People re–identification in camera networks based on probabilistic color histograms

Angela D'Angelo[a] and Jean-Luc Dugelay[a]

[a]EURECOM, Multimedia Communications Department, Sophia Antipolis, France

ABSTRACT

People tracking has to face many issues in video surveillance scenarios. One of the most challenging aspect is to re–identify people across different cameras. Humans, indeed, change appearance according to pose, clothes and illumination conditions and thus defining features that are able to robustly describe people moving in a camera network is a not trivial task. While color is widely exploited in the distinction and recognition of objects, most of the color descriptors proposed so far are not robust in complex applications such as video surveillance scenarios.

A new color based feature is introduced in this paper to describe the color appearance of the subjects. For each target a probabilistic color histogram (PCH) is built by using a fuzzy K–Nearest Neighbors (KNN) classifier trained on an ad-hoc dataset and is used to match two corresponding appearances of the same person in different cameras of the network. The experimental results show that the defined descriptor is effective at discriminating and re-identifying people across two different video cameras regardless of the viewpoint change between the two views and outperforms state of the art appearance based techniques.

Keywords: People tracking, people re-identification, video surveillance, probabilistic color histogram

1. INTRODUCTION

The proliferation of video cameras in public places such as airports, train stations, parking lots and shopping malls create many opportunities for public safety applications, including surveillance for threat detection, customers tracking in a store, detecting of unusual events, monitoring of elderly people at home, etc. Manual supervision of these videos, however, is a cumbersome task due to the large volumes of information. Therefore, it would be advantageous to automate this procedure by using a computer vision system that is able to extract the useful information represented in the video and to perform specific tasks depending on the security scenario.

Despite video surveillance has been in use for decades, the development of systems that can automatically detect and track people is still an active research area. Specifically, in many video-surveillance applications, it is desirable to determine if a presently visible person has already been observed somewhere else in the cameras network. This problem has been addressed in the scientific literature with the term of people re–identification, the ultimate goal of such a system is to correctly tag all instances of the same visual object at any given location and at any given time instant. In most cases, such a system relies on an appearance-based model that depends on several factors, such as illumination conditions, different camera angles and pose changes. Among the possible features, it is widely recognized that color is an important and powerful cue in the distinction and recognition of objects. Most of the color descriptors proposed so far, however, are based on simple approaches like color histograms in different color spaces, color correlograms or dominant colors. Unfortunately these features are not robust to the challenges mentioned above and typical of video surveillance scenarios.

To counter this problem a new color descriptor is introduced in this paper. For each target a probabilistic color histogram (PCH) is introduced to represent the color appearance of the subject. Such histogram is

Further author information:
A.D.: E-mail: angela.dangelo@eurecom.fr, Telephone: +33 (0)4 93 00 81 55
J.L.D.: E-mail: Jean-Luc.Dugelay@eurecom.fr, Telephone: +33 (0)4 93 00 81 41

built by using a fuzzy K–Nearest Neighbors (KNN) classifier and is then used to match the appearances of the same person in two different cameras of the network. The experimental results will demonstrate the robustness of the proposed descriptors in re–identifying people across diffent cameras.

The paper is organized as follows. After a section dedicated to related works, in Section 3 the proposed color descriptor is defined. A fuzzy color quantization approach based on the concept of culture colors is introduced and the use of probabilistic color histogram in people tracking is then explained. The experimental results are presented in Section 4, the performance are analyzed and compared with state of the art systems. Finally, in the conclusion, we summarize the contributions of our approach and we derive some ideas for future works.

2. RELATED WORKS

The attempts of addressing the re–identification problem in video-surveillance scenarios have been manifolds in the last years. The state of the art techniques can be divided in two main categories. The first group includes all those methods which take into account appearance based models along with spatial knowledge of the network, while to the second group belong all the techniques that are uniquely based on the appearance of the subjects without any spatial information of the cameras. The second class of algorithms is more challenging and most of the researchers are working in this direction. Since the proposed approach belongs to this second category, and in order to clarify our contribution, a brief description of the state of the art appearance based models follows.

Template methods are based on the comparison of image pairs according to a simple distance metric. These techniques have the advantage of being straightforward, but they fail as pose and viewpoint deviate from the template. The lack of robustness can be partially fixed by aligning the query image to the template. This strategy is efficient when dealing with rigid objects as vehicles,[1] nevertheless the efficency drops as not rigid targets (e.g. humans) are considered.

