

Person re-identification by fuzzy space color histogram

Zong Jie Xiang · Qiren Chen · Yuncai Liu

Published online: 9 November 2012

© Springer Science+Business Media New York 2012

Abstract In this paper we propose a new feature for person re-identification with or without the full gallery of persons: the Fuzzy Space Color Histogram. This feature contains both the space and color information, and is characterized by fuzzy quantization. We can optionally integrate our feature with Fuzzy Foreground when accurate segmentation is unavailable. Intensive experiments on three typical datasets have demonstrated that our method achieves promising results in real time.

Keywords Re-identification · Camera network · Gallery · Fuzzy space color histogram

1 Introduction

Person Re-identification is a critical technique in surveillance systems that use a network of cameras. As one camera can only cover a limited field of view (FoV), multiple cameras are deployed, composing a network which records human activities. To automatically process the immense volume of data, algorithms are required to track people within one camera's FoV, as well as across multiple cameras. Assuming that intra-camera tracking is adequate, the problem is reduced to re-identification, i.e. associating two instances of the same person on different cameras at the same time (overlapping FoV) or at different times (non-overlapping FoV).

Re-identification is the process of discovering the correspondence between gallery images and probe images. With the full gallery available, wherein all appearing persons have been previously modeled and recorded in the gallery, re-identification can be solved as a ranking problem. Adversely, without the availability of the full gallery, we need to verify if the current appearing person is the given person of interest; in that case re-identification becomes a binary classification problem.

A histogram is one of the classic methods used for re-identification. However, original histograms lose all of the space information of image signals, and are sensitive to small shifts in color values. To overcome these issues, we attempt to develop a new type of histogram

Z. J. Xiang (✉) · Q. Chen · Y. Liu
Department of Automation, Shanghai Jiao Tong University, Shanghai, China
e-mail: zongzong_1984_111@hotmail.com

Q. Chen
e-mail: chenqiren1234@gmail.com

Y. Liu
e-mail: whomliu@sjtu.edu.cn

which incorporates space information and is characterized by soft quantization. In Gray's work [8], the authors recommend using a hand-localized histogram for person re-identification. The histogram incorporates space information, but it is rigid in quantization. In [5] a probabilistic color histogram is proposed, but the histogram does not contain space information. As a remedy, the authors use it for the torso and legs of the body. In this work, we propose a novel Fuzzy Space Color Histogram (FSCH) which incorporates both space and color information, and is characterized by soft quantization. We can optionally integrate Fuzzy Foreground into our feature when accurate segmentation is unavailable.

The rest of the paper is organized as follows: In Section 2, we review the related works. In Section 3 we analyze the two frameworks of re-identification application. We propose FSCH for re-identification and describe its mathematical insight in Section 4. In Section 5, we test our algorithm; and we conclude the work in Section 6.

2 Related works

Various methods have been proposed to solve the re-identification problem. The authors of [13, 15, 16] recover the topology of a camera network and estimate the duration of an individual's passage from one camera's FoV to another. By employing the topology information, we can significantly reduce the number of potential matching candidates and improve the re-identification accuracy. The Brightness Transfer Function (BTF) between cameras is useful when working with different levels of illumination. Javed et al. have proposed an online method to compute the BTF [14]. More recently, a learning method was proposed in [21] to obtain effective feature comparison functions, regardless of the feature used.

Other researchers focus on features. The authors of [3] re-identify persons according to their faces. Works [11] and [19] are based on the walking pattern of pedestrians. Wang et al. use a 1D signal to describe human contours, compress the walking patterns using Principle Component Analysis (PCA), and re-identify people by PCA parameters [19]. Gray et al. [8] compare several useful features such as histograms and correlograms [12], and recommend using a hand-localized histogram in person re-identification. In [5], a new feature called a probabilistic color histogram is proposed and achieves satisfactory results. The authors of [7] devise a triangulated model which records the spatial distribution of local features over time. They call it spatiotemporal appearance. Reference [9] presents a method which uses AdaBoost to select good features from a feature space, which was intuitively defined according to the problem. In [6], the authors develop a signature composed of the overall chromatic content, the spatial arrangement of colors, and recurrent local motifs. The three parts of the signature are weighted by exploiting symmetry and asymmetry perceptual principles. Researchers have also explored the usage of interest point descriptors such as SIFT and SURF. In [10] and [17], interest point descriptors are used and re-identification is solved as a voting problem. However, the voting mechanism for their methods [10, 17] requires the availability of the full gallery, and as such, cannot be applied when the full gallery is unavailable. A grid cascade of dense region descriptors has been proposed in [1]. In their paper, covariance matrices of features, and some descriptors such as SIFT and SURF have been evaluated. Unlike the methods mentioned above, the researchers of [2] establish a 3D model of pedestrians to achieve viewpoint invariance. Their 3D model is vertices-based, and the colors at the vertices are used as the features.

Although most of the mentioned features [2, 5–9, 19] can be applied without the full gallery, these papers have not discussed the re-identification strategy when the full gallery is unavailable. However, in application, the re-identification system usually works without the

full gallery. In this paper, we propose our feature, and construct and evaluate our re-identification system under both situations (with and without the full gallery).

3 Frameworks

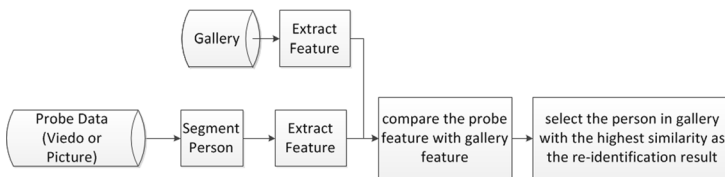
In this section, we discuss two frameworks (see Fig. 1) of the re-identification systems. The first framework is used when a full gallery of persons is available, and re-identification can be solved as a ranking problem. The system compares the probe person with each person in the gallery, and selects the person with the highest similarity in the gallery as the re-identification result. The requirement of the availability of the full gallery limits the usage of this framework in a relatively closed environment such as a prison, a laboratory, or a military restricted zone.

