank and to weight the features according to each scenario, so the difference of feature rankings between mono- and multi-camera scenarios can be observed.

Thus, the presented work implies an optimal system configuration according to each scenario. In addition, the results obtained demonstrate the promising potential of the soft-biometric approach.

### Appendix

In this section, full results are presented. Table 9 shows all results obtained in mono-camera scenario (watch-list

**Table 9** Full obtained results in mono-camera scenario

Database	Method	Weighted	Identification rate (%)	False alarm rate (%)
PETS 2006	PCA-FV	Non-parametric	99.82	0.18
	PCA-FV	Parametric	99.63	0.37
	DM	Non-parametric	99.45	0.55
	DM	Parametric	99.27	0.73
	KA	Non-parametric	99.09	0.91
	KA	Parametric	98.90	1.10
	PCA-FP	Non-parametric	98.72	1.28
	CM	Non-parametric	98.72	1.28
	Without weighted	–	98.54	1.46
PETS 2009	DM	Parametric	93.01	2.10
	DM	Non-parametric	93.01	2.80
	PCA-FV	Non-parametric	90.56	3.50
	CM	Non-parametric	90.21	3.50
	KA	Parametric	90.21	3.85
	PCA-FV	Parametric	89.86	3.85
	PCA-FP	Non-parametric	89.51	3.50
	KA	Non-parametric	89.16	3.50
	Without weighted	–	88.81	3.50
CAVIAR	PCA-FP	Non-parametric	98.21	0.51
	PCA-FV	Non-parametric	97.83	0.13
	DM	Parametric	97.70	0.13
	PCA-FV	Parametric	97.57	0.26
	DM	Non-parametric	97.06	0
	CM	Non-parametric	96.93	0.13
	Without weighted	–	96.68	2.04
	KA	Non-parametric	96.17	0
	KA	Parametric	95.91	0
MUBA	KA	Non-parametric	92.30	6.59
	PCA-FP	Non-parametric	91.93	6.46
	PCA-FV	Parametric	91.21	7.40
	CM	Non-parametric	91.15	7.39
	PCA-FV	Non-parametric	90.20	8.75
	DM	Non-parametric	89.83	7.45
	DM	Parametric	89.56	6.99
	KA	Parametric	87.65	6.63
	Without weighted	–	87.10	10.53

**Table 10** Full obtained results in multi-camera scenario

Database	Method	Weighted	Identification rate (%)	False alarm rate (%)
PETS 2006	DM	Parametric	96.16	3.78
	DM	Non-parametric	96.16	3.84
	PCA-FV	Parametric	95.51	4.37
	KA	Parametric	95.26	4.68
	PCA-FP	Non-parametric	94.89	5.05
	KA	Non-parametric	94.58	5.30
	CM	Non-parametric	94.58	5.30
	Without weighted	–	94.58	5.36
	PCA-FV	Non-parametric	94.42	5.42
PETS 2009	PCA-FP	Non-parametric	94.24	5.44
	DM	Parametric	94.09	5.52
	KA	Parametric	94.01	5.44
	KA	Non-parametric	94.01	5.76
	Without weighted	–	93.93	5.76
	CM	Non-parametric	93.85	5.76
	DM	Non-parametric	93.53	5.91
	PCA-FV	Non-parametric	92.11	7.65
	PCA-FV	Parametric	91.20	8.40
MUBA	PCA-FV	Parametric	93.10	6.90
	KA	Parametric	92.68	7.25
	KA	Non-parametric	92.01	7.99
	DM	Parametric	91.91	8.06
	DM	Non-parametric	91.10	8.90
	CM	Non-parametric	90.62	9.35
	PCA-FV	Non-parametric	90.20	9.74
	Without weighted	–	90.10	9.83
	PCA-FP	Non-parametric	90.04	9.83
SAIVT-SoftBio	KA	Parametric	89.33	10.66
	PCA-FV	Parametric	89.24	10.75
	DM	Parametric	88.94	11.05
	Without weighted	–	88.86	11.13
	KA	Non-parametric	88.84	11.15
	CM	Non-parametric	88.57	11.42
	DM	Non-parametric	88.02	11.97
	PCA-FP	Non-parametric	87.52	12.47
	PCA-FV	Non-parametric	87.09	12.90
CAVIAR4REID	PCA-FV	Parametric	75.23	24.76
	DM	Non-parametric	73.85	26.15
	DM	Parametric	72.71	11.05
	PCA-FV	Parametric	72.64	27.36
	DM	Non-parametric	72.56	27.44
	CM	Non-parametric	71.85	28.15
	KA	Parametric	71.38	28.62
	KA	Non-parametric	70.71	29.29
	Without weighted	–	63.13	36.87

approach). In this table, the results are sorted from best to worst.