Subspace methods together with manifolds are sometimes used to identify people thanks to their property to be pose and viewpoint invariant. Lee et al.,[2] as an example, use a full subspace of nonrigid objects approximated by nonlinear appearance manifold. While this method is efficient in describing local features, e.g. face features, it becomes too complex to accurately describe a full body.

A class of methods is based on *localized features*, that is a bag of features is used to create a distinctive trait representing the subject to be re–identified. In this context, the promising work from Gray et al.[3] uses an ensemble of local features mainly based on colors and textures, together with a machine learning approach, to compute the signature for each subject. The drawback of this technique is that the increment of performance rates implies the rise of dimensionality of the feature space and the higher complexity of the learning method.

Another common way to discriminate among different targets is represented by *histogram–like methods*.[4] The scientific literature in the field of CBRS (Content Based Retrieval Systems) have already showed that such techniques are well suited for retrieving images similar from a content point of view, but they are strongly affected by changes in appearance. The main advantage of the histogram–like methods, which make them so popular in the literature, is that they are straightforward and fast to compute.

In this paper we try to overcome the previously mentioned limitiations by introducing a new robust color based descriptor to model the visual appearance of the subjects. The desidered properties of the proposed descriptor are: being invariant to changes in the viewpoints; being robust to variations in illumination conditions and to different sensors; being adaptive, that is independent from the topology of the camera network; being fast to compute and easy to be adopted in real applications.

3. VISUAL SIGNATURE

As already explained in the previous section, the goal of the proposed work is to introduce a new visual signature based on color to reliably describe and identify subjects in a camera network. Fig.1 shows an

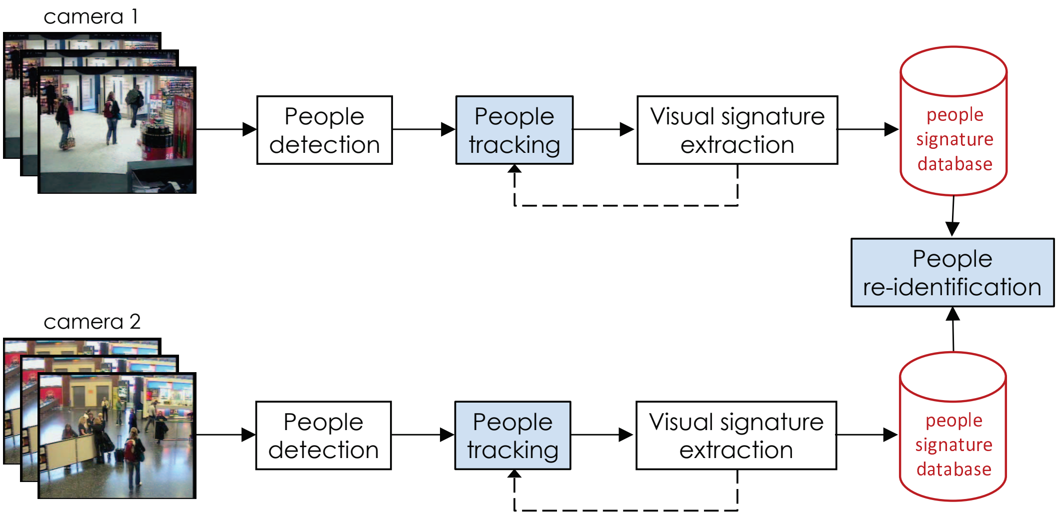


Figure 1. The block diagram presents a possible application of the proposed visual signature. After detection and tracking of targets in the scene, a visual signature is computed. For each camera, a database is created for all the people crossing the camera FOV. The re–identification module computes the distance between two different persons exploiting the corresponding signatures in the databases. The visual signature can also be used to improve the robustness of the tracking.

example of how the proposed signature could be integrated in a video surveillance system and used to re–identify subjects across the fields of view (FOVs) of different cameras.

For each camera of the network, after detection and tracking of targets in the scene, the visual signature of each subject is computed and stored in a database. The people re–identification process is performed by comparing the visual signatures from all the cameras in the network. Even if out of the scope of this paper, the color descriptor can also be used to improve the robustness of the tracking (dotted line in the figure). As a matter of fact, despite the existence of several works aiming at introducing robust people tracking algorithms, this is still a challenging task in video surveillance scenarios. Exploiting color information with a robust feature can thus be useful to improve the results of the tracking in case of complex scenarios like body occlusions.