The second framework can be used in open environments such as streets, where full galleries are usually unavailable. An operator must select the person of interest by human-machine interaction. The re-identification system collects and records the features of this person. When the target person reappears, the system will re-identify and continue to track him. A typical application of this framework is to implement it when searching for a missing person in streets.

There is one main difference between framework 1 and framework 2. Assume that one probe person is very similar to a different person in the gallery. Framework 1 tends to reject him because the true corresponding person in the gallery tends to present even higher similarity, whereas framework 2 would probably consider the two persons to be the same because the similarity between them exceeds the similarity threshold. For this reason, framework 2 is much more challenging.

For feature comparison, different methods can be applied depending on their mathematical representation. In most cases, features take the form of vectors of equal length. Euclidean Distance, Bhattacharyya Distance, Correlation, etc. can be used as metrics to compute the similarity. In other cases, one feature may consist of several sub-features.

Framework 1: Re-identification with the full Gallery



Framework 2: Re-identification without Predefined Gallery, but with Operator

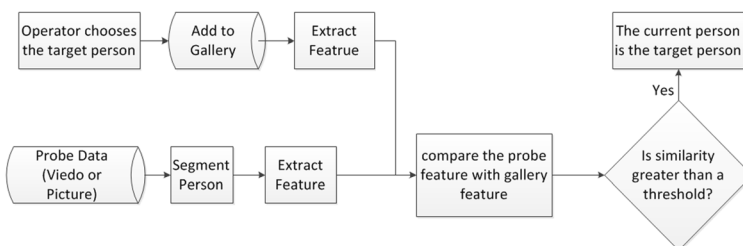


Fig. 1 The frameworks

Feature comparison is accomplished by comparing pairs of sub-features using the basic metrics mentioned above.

4 Features for re-identification

As shown in both of the frameworks in the previous section, feature extraction is the core step. In this section, we propose a new feature called the Fuzzy Space Color Histogram, and discuss its specific usage in person re-identification. We further integrate our feature with Fuzzy Foreground, so that it can be used when the segmentation results are unsatisfactory. Before moving on, we assume that a bounding box of the person is available (For videos, it can be obtained by background subtraction).

4.1 Fuzzy space color histogram

We combine a 3D color histogram with 2D space information (image coordinate (x,y)), resulting in a 5D space color histogram. We then obscure this 5D space color histogram. As is depicted in Fig. 2, we divide the space of this dimension into nine bins. The center of each bin marks the maximal membership degree. Other membership functions can also be applied, such as the Gaussian function, the trapezoid function, etc.

In our histogram, one variable usually belongs to two adjacent bins with different membership degrees (see Fig. 2), which is different from the ‘Winner Takes All’ strategy in the original histogram. The following is the mathematical explanation of our FSCH.

We assume a normalized 3D color histogram (see (1)) with each bin mapped to only one color value; thus the histogram contains 256^3 bins, reaching the maximum number of bins.

$$\sum_{R,G,B} p(R, G, B) = 1 \tag{1}$$

Here, $p(R,G,B)$ is the 3D color histogram.

Equation (2) is for the space color histogram, which is a 5D histogram. Similar to (1), we restrict it to a histogram with each bin mapped to only one color value or space value. This restriction is not a useful setup in application, but we only use it to derive our mathematical explanation.

$$\sum_{R,G,B,x,y} p(R, G, B, x, y) = 1 \tag{2}$$

The space color histogram can be expressed by (3).

$$p(R, G, B, x, y) = \begin{cases} 1/TotalPixels, & image(x, y) = [R, G, B] \\ 0, & rest \end{cases} \tag{3}$$

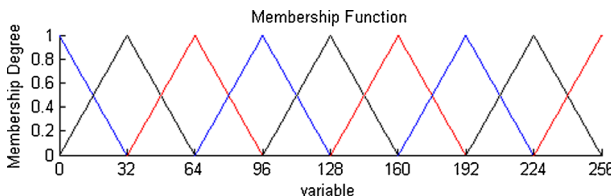


Fig. 2 The member function example in one dimension

TotalPixels is the total number of pixels in the image, or the masked area when using masks.

Equations (1–3) are used for the rigid histogram. We need to transform it into a fuzzy histogram. Here we define the unary membership function $w_i(x)$ as a function that satisfies:

- (1) $w_i(x) = w_i(-x)$
- (2) $w_i(0) = 1$
- (3) $w_i(x_1) \leq w_i(x_2)$, given $|x_1| > |x_2|$

When classifying a pixel into a bin, we weigh it in each dimension. This process can be expressed by (4).

$$\begin{aligned} \tilde{p}(R_C, G_C, B_C, x_c, y_c) &= \sum_{R,G,B,x,y} [p(R, G, B, x, y) \cdot w(R - R_C, G - G_C, B - B_C, x - x_c, y - y_c)] \end{aligned} \tag{4}$$

The subscript c denotes the bin center. The function w in (4) is the weight, representing the final membership degree. It is the product of 5 unary membership functions (see (5)).

$$\begin{aligned} w(R - R_C, G - G_C, B - B_C, x - x_c, y - y_c) &= w_1(R - R_C) \cdot w_2(G - G_C) \cdot w_3(B - B_C) \cdot w_4(x - x_c) \cdot w_5(y - y_c) \end{aligned} \tag{5}$$

Where the function w_1, w_2, w_3, w_4, w_5 are the unary membership functions. These functions satisfy (6).

$$w_1(R_c - R) = w_1(R - R_c) \tag{6}$$

From the above derivation, we can conclude that for each bin in FSCH, its value is the convolution of $p(R, G, B, x, y)$ and the weight function $w(R, G, B, x, y)$ at the bin center R_C, G_C, B_C, x_c, y_c . It yields the following, more compact equation (see (7)). However, when implementing the algorithm, we use (4) as the core, and (3) and (5) as the supplementaries.

$$\tilde{p}(R_c, G_c, B_c, x_c, y_c) = p(R, G, B, x, y) * w(R, G, B, x, y)|_{R_c, G_c, B_c, x_c, y_c} \tag{7}$$

The pseudo code for computing a FSCH is presented in Algorithm 1.