Table 10 shows all results obtained in multi-camera scenario (watch-list approach). In this table, the results are sorted from best to worst.

#### Competing interests

The authors declare that they have no competing interest.

#### Acknowledgements

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 312797.

#### Author details

<sup>1</sup>Centro de Investigación en Geografía y Geomática "Ing. Jorge L. Tamayo", A.C., Contoy 137 Esq. Chemax, Col. Lomas de Padierna, Delegación Tlalpan, 14240 Mexico D.F., Mexico. <sup>2</sup>Rey Juan Carlos University, Tulipan S/N, 97200 Móstoles, Madrid, Spain.

Received: 31 October 2014 Accepted: 3 July 2015

Published online: 18 August 2015

#### References

- L An, X Chen, M Kafai, S Yang, B Bhanu, in *Distributed Smart Cameras (ICDSC), 2013 Seventh International Conference On*. Improving person re-identification by soft biometrics based reranking (Palm Springs, CA, 2013), pp. 1–6
- J-H Kao, C-Y Lin, W-H Wang, Y-T Wu, A unified hierarchical appearance model for people re-identification using multi-view vision sensors. *Adv. Multimedia Inform. Process. - PCM*. **5353**, 553–562 (2008)
- L Lamontagne, M Marchand (eds.), *Advances in artificial intelligence, 19th conference of the Canadian society for computational studies of intelligence, canadian ai 2006, Québec City, Québec, Canada, June 7-9, 2006, proceedings*, (Canadian Conference on AI, 2006), pp. 445–456
- B Kamgar-Parsi, W Lawson, B Kamgar-Parsi, Toward development of a face recognition system for watchlist surveillance. *Pattern Anal. Mach. Intell. IEEE Trans.* **33**(10), 1925–1937 (2011)
- L Ma, X Yang, D Tao, Person re-identification over camera networks using multi-task distance metric learning. *Image Process. IEEE Trans.* **23**(8), 3656–3670 (2014)
- H-M Moon, SB Pan, in *International Conference on Computer Communication Networks*. A new human identification method for intelligent video surveillance system (Zurich, 2010), pp. 1–6
- E Martinson, W Lawson, JG Trafton, in *Human-Robot Interaction (HRI), 2013 8th ACM/IEEE International Conference On*. Identifying people with soft-biometrics at fleet week (Tokyo, 2013), pp. 49–56
- A Li, L Liu, S Yan, in *Person Re-Identification. Advances in Computer Vision and Pattern Recognition*, ed. by S Gong, M Cristani, S Yan, and CC Loy. Person re-identification by attribute-assisted clothes appearance, (2014), pp. 119–138. doi:10.1109/TCSVT.2014.2352552
- UP Koichiro Niinuma, AK Jain, Soft biometric traits for continuous user authentication. *IEEE Trans. Inform. Forensics Secur.* **5**, 771–780 (2010)
- E Monari, J Maerker, K Kroschel, in *International Conference on Advanced Video and Signal-Based Surveillance*. A robust and efficient approach for human tracking in multi-camera systems (Genova, 2009), pp. 134–139
- L Goldmann, M Karaman, JTS Minquez, T Sikora, in *7th International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS 2006)*. Appearance-based person recognition for surveillance applications, (2006)
- VLA Vanacloig, JAR Ortega, GA García, JMV González, in *ICPR*. People and luggage recognition in airport surveillance under real-time constraints (Tampa, FL, 2008), pp. 1–4
- N Gkalelis, A Tefas, I Pitas, in *ICIP*. Human identification from human movements (Cairo, 2009), pp. 2585–2588
- C Wang, J Zhang, L Wang, J Pu, X Yuan, Human identification using temporal information preserving gait template. *Pattern Anal. Mach. Intell. IEEE Trans.* **34**, 2164–2176 (2012)