The proposed signature is computed in two steps. In the first step, a color quantization of the subject is performed to introduce robustness in case of complex illumination conditions. The second step is the evaluation of the color descriptor based on the quantized versions of the body parts. These processes are explained in details in the following paragraphs.

3.1 Fuzzy color quantization

Color is an important and powerful cue in the distinction and recognition of objects, however defining a robust color descriptor could be a not easy task. This is particularly true in the case of video-surveillance systems. To simplify the analysis, many color constancy models assume a single camera, a frontal surface orientation or a spatially–invariant illumination. In the reality, unluckily, we must account for spatially–distributed surveillance cameras operating under different lighting conditions and with varying color sensitivity. The complexity of such modeling for identifying colors in video sensors, makes the task very difficult.

In order to achieve robustness in a so complex scenario, the first step of the proposed approach consists of performing a color quantization based on eleven colors: black, white, red, yellow, green, blue, brown, purple, pink, orange, and grey. These colors are usually referred to as *culture colors*[5] and describe the set of color terms that can be considered as universal constants among languages and cultures. Culture colors represent the way the human beings perceive and describe colors. One might argue that having a finer quantization may better discern different objects. Berlin et al.[5] showed that finer quantization leads to less reliable color prediction, and can be counter-productive in improving prediction accuracy.

Color quantization in the eleven culture colors is performed in the Lab color space using a fuzzy k-nearest neighbor (KNN) clustering algorithm.[6] In fuzzy clustering, data elements can belong to more than one cluster, and associated with each element n is a membership vector $\mathbf{u}(n) = \{u_1(n), u_2(n), ..., u_C(n)\}$ indicating the strength of the association between that data element n and all the possible C clusters. The possible clusters in our case are the eleven previously defined culture colors, thus $C = 11$.

The KNN algorithm is trained on a dataset of samples describing pixels in the Lab color space and associated to the eleven culture colors. The design of the training set is crucial in the proposed approach. To this aim, we follow the approach described by D'Angelo et al.[7] consisting of collecting samples describing the culture colors under various lighting conditions and from different sensors camera. In order to obtain a so diversified dataset of colors, the authors collected pixel samples from video clips of sport teams with the color of the uniforms corresponding to the culture colors. The video clips of the selected teams were randomly chosen from the web. This procedure allow to obtain a great number of samples in real illumination conditions and with high probability from different sensors and thus to obtain a quantization process as much as possible robust to variations in illumination. Following the described approach, we collected about 1200 samples that we use as training set for the discussed classifer.

Subjects walking across the field of views of a camera network can thus be identified using the color information of the clothes they wear. Generally we can distinguish two main colors (or set of colors) describing the upper and lower parts of the subjects (let us call them segments) that can be automatically extracted from the bounding boxes of the detected subjects. Based on these considerations, the fuzzy KNN classifier is applied to each pixel of the Lab version of the selected segments using the training set designed as above. The classifier assigns to each pixel a label corresponding to one of the classes describing the culture colors, as we can observe in the example shown in Fig.2, where it is applied to the torso of a subject extracted from the CAVIAR* database.

(a) (b) (c)

Figure 2. Example of fuzzy color quantization: (a) bounding box of the detected people; (b) extracted torso; (c) quantized torso through fuzzy KNN classifier.

3.2 Probabilistic color histogram

The advantage of using a soft classifier (like the fuzzy classifier) with respect to the hard classifier is that to each pixel n is also associated the membership vector $\mathbf{u}(n) = \{u_1(n), u_2(n), .., u_{11}(n)\}$ describing the probabilities of the membership of the pixel to the 11 possible classes.