Algorithm 1: Computation of FSCH

```

Initialize our histogram: set all bins in FSCH to zero
for every pixel in the bounding box
  for each dimension
    compute the current pixel's membership degrees at the bins in this dimension
    record the bin index where the corresponding membership degree is not zero
    record the membership degree
  end
  for each combination of the recorded bins at each dimension
    calculate the bin index at the histogram
    compute the final membership degree
    add the final membership degree to the histogram bin
  end
end
normalize the histogram
    
```

Typically, in each dimension one pixel belongs to two neighboring bins with different membership degrees. Since our FSCH is a 5D histogram, one pixel would finally belong to 32 (2^5) bins with non-zero membership degrees.

Let $nPixels$ be the number of pixels. The number of dimensions equals 5. Each pixel belongs to two bins in each dimension. From the above analysis, $nPixels$ is the only unknown factor. Therefore, the computational complexity for computing FSCH is $O(nPixels)$, which is the same as the original histogram. However, FSCH indeed incurs more computational cost. For each pixel, our algorithm must perform 10 ($5*2$) computations of membership degrees plus 32 (2^5) computations of the final membership degrees. Nevertheless, it is still efficient; our test shows that computing a FSCH for a 70×25 image patch takes 1.14 milliseconds, and our codes have potential for optimization.

4.2 Usage in person re-identification

The details of feature extraction in person re-identification are as follows: For vertical dimension y , we compute the centroid $y_{centroid}$ and standard deviation σ . This dimension is divided into three bins, whose bin centers are $y_{centroid} - \sigma$, $y_{centroid}$, and $y_{centroid} + \sigma$ respectively. In doing so, we reduce the quantization error in space, which commonly appears in localized histograms due to the inaccurate segmentation of body parts.

We ignore horizontal dimension x and thus FSCH is degenerated into a 4D histogram. However, it does not conflict with our conception of FSCH. It is equal to the arrangement of a single bin covering the whole space with a flat (constant) membership function. In doing so, our signature acquires an increased ability in conquering viewpoint changes. In our initial proposition, FSCH is formulated as a 5D histogram, whereas our experiment shows that removing the horizontal dimension x would benefit its specific usage in person re-identification, wherein people exhibit substantial variation in viewpoints. In other possible applications, a 5D FSCH may outperform a 4D one.

For feature comparison, we use Correlation as the metric. Let H_1 and H_2 be the feature vectors; their similarity can be computed by (8) and (9).

$$similarity(H_1, H_2) = \frac{\sum_I (H_1(I) - \bar{H}_1)(H_2(I) - \bar{H}_2)}{\sqrt{\sum_I (H_1(I) - \bar{H}_1)^2 \sum_I (H_2(I) - \bar{H}_2)^2}} \quad (8)$$

$$\bar{H}_k = \frac{1}{N} \sum_J H_k(J) \quad (9)$$

Where N is the length of the feature vector.

Although the feature is not completely viewpoint and illumination invariant, the experiments show that its performance is comparable to state of the art work, and with much lower computational cost.

Our feature is stable to the poses: standing, walking and running. To address other poses and occlusion, one should take full advantage of the intra-camera tracking module. From the moment a person enters into the FoV of one camera to the moment he leaves the FoV, the pose change (except for standing, walking, and running) and occlusion do not usually last throughout the entire frame sequence. Re-identification can be done according to the frames devoid of pose change and occlusion.

4.3 Integration with fuzzy foreground

Since accurate segmentation is always difficult, we attempt to incorporate Fuzzy Foreground into our histogram. Equation (5) can be replaced by (10).

$$\begin{aligned} & w(R - R_C, G - G_C, B - B_C, x - x_c, y - y_c) \\ &= w_1(R - R_C) \cdot w_2(G - G_C) \cdot w_3(B - B_C) \cdot w_4(x - x_c) \cdot w_5(y - y_c) \\ &\quad \cdot w_{\text{foreground}}(x, y) \end{aligned} \quad (10)$$

Here $w_{\text{foreground}}(x, y)$ is the foreground likelihood at $\text{pixel}(x, y)$.

When processing videos, we segment the foreground using the Gaussian Mixture Model (GMM) and obtain the bounding box for each person. For each pixel within the bounding box, we compute its probability of being in the foreground, instead of classifying a pixel rigidly as either foreground or background. Like what was done in [20], we classify image pixels into super-pixels (see Fig. 3a). Super-pixels [18] are regions which exhibit similar properties. We obtain the super-pixels using the software Edge Detection and Image Segmentation (EDISON) System [4], available at <http://coewww.rutgers.edu/riul/research/code/EDISON/doc/help.html>. For each super-pixel, we compute the difference between the image and the background (see Fig. 3b) using color histogram comparison. We use this difference to determine the foreground likelihood of the super-pixel (see Fig. 3c).