- P Tome, J Fierrez, R Vera-Rodriguez, MS Nixon, Soft biometrics and their application in person recognition at a distance. *Inform. Forensics Secur. IEEE Trans.* **9**(3), 464–475 (2014)
- D Thirde, L Li, J Ferryman, in *9th IEEE International Workshop on PETS*. Overview of the PETS 2006 challenge, (2006), pp. 47–50
- A Ellis, A Shahrokni, J Ferryman, in *11th IEEE International Workshop on PETS*. Overall evaluation of the PETS 2009 results, (2009), pp. 117–124
- CAVIAR, *CAVIAR Benchmark Data*, (2004). Online: <http://groups.inf.ed.ac.uk/vision/CAVIAR>
- A Bialkowski, S Denman, S Sridharan, C Fookes, P Lucey, in *Digital Image Computing Techniques and Applications (DICTA), 2012 International Conference On*. A database for person re-identification in multi-camera surveillance networks (Fremantle, WA, 2012), pp. 1–8
- DS Cheng, M Cristani, M Stoppa, L Bazzani, V Murino, in *British Machine Vision Conference (BMVC)*. Custom pictorial structures for re-identification, (2011), pp. 68.1–68.11
- AM Rohaly, PJ Corriveau, JM Libert, AA Webster, V Baroncini, J Beerends, J-L Blin, L Contin, T Hamada, D Harrison, AP Hekstra, J Lubin, Y Nishida, R Nishihara, JC Pearson, AF Pessoa, N Pickford, A Schertz, M Visca, AB Watson, S Winkler, Video quality experts group: current results and future directions, 742–753 (2000). doi:10.1117/12.386632
- J You, MM Hannuksela, M Gabbouj, in *IEEE International Conference on Image Processing*. An objective video quality metric based on spatiotemporal distortion (Cairo, 2009), pp. 2229–2232
- C Conde, D Moctezuma, IM de Diego, E Cabello, Hogg: Gabor and hog-based human detection for surveillance in non-controlled environments. *Neurocomputing*. **100**, 19–30 (2013)
- A Vadivel, S Sural, AK Majumdar, An integrated color and intensity co-occurrence matrix. *Pattern Recognition Letters*, 974–983 (2007). doi:10.1016/j.patrec.2007.01.004
- S Moore, R Bowden, Local binary patterns for multi-view facial expression recognition. *Comput. Vis. Image Underst.* **115**(4), 541–558 (2011)
- IT Jolliffe, Discarding variables in a principal component analysis. ii: Real data. *J. R. Stat. Soc. Ser. C (Applied Statistics)*. **21**(2), 21–31 (1973)
- N Cristianini, J Kandola, A Elisseeff, J Shawe-Taylor, On kernel-target alignment. *Adv. Neural Inform. Process. Syst.* **19**, 367–373 (2002)
- PH Gosselin, F Precioso, S Philipp-Foliguet, Incremental kernel learning for active image retrieval without global dictionaries. *Pattern Recognit.* **44**(10-11), 2244–2254 (2011)
- PJ Phillips, P Grother, RJ Micheals, DM Blackburn, E Tabassi, M Bone, NF Dr, U Kingdom, PJ Phillips, P Grother, RJ Micheals, DM Blackburn, E Tabassi, M Bone, Face recognition vendor test 2002: Evaluation report (2003). doi:10.1109/AMFG.2003.1240822
- RM Bolle, JH Connell, S Pankanti, NK Ratha, AW Senior, in *Fourth IEEE Workshop on Automatic Identification Advanced Technologies, 2005*. The relation between the ROC curve and the cmc curve, (2005), pp. 15–20. doi:10.1109/AUTOID.2005.48
- J Demser, Statistical comparisons of classifiers over multiple data sets. *J. Mach. Learn. Res.* **7**, 1–30 (2006)
- F Wilcoxon, Individual comparisons by ranking methods. *Biometrics Bulletin*, 80–83 (1945). <http://www.jstor.org/stable/3001968>

Submit your manuscript to a SpringerOpen® journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](http://springeropen.com)