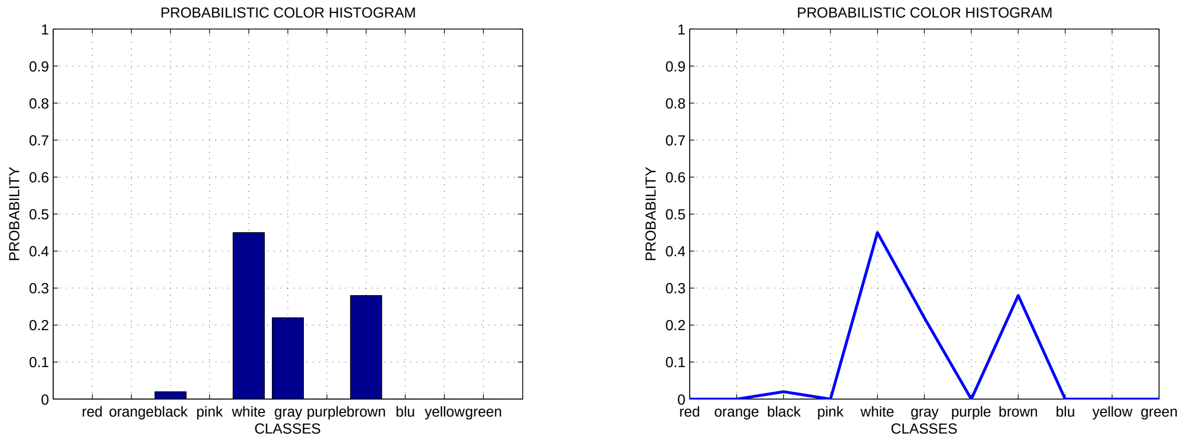
To each quantized segment X is assigned a color descriptor which is based on the definition of a **probabilistic color histogram** (PCH) $\mathbf{H}(X) = \{H_1(X), H_2(X), .., H_{11}(X)\}$ described as follows:

$$H_c(X) = \frac{1}{N} \sum_{n=1}^{N} u_c(X_n), \quad c = 1, ..., 11 \tag{1}$$

*available for download at http://homepages.inf.ed.ac.uk/rbf/CAVIARDATA1/

and representing for each class c the sum of the probabilities that all the pixels in the segment (in this case the torso) belong to that class.

The PCH corresponding to the quantized torso in Fig.2.(c) is shown in Fig.3.(a). As expected the dominant color is white. The same PCH is plotted using a different representation in Fig.3.(b). This representation is introduced because it will be used later in the paper to analyze the behaviour of the PCH distribution in consecutive frames.



(a) (b)

Figure 3. Example of probabilistic color histograms using two different representations.

The probabilistic color histogram is the visual signature associated to each subject in the video and can be used, as explained at the beginning of this section, to improve the robustness of the tracking or to re–identify people across different cameras. In the first case we talk about the *intra camera scenario*, that is the color information of a subject could be exploited to improve the efficiency of the tracking in the camera. In the second case we are in the *inter camera scenario*. With this term we refer to problem of recognizing people walking in the FOVs of a camera network. While the final goal of this paper is to provide a robust color descriptor with application to the re-identification problem, in the next paragraphs we will explain how to use the probabilistic color histogram described by eq.(1) in both the scenarios.

3.2.1 Intra camera scenario

In the first scenario we could use the probabilistic color histogram to compare two segments X_1 and X_2 extracted from the bounding boxes of a subject in consecutive frames in order to decide if they belong or not to the same person. We can expect, in fact, that the torsos (or legs) of a subject extracted from consecutive frames have similar PCHs. This is confirmed if we look at Fig.4 where the PCHs referring to the torsos of a subject extracted from 46 consecutive frames are plotted. As expected the extracted torsos have similar distributions. If this were not the case, we would have expected an occlusion on the detected person, or more generally an error in the detection step.

The comparison between consecutive frames is done using a distance metric to check the similarity between the corresponding probabilistic color histograms. If the distance of the two histograms is lower than a fixed threshold T, then the two analyzed segments belong to the same person. In the experimental results we introduce two distance metrics that can be used to this aim.

3.2.2 Inter camera scenario

In the second scenario of re–identification across different cameras, to each person p walking throughout a camera, a visual signature is assigned which is obtained by averaging the probabilistic color histograms associated with the segments extracted from each frame, as follows:

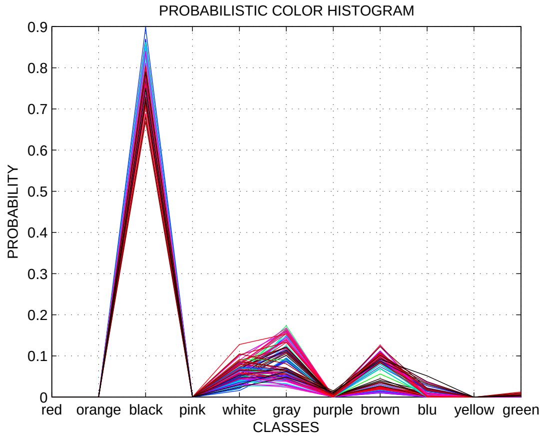


Figure 4. Example of PCHs of torsos extracted in 46 consecutive frames.