5 Experiments

We evaluate our re-identification method on three datasets. The first dataset is comprised of multi-camera videos. We implement the system under framework 2. The other two datasets are the VIPeR dataset and the SARC3D dataset. We implement our system under framework 1. The methods we use for comparison include: FSCH (proposed), 3D color histogram, space color histogram, and fuzzy color histogram. Apart from the above methods, we also compare our work with Symmetry-Driven Accumulation of Local Features (SDALF) [6] on all three datasets, and SARC3D [2] on the SARC3D dataset. The code for SDALF is publicly available at <http://www.lorisbazzani.info/code-datasets/sdalf-descriptor/>. All the experiments are carried out using an Intel CPU core 2.67 GHz.

The Receiver Operating Characteristic (ROC) curve and the Cumulative Match Characteristic (CMC) curve are employed to evaluate the performance of the re-identification system. For both curves, the Area Under Curve (AUC) is used as the numerical index. A

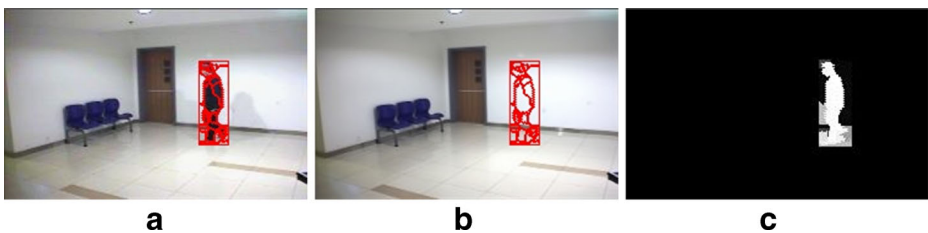


Fig. 3 **a** shows the super-pixels. The boundary of every super-pixel is depicted; **b** shows the same super-pixels in the background; **c** is the foreground likelihood computed from (a) and (b), with higher likelihood indicated by higher brightness



Fig. 4 The cameras

larger AUC indicates a better performance. The ROC curve is used when we do not have a full gallery. It shows the relationship between a false positive rate and a true positive rate with respect to the similarity threshold. The CMC curve is used when the full gallery is available. It depicts the relationship between the accuracy and the threshold of rank.

5.1 Experiment on multi-camera videos

Our multi-camera dataset consists of six videos (see Fig. 4) taken simultaneously at different locations in a building. Cameras 1,2,4 and 5 have side viewpoints of pedestrians, camera 3 has a back and front view and camera 6 has a 45°view vertically and horizontally. In total, nine people (see Fig. 5) cross the FoVs of the six cameras. The size of these images is 248×156 , and the videos are recorded at 12 frames per second.

Because the only materials we have are the six videos and we do not have a gallery, it is practical to implement our re-identification method under framework 2. We present our result in ROC curves (see Fig. 6). AUC is used to evaluate the performance (see Table 1). We can conclude from Fig. 6a–d and from Table 1 that: (1) FSCH is the best among the histogram-based methods under all configurations; (2) The algorithm works better in RGB space than in HSV space, due to the minimal illumination difference between cameras.; (3) Fuzzy Foreground has a positive effect on person re-identification. More experiments should be conducted to verify if the above results recur in other circumstances.

In Fig. 6e, our proposed method (in the optimal configuration: RGB space, Fuzzy Foreground) is compared to the state of the art work: SDALF. The AUC of our method is 0.931 in RGB space with Fuzzy Foreground, which is slightly higher than the AUC of 0.923 of SDALF in HSV space



Fig. 5 The appearances of the 9 individuals wandering through the camera networks

