Research Article

Reidentification of Persons Using Clothing Features in Real-Life Video

Guodong Zhang,[1] Peilin Jiang,[2] Kazuyuki Matsumoto,[1] Minoru Yoshida,[1] and Kenji Kita[1]

[1]*Faculty of Engineering, Tokushima University, Tokushima 7708506, Japan*
[2]*Xian Jiao Tong University, No. 28, Xianning West Road, Xian, China*

Correspondence should be addressed to Guodong Zhang; zhang-g@hotmail.co.jp

Academic Editor: Qiushi Zhao

Person reidentification, which aims to track people across nonoverlapping cameras, is a fundamental task in automated video processing. Moving people often appear differently when viewed from different nonoverlapping cameras because of differences in illumination, pose, and camera properties. The color histogram is a global feature of an object that can be used for identification. This histogram describes the distribution of all colors on the object. However, the use of color histograms has two disadvantages. First, colors change differently under different lighting and at different angles. Second, traditional color histograms lack spatial information. We used a perception-based color space to solve the illumination problem of traditional histograms. We also used the spatial pyramid matching (SPM) model to improve the image spatial information in color histograms. Finally, we used the Gaussian mixture model (GMM) to show features for person reidentification, because the main color feature of GMM is more adaptable for scene changes, and improve the stability of the retrieved results for different color spaces in various scenes. Through a series of experiments, we found the relationships of different features that impact person reidentification.

1. Introduction

As public security technology has become increasingly intelligent, surveillance cameras have been set up in public places such as airports and supermarkets. These cameras provide huge amounts of nonoverlapping video data. It is often necessary to track an object or person of interest that appears on video from multiple cameras under different illumination conditions [1–3]. When searching for moving people in surveillance video data, object retrieval systems for intelligent video surveillance experience the following problems.

(1) Object retrieval results in video surveillance depend on motion segmentation and video analysis. Digital video is a series of images, constituted by frames that contain rich information. If an image frame contains moving objects, then object retrieval detection can be used to segment a moving target [4]. Object retrieval results depend on the object segmentation. If video analysis cannot separate the foreground and moving objects, the target object cannot be retrieved from the many irrelevant foreground objects. A good object retrieval system should adapt to various levels of video quality for foreground detection, which could eliminate unrelated objects and retrieve the target [5].

(2) Specific object retrieval in video surveillance faces technical limitations. The moving objects of interest in surveillance video are often persons and cars. Facial features are the most distinctive elements for person recognition, and relatively mature methods are available for this process. However, low camera resolution often makes it difficult to extract perceivable information about facial expression [6]. The mature technology of video object retrieval based on facial features should receive more technical exploration.

(3) External factors greatly influence objects appearance under video surveillance. A robust object retrieval system should be able to compensate for the following factors.

(i) Person pose variation: a moving person may have arbitrary poses (Figure 1(a)).

Figure 1: Images showing the same person in different camera views: (a) pose change, (b) illumination change, (c) occlusion, and (d) low resolution.

  (ii) Varying illumination conditions: illumination conditions usually differ between camera views (Figure 1(b)).
 (iii) Occlusion: a person body parts may be occluded by other subjects, such as a carried bag, in one camera view (Figure 1(c)).
 (iv) Low image resolution: due to surveillance camera performance, images of a moving person often have low resolution (Figure 1(d)).

The color histogram is a tool used to describe the color composition of an image [7]. The histogram shows the appearance of different colors and the number of pixels for each color in an image. Colors possess better immunity to the noise jamming of images and are robust against image degradation and scaling. We selected a global color approach to body features for person reidentification in surveillance video. Extracting the color information of the person makes the method clear and simple. Because color statistic features lose information about color spatial distribution, we combined this approach with the spatial pyramid matching (SPM) model. We tested our method in the RGB, HSV, and UVW color spaces using real video images. We present related work on person reidentification and feature analysis in Section 2. We offer details on our proposed method in Section 3. We report and discuss the experimental results in Section 4, and we give conclusions and suggestions for future work in Section 5.

2. Related Works

For the past few years, object retrieval techniques using content-based video retrieval have received significant theoretical and technological support. Many researchers have examined person reidentification, and the related literature is extensive [8, 9]. This section discusses feature modeling and effective matching strategies, which are important methods for person reidentification.

2.1. Color Feature. Color features are one of the low-level feature types that have been widely used in content-based image retrieval (CBIR). Compared with other features, color exhibits little dependence on image rotation, translation,

scale change, and even the shape change. Color is thus thought of as almost independent of the images dimensions, direction, and view angles. Most representations in previous approaches are based on appearance. Gray and Tao [10] used a similarity function that was trained from a set of data. These authors focused on the problems of unknown viewpoint and pose. The method is robust to viewpoint change because it is based on the ensemble of localized features (ELF). Farenzena et al. [11] presented an appearance-based method based on the localization of perceptually relevant human parts. The information features contain three parts: overall chromatic content, the spatial arrangement of colors into stable regions, and the presence of recurrent local motifs with high entropy. The method is robust to pose, viewpoint, and illumination variations. Zhao et al. [12] transformed person reidentification into a distance learning problem. Using the