$$\bar{\mathbf{H}}(p) = \frac{1}{F}\sum_{f=1}^{F}\mathbf{H}(X_p(f)) \tag{2}$$

where F is the number of frames, and $X_p(f)$ is the segment of the person p extracted in the frame f. Obviously, the averaging operation across different frames helps to provide a more robust visual signature associated to each person than using only one frame. Sometimes, in fact, color appearance could vary also in consecutive frames due to sudden changes in the illumination conditions or to changes of the pose of the subject respect to the camera. An example of this variability is shown in Fig.5 (extracted from the TRECVID dataset†) in which we can observe how the colors of the shirt is drastically changing between the two cameras of the network but also in consecutive frames of the same camera. The PCHs of the torso extracted in 46 consecutive frames from the first camera (Fig.5.(a)) are shown in Fig.6 (each plot corresponds to a different frame). While the segments contain almost the same colors (black, white, gray, brown, green), the histograms distribution drastically changes from one frame to another. By averaging the PCHs we can decrease this variability and increase the probability of the correct re–identification.

In the inter camera scenario, the re–identification is obtained by evaluating the distance between $\bar{\mathbf{H}}(p_1)$ and $\bar{\mathbf{H}}(p_2)$, the average visual signatures of two persons p_1 and p_2. Like for the intra-camera scenario, if the distance is lower than the threshold T, then the two persons match. Obviously, the higher the value of T the higher the correct identification (true positive) rate is but the higher the false alarm rate is.

As explained in the introduction, in this work we decided to focus our attention on the inter camera scenario, that is the problem of people re–identification in a camera network, since this is a challenging problem for the mentioned reasons. Thus, the following section will be oriented at evaluating the performance of the proposed color feature as a visual signature useful to re–identify people across different cameras.

4. EXPERIMENTAL RESULTS

In this section we summarize the experimental results that validates our people re–identification algorithm. The first part will be devoted to the dataset description: a new challenging dataset has been used in order to have a full validation of our system. Secondly we will describe the benchmarking strategy. A full list of baseline algorithms will be described and their characteristics presented for the sake of completeness. Finally we will present our experimental results and achievements using the state of the art validation tool.

†the dataset is available for download at http://www-nlpir.nist.gov/projects/trecvid/trecvid.data.html.

Figure 5. Example of extracted bounding boxes from the TRECVID dataset: (a) frame 0, 14, 33 and 45 from the first camera; (a) frame 0, 14, 33 and 45 from the second camera.

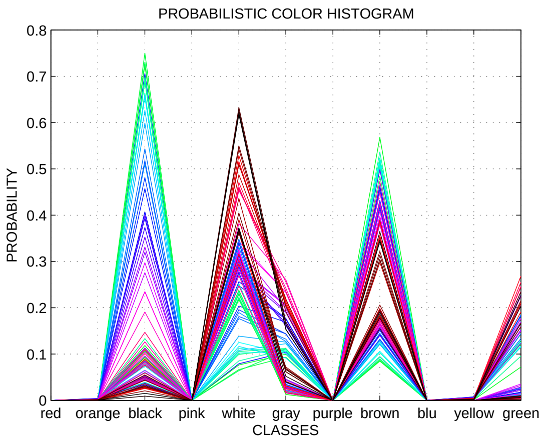


Figure 6. Examples of PCHs of torsos extracted in 46 consecutive frames.

4.1 Dataset

The first requirement to test the efficiency of a re-identification system is to select a dataset comparable to a real video surveillance scenario. Given a single image, indeed, the chance of choosing the correct match is inversely proportional to the size of the dataset. Unfortunately, a limitation of most of the state of the art works is that the experimental results are performed on very limited database ($10 - 40$ subjects) and on frontal view images, far away from real scenarios.

To the best of our knowledge, the most challenging dataset is VIPeR (Viewpoint Invariant Pedestrian Recognition),[8] introduced in the research field of *pedestrian tracking* which can be considered as a subset of the people tracking topic. The authors have collected two views of 632 individuals seen from widely differing viewpoints. Most of the subjects in the database contain a viewpoint change of 90 degrees or more, making recognition very challenging. The method presented in this paper is evaluated using their public dataset[‡]. Some examples from this dataset can be found in Fig.7.

Figure 7. Some examples from the VIPeR dataset. Each column is one of the 632 pairs of images.

Other than introducing the new dataset, Gray et al.[8] in their paper perform an extensive comparison of state of the art systems. We will refer to their experimental results to test the efficiency of the proposed method.