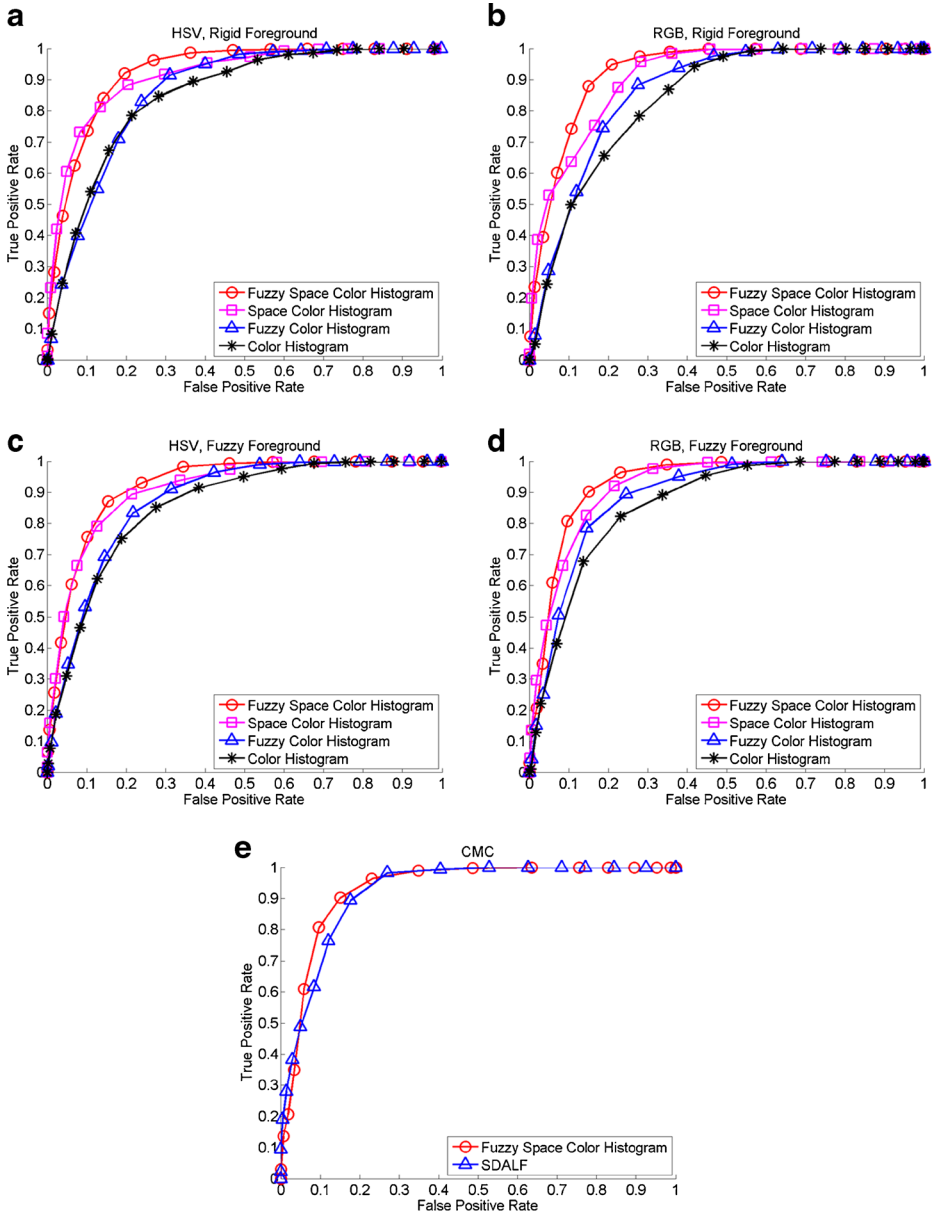


Fig. 6 ROC curves: **a** All histograms are computed in HSV space with rigid foreground; **b** in RGB space with rigid foreground; **c** in HSV space with Fuzzy Foreground; **d** in RGB space with Fuzzy Foreground; **e** comparison between the proposed feature (in the optimal configuration) and SDALF

with rigid foreground. Though we have only tested SDALF in HSV space and with rigid foreground, SDALF should optimally be used in HSV space according to the description in [6], and the authors have not discussed the usage of Fuzzy Foreground with their feature. With regard to the computational cost, our method processes these videos in real time while SDALF must spend approximately 2 hours to process 3.5 minutes of the same video footage.

Table 1 The AUC under different configurations (video dataset)

configuration	3D Color Histogram	Fuzzy Color Histogram	Space Color Histogram	FSCH	SDALF
HSV, Rigid Foreground	0.831	0.858	0.898	0.904	0.923
RGB, Rigid Foreground	0.837	0.861	0.907	0.926	/
HSV, Fuzzy Foreground	0.855	0.874	0.908	0.920	/
RGB, Fuzzy Foreground	0.864	0.889	0.919	0.931	/

5.2 Experiment on the VIPeR dataset

The VIPeR dataset contains 632 pairs of images (one pair for each person), taken from different viewpoints (see Fig. 7). We do not use Fuzzy Foreground since the foreground masks of these images are released with the codes of SDALF. We implement our re-identification method under framework 1. In this framework, the CMC curve is more suitable for evaluating the method. We use AUC to numerically compare the performance of the methods.

Figure 8 shows the CMC curves under different configurations. The AUCs are listed in Table 2 (area normalized). The results show that: (1) FSCH performs best among the histogram-based methods in all configurations. However, in RGB space, FSCH is close to the space color histogram, and the fuzzy color histogram is almost the same as the color histogram. This is due to the fact that the illumination variation (see Fig. 7) is so great in this dataset, that it overwhelms the quantization error which can be easily removed by a fuzzy histogram. (2) Our system works better in HSV space than in RGB space. This is due to the large illumination variation in the VIPeR dataset.

Apart from histogram-based features, it is also important to compare our feature with state of the art work. For fair comparison, we use exactly the same rigid foreground masks



Fig. 7 Some examples from the VIPeR dataset. Each column is one of 632 same-person example pairs. Note the wide range of viewpoint and illumination changes. This figure is extracted from [8]

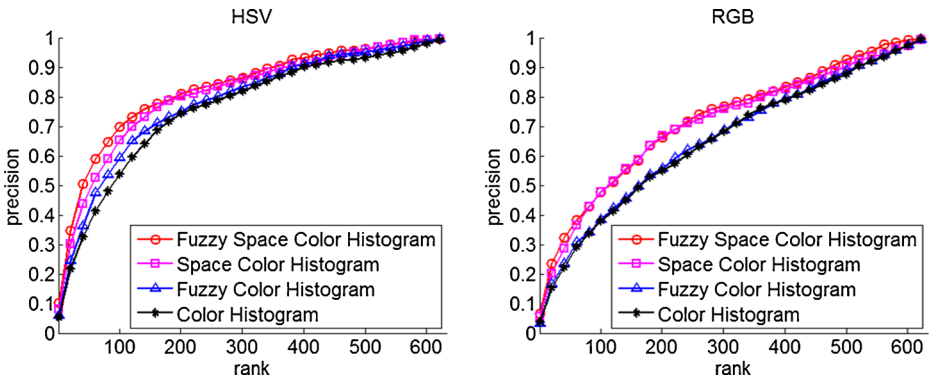


Fig. 8 CMC curves: **a** All histograms are computed in HSV space; **b** in RGB space

employed by the author of [6]. The CMC curves for our feature (in HSV space) and for SDALF are depicted in Fig. 9. We can see that the AUC of SDALF is slightly higher than the AUC of our method. However, the computational cost of the extraction and matching of our feature is significantly lower than that of SDALF. It takes about 4940 seconds to extract SDALF for all 1264 pictures in the VIPeR dataset, whereas it takes only 4.2 seconds to extract FSCH for these images. For SDALF it takes 5380 seconds to compare all 399424 (632^2) pairs of images, whereas for FSCH, it takes only 0.6 seconds.