4.2 Benchmark

As already discussed in Section 2, the state of the art is manifold in the attempted approaches, but mainly two methodologies were found. Since the nature of our re–identification approach is only based on the appearance model, disregarding any additional independent information coming from other variables or properties of the camera network, we aim at comparing our work with similar approaches.

The methods chosen in[8] for evaluation are all belonging to the latter family of techniques presented in Section 2. They include templates, histograms, correlograms, subspace projections, and several variations of each. Some details about each method are provided as follows:

- Histograms, similarly to,[4] were extracted from the entire image using YCbCr colorspace. Each channel was considered separately and later concatenated to create a single dimensional histogram. Optimal performance was achieved with 128 bins per channel.
- 3D Histograms were calculated similarly to 1D ones but the channels were considered jointly. Optimal performance was achieved with 16 bins per channel.
- Correlograms, as in,[9] were computed using 4 distance bins, and 2 to 5 color bins. Optimal performance was achieved with 4 color bins.
- Both histograms and correlograms are also used in their hand localized version (using 4, 16 and 128 bins). Separated areas of the silhouette were taken into account (corresponding to the head, shirt, and legs areas) and the three area histograms were concatenated to compose a signature.
- Template methods use a simple sum of squared distance between two images, as explained in.[1,10] Gaussian weights are also used in order to take more in emphasize the center of each image trying to reduce the effect of the background.

[‡]available for download at http://vision.soe.ucsc.edu

- Subspace methods were computed using principle component analysis directly on the training set, as in.[11] A simple reprojection of the data was done in the n principal components and compared using the sum of squared distance. Performance for this method was the worst and did not increase significantly for values of n higher than 64.

4.3 Results

The proposed probabilistic color histogram should be applied to the extracted bounding boxes of the subjects walking in fields of view of the camera network. Generally it is possible to distinguish two main colors or set of colors describing the subjects: one referring to the torso, the other one referring to the legs of the subjects. In order to obtain a full description of the people, we decided to use the color information provided by both the segments. The torsos and the legs of the subjects are automatically selected from the bounding boxes in the VIPeR dataset by extracting the middle and the bottom two-fifth parts of the images respectively. While the size of the provided bounding boxes is constant in the dataset, the position of the torsos and legs is not (due to variations of the viewpoints in the images). That means that most of the images will contain many pixels from the background that will add undesiderable variability to the evaluation of the PCHs. Moreover we can expect that this variability is higher in the segments referring to the legs since the people are recorded while walking. For this reason we analyze the performance of the system using both the segments.

A quite popular approach to evaluate the performance of a re–identification algorithm is to consider recognition as a ranking problem. In this framework a ranking is induced on the elements of the dataset and the probability that the correct match has a rank equal to or less than some values is plotted over the size of the test set. This performance metric is known as the cumulative matching characteristic (CMC) curve, which is analogous to the ROC curve for detection problems.

The VIPeR dataset consists of 632 pairs of images. To be consistent with the results shown in,[8] in order to facilitate the comparison, we split the database in two parts each consisting of 316 pairs of images. The first part was used as a training set to analyze the efficiency of the Euclidean and Bhattacharyya metrics in evaluating the similarity between histograms [§], while the second part is used as testing set. For sake of clarity:

$$Eucl\_distance(a, b) = \|a - b\| \tag{3}$$

$$Bhat\_distance(a, b) = \sum \sqrt{ab} \tag{4}$$

The results we obtained are depicted in Fig.8 which shows the CMC curves for both metrics. Additionally different weights were assigned to the torsos and the legs of the subjects. The parameters $a = 1, b = 0$ refer to the case of taking into account only the torsos of the subjects, while with $a = 1, b = 1$ both the distances on the torso and the legs are used for the results. It follows that in the case $a = 2, b = 1$ or $a = 3, b = 1$, the torso is weighted more than the legs.

We can observe how in all the cases the Bhattacharyya distance provides better results than the Euclidean distance. Moreover the best CMC curve is obtained weighting the torso twice than the legs, thus we decide to adopt the corresponding parameters in the test. It is worth pointing out that this result can not be considered general but strictly depending on the selected database and, mostly, on the body parts extraction method.

The results obtained by applying the defined PCH on the VIPeR dataset (the testing set) are shown in Fig.9. While the efficiency of the PCH is obvious from the image, a comparison with the state of the art systems is required to quantify the performance of the proposed color descriptor.