5.3 Experiment on the SARC3D dataset

Finally, we test our method on the SARC3D dataset [2]. The dataset contains images of 50 individuals and their masks. For each person, there are four images taken from four different viewpoints (back, front, left, right). The re-identification algorithm is implemented under framework 1. We choose images from one viewpoint as probe images, and images from the other three viewpoints as the gallery. For example, if a probe image is from the front view, the gallery is composed of images of all persons from the back, left and right views. We rank all persons in the gallery according to the similarities between the probe image and all persons in the gallery in a descending order. In our experiment, we remove six persons from the dataset since the sizes of their masks are different from those of the original images, thus we use 44 persons for re-identification. All of the tests are conducted in RGB space. HSV color space is not applied because too many persons are wearing black clothing, which produces huge disruptions in the Hue channel. If one color is close to grey, its hue value becomes unstable. For example, the color RGB(24,25,26) is equivalent to HSV(105,20, 26), and a similar color RGB(26,24,25) is equivalent to HSV(165,20, 26). The two colors are close to each other, whereas their hue values vary enormously.

Table 2 The AUC under different configurations (VIPeR dataset)

Configuration	3D Color Histogram	Fuzzy Color Histogram	Space Color Histogram	FSCH	SDALF
HSV, Rigid Foreground	0.764	0.786	0.812	0.830	0.860
RGB, Rigid Foreground	0.657	0.660	0.713	0.725	/

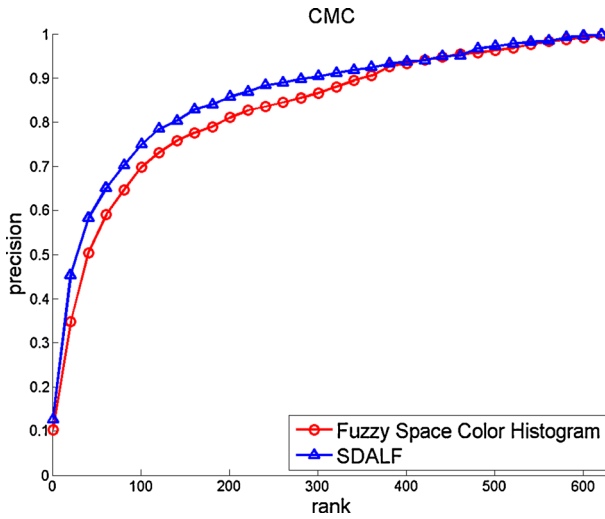


Fig. 9 Comparison with SDALF

Figure 10 shows the CMC curves. All curves are zoomed in to get a clear view. Table 3 shows the AUC results. Although in one test (probe image in right view), FSCH is

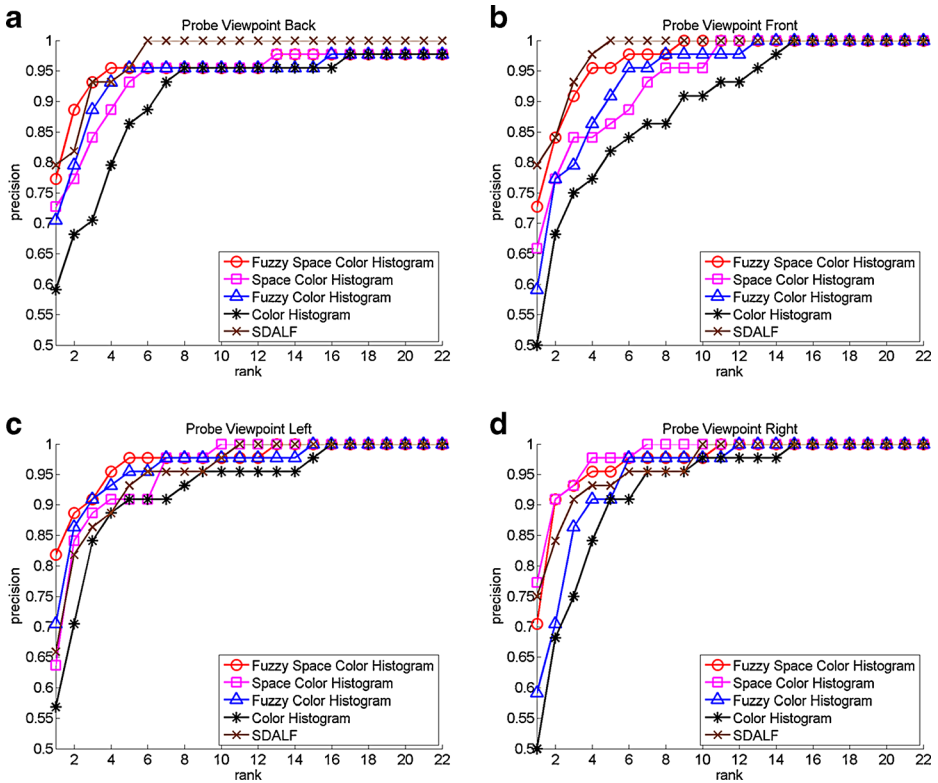


Fig. 10 CMC curves: a probe image is from the back view; b front view; c left view; d right view

Table 3 AUC when the probe image is from a specific viewpoint

Probe Image Viewpoint	3D Color Histogram	Fuzzy Color Histogram	Space Color Histogram	FSCH	SDALF
back	0.956	0.972	0.970	0.978	0.989
front	0.952	0.975	0.972	0.987	0.991
left	0.966	0.983	0.981	0.987	0.979
right	0.967	0.977	0.992	0.988	0.983
average	0.960	0.976	0.978	0.985	0.986

outperformed by the space color histogram, according to the average AUCs presented in Table 3 we can still reasonably conclude that FSCH is superior to the other histogram-based methods evaluated here. Figure 11 shows some examples of re-identification results.

The detailed comparison between our results and those of SDALF is shown in Fig. 10 and Table 3. In Fig. 12a we present the average CMC curves of our method and of SDALF. We can see that the overall performance of our feature is comparable to SDALF. We spend less than 1 second to extract all 176 (44*4) features and compare them using our method, whereas it takes 534 seconds to extract these features and 85 seconds to compare them using SDALF.

We also compare our results with SARC3D [2]. For each person, SARC3D uses two images (viewpoints) to create a probe 3D model, and the other two images (viewpoints) to establish a gallery model, whereas we use one viewpoint as the probe image, and the other three viewpoints as the gallery. It is difficult to say which configuration would be easier for re-identification. Figure 12b shows the comparison

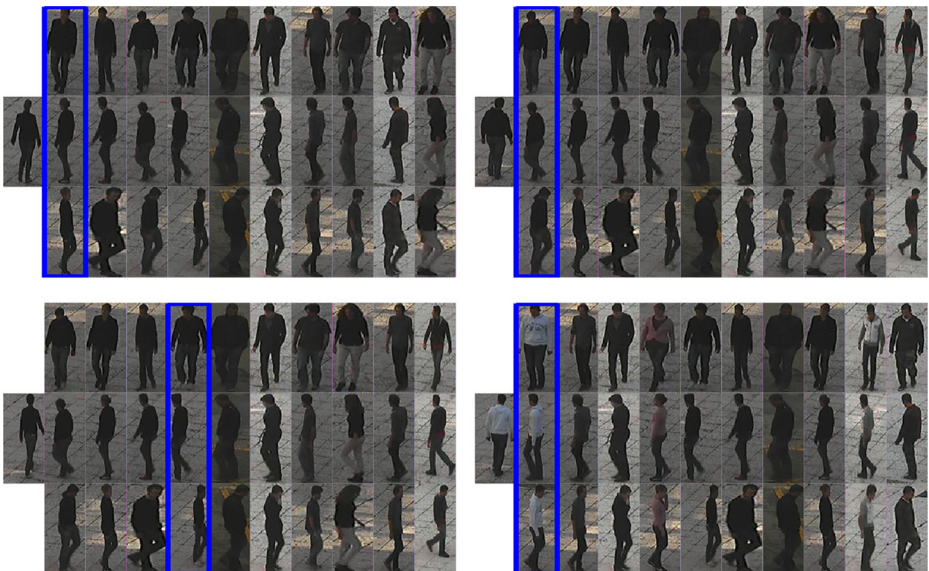


Fig. 11 These groups of images show 4 examples of the re-identification results. In each group, the leftmost person is the probe person from the back view; on its right are the 10 most similar persons in the gallery in the other three views, descending in similarity from left to right. The rectangles mark the correct correspondences

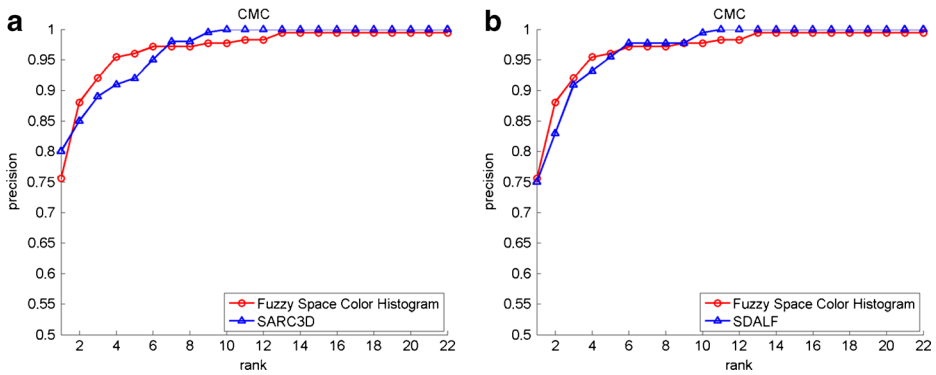


Fig. 12 a comparison between FSCH and SDALF; b comparison between FSCH and SARC3D

between our average CMC curve and that of SARC3D. It suggests that our method is comparable to SARC3D. An obvious disadvantage of SARC3D is that it requires at least two viewpoints of each person to build a 3D model. However, the paper [2] has not discussed the method to automatically associate the two different viewpoints of the same person before creating the 3D model.

5.4 Comments on all experiments

Our extensive experiments have shown that our feature is comparable to the state of the art work SDALF in its accuracy. Furthermore, the efficiency of our feature far surpasses that of SDALF, as it can be computed and compared in real time, resulting in a computational cost thousands of times less than that of SDALF.

However, there still remain several issues to be addressed. Firstly, is Fuzzy Foreground a useful element in the process of re-identification? Generally, Fuzzy Foreground promotes re-identification only when it describes the real foreground better than rigid foreground, as is the case in the experiment in Section 5.1, where Fuzzy Foreground is obtained according to the difference between the image and the background, and rigid foreground is insufficient. In the experiments in Section 5.2 and 5.3, rigid foregrounds are accurate; hence Fuzzy Foregrounds can not compete with the accurate rigid foregrounds.

It can also be concluded from the experiments that RGB is better for our video dataset, and for the VIPeR dataset HSV is better. The advantage of HSV space is its capacity to compensate for illumination differences, whereas its disadvantage is its inability to manage the color grey as the channel Hue becomes interrupted and impedes the performance of the feature. Therefore, when the illumination difference is more significant than the presence of the color grey, HSV space would be superior. In any other case, RGB space is suggested. However, for videos, there would be no need to use HSV space, since BTF can be learned [14] from the occurrences of pedestrians.