In Fig.10 the CMC curves relative to state of the art systems (left) and a magnified version of them (right) are reported. Both the images are extracted from.[8] For an easy comparison with the proposed probabilistic color histogram, in the image on the right the red dots refer to the performance of the PCH, as reported in Fig.9 and zoomed in the range [0 100]. We can observe how the proposed PCH overperforms the analyzed state of the art systems.

[§]Obviously, more complex distance metrics could be defined to evaluate the similarity between histograms.

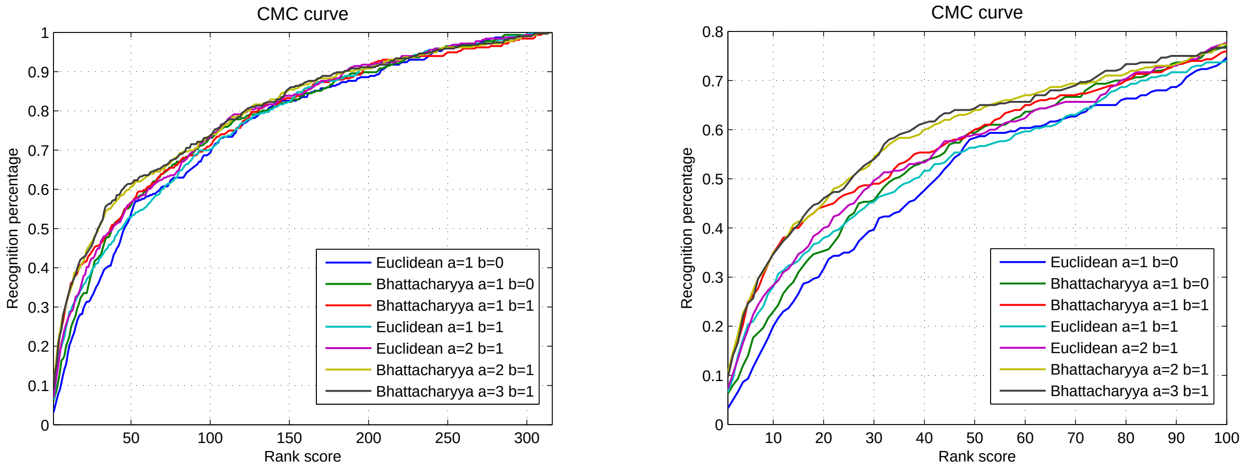


Figure 8. CMC curves for Euclidean and Bhattaracyya distances with different parameters (left) and a zoomed version of them (right).

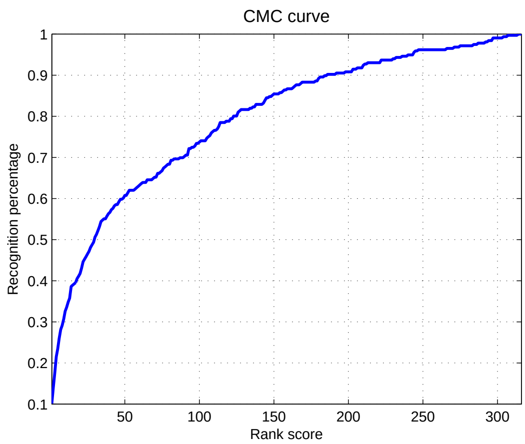


Figure 9. CMC curve obtained with the prposed PCH.

5. CONCLUSIONS

In this paper we have proposed a new color feature based on the definition of Probabilistic Color Histogram. A fuzzy color quantization based on a set of culture colors is first applied on selected segments from the bounding boxes and the PCH is then evaluated and used as a visual signature for re-identification. The experimental results are performed on a challenging dataset, providing 632 pairs of images seen from widely different viewpoints. It has been shown that the defined descriptor is effective at discriminating and re-identifying people across two different cameras regardless of the viewpoint change between the two views and outperforms state of the art appearance based techniques. The proposed approach can be integrated in a video-surveillance system and can be used to assist a human operator in the task of people re–identification.

Future works include the introduction of other features describing the people appearance and the definition of a multimodal system to fuse the information provided by all the features. A machine learning algorithm could then be used to find the best representation.