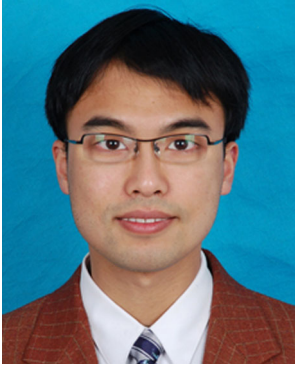
Furthermore, though the experiments on the VIPeR and SARC3D datasets are both evaluated by the CMC curve, their AUCs vary substantially from 0.830 to 0.985. This is because in the VIPeR dataset the illumination differences are significantly larger than those in the SARC3D dataset.

6 Conclusions

We propose FSCH with Fuzzy Foreground (optionally) as a feature to re-identify persons with or without the availability of the full gallery. The results of our intensive experiments conducted under framework 1 and framework 2 verify that the performance of our feature is comparable to the SDALF [6] and SARC3D models [2]. The computation and matching of the proposed feature are extremely efficient and the overall re-identification system can run in real time. Thus, due to its efficiency and comparability to state of the art work, it can be concluded that our proposed feature, FSCH, would be superior to current methods in re-identifying persons across multiple cameras.

References

1. Alahi A, Vanderghelynst P, Bierlaire M, Kunt M (2010) Cascade of descriptors to detect and track objects across any network of cameras. *Comput Vis Image Underst* 114:624–640
2. Baltieri D, Vezzani R, Cucchiara R (2011) SARC3d: a new 3d body model for people tracking and re-identification. *Int Conf Image Anal Process*
3. Bauml M, Bernardin K, Fischer M, Ekenel H, Stiefelhagen R (2010) Multi-pose face recognition for person retrieval in camera networks. *IEEE Int Conf Adv Video Signal Based Surveill*
4. Comanicu D, Meer P (2002) Mean shift: A robust approach toward feature space analysis. *IEEE Trans Pattern Anal Mach Intell* 24:603–619
5. D'Angelo A, Dugelay J (2011) People re-identification in camera networks based on probabilistic color histograms. *Proc of Electron Imaging*
6. Farenzena M, Bazzani L, Perina A, Murino V, Cristani M (2010) Person re-identification by symmetry-driven accumulation of local features. *IEEE Conf Comput Vis Pattern Recog*
7. Gheissari N, Sebastian T, Hartley R (2006) Person reidentification using spatiotemporal appearance. *IEEE Conf Comput Vis Pattern Recog*
8. Gray D, Brennan S, Tao H, (2007) Evaluating appearance models for recognition, reacquisition, and tracking. *IEEE Int Work Perform Eval Tracking Surveill*
9. Gray D, Tao H (2008) Viewpoint invariant pedestrian recognition with an ensemble of localized features. *Eur Conf Comput Vis*
10. Hamdoun O, Moutarde F, Stanculescu B, Steux B (2008) Person re-identification in multi-camera system by signature based on interest point descriptors collected on short video sequences. *ACM/IEEE Int Conf Distrib Smart Cameras*
11. Havasi L, Szlavik Z, Sziranyi T (2005) Eigenwalks: Walk detection and biometrics from symmetry patterns. *IEEE Int Conf Image Process*
12. Huang J, Kumar S, Mitra M, Zhu W, Zabih R (1999) Spatial color indexing and applications. *Int J Comput Vis* 35:245–268
13. Javed O, Shafique K, Rasheed Z, Shah M (2008) Modeling inter-camera space–time and appearance relationships for tracking across non-overlapping views. *Comp Vision Image Underst* 109:146–162
14. Javed O, Shafique K, Shah M (2005) Appearance modeling for tracking in multiple non-overlapping cameras. *IEEE Conf Comput Vis Pattern Recognit*
15. Loy C, Xiang T, Gong S (2010) Time-delayed correlation analysis for multi-camera activity understanding. *Int J Comput Vis* 90:106–129
16. Makris D, Ellis T, Black T (2004) Bridging the gaps between cameras. *IEEE Conf Comput Vis Pattern Recognit*
17. Oliveira I, Pio J (2009) People reidentification in a camera network. *IEEE International Conference on Depend Auton Secure Comput*
18. Ren X, Malik J (2003) Learning a classification model for segmentation. *Int Conf Comput Vis*
19. Wang L, Tan T, Ning H, Hu W (2003) Silhouette analysis-based gait recognition for human identification. *IEEE Trans Pattern Anal Machine Intell* 25:1505–1518
20. Zheng E, Chen Q, Yang X, Liu Y (2009) Robust 3d modeling from silhouette cues. *IEEE Conf Acoust Speech Signal Process*
21. Zheng WS, Gong S, Xiang T (2011) Person re-identification by probabilistic relative distance comparison. *IEEE Conf Comput Vis Pattern Recog*



Zong Jie Xiang received his MS degree at 2009 in Shanghai University, China. He is currently a PhD candidate at the Institute of Image Processing and Pattern Recognition at Shanghai Jiao Tong University, Shanghai, China. His research interests include pattern recognition, computer vision.



Qiren Chen is currently a college student at Shanghai Jiao Tong University. He takes an internship at the Institute of Image Processing and Pattern Recognition of Shanghai Jiao Tong University.



Yuncai Liu received his PhD in the Department of Electrical and Computer Science Engineering from the University of Illinois at Urbana-Champaign in 1990 and worked as an associate researcher at the Beckman Institute of Science and Technology from 1990 to 1991. Since 1991, he was a system consultant and then chief consultant of research at Sumitomo Electric Industries, Ltd., Japan. In October 2000, he joined Shanghai Jiao Tong University as a distinguished professor. His research interests are in image processing and computer vision, especially in motion estimation, feature detection and matching, and image registration. He also has made great progress in the research of intelligent transportation systems.