We would like to point out that the proposed color descriptor, thanks to its property to be robust to variation in illumination conditions and to viewpoint changes, can be exploited in several applications of image and video processing. A straightforward example is the use of PCH in image indexing and retrieval

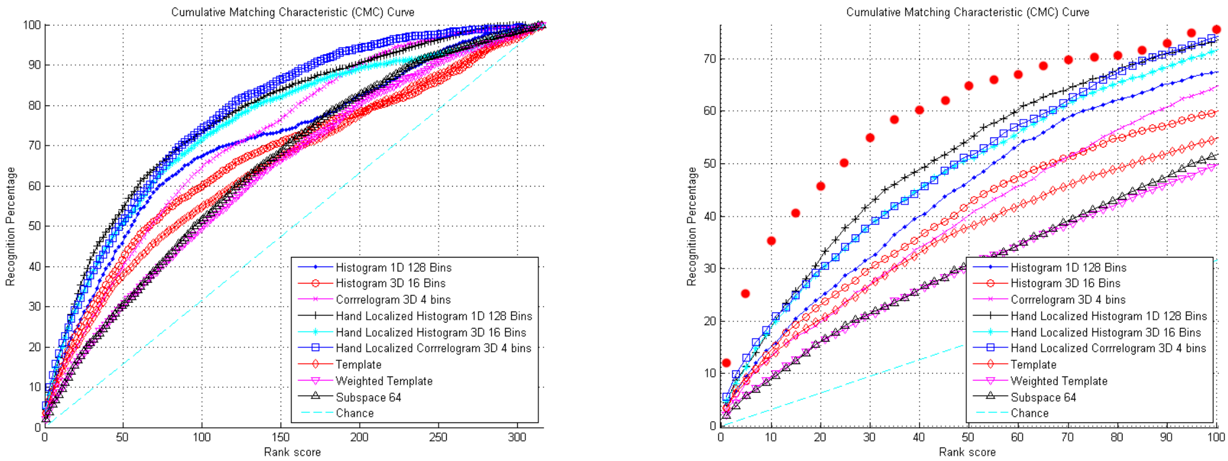


Figure 10. CMC curves relative to state of the art systems (left) and a zoomed version of them (right). Both the images are extracted from.[8] In the image on the right the red dots refer to the performance of the PCH, as reported in Fig.9.

instead of the more popular but not so effective histograms or color correlograms.

Acknowledgment

This work has been partially funded by the FR OSEO BIORAFALE and FR ANR VIDEO-ID projects.

REFERENCES

[1] Guo, Y., Hsu, S., Shan, Y., Sawhney, H., and Kumar, R., "Vehicle fingerprinting for reacquisition and tracking in videos," in [*CVPR '05: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 2*], 761–768, IEEE Computer Society, Washington, DC, USA (2005).

[2] Lee, K.-C., Ho, J., Yang, M.-H., and Kriegman, D., "Video-based face recognition using probabilistic appearance manifolds," in [*Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference*], **1**, I–313–I–320 vol.1 (June 2003).

[3] Gray, D. and Tao, H., "Viewpoint invariant pedestrian recognition with an ensemble of localized features," in [*ECCV '08: Proceedings of the 10th European Conference on Computer Vision*], 262–275, Springer-Verlag, Berlin, Heidelberg (2008).

[4] Javed, O. and Shah, M., "Tracking in multiple cameras with disjoint views," *Automated Multi-Camera Surveillance: Algorithms and Practice* , 1–26 (2008).

[5] Berlin, B., Birren, F., Gage, J., Gatto, J., Hoban, T., and Kuppers, H., "Basic color terms; their universality and evolution," (1993).

[6] Keller, J., GRAY, M., and GIVENS, J., "A Fuzzy K-Nearest Neighbor Algorithm ," *IEEE Transactions on Systems, Man, and Cybernetics* **15**(4), 581 (1985).

[7] D'Angelo, A. and Dugelay, J.-L., "A statistical approach to culture colors distribution in video sensors," in [*VPQM 2010, 5th International Workshop on Video Processing and Quality Metrics for Consumer Electronics, January 13-15, 2010, Scottsdale, AZ, USA*], (01 2010).

[8] Gray, D., Brennan, S., and Tao, H., "Evaluating Appearance Models for Recognition, Reacquisition, and Tracking," in [*10th IEEE International Workshop on PETS*], (2007).

[9] Huang, J., Kumar, S., Mitra, M., and Zhu, W., "Image indexing using color correlograms," (June 12 2001). US Patent 6,246,790.

[10] Shan, Y., Sawhney, H., and Kumar, R., "Vehicle identification between non-overlapping cameras without direct feature matching," in [*Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*], **1**, 378–385, IEEE (2005).

[11] Leclercq, T., Khoudour, L., Macaire, L., Postaire, J., and Flancquart, A., "Compact color video signature by principal component analysis," *Computer Vision and Graphics* , 814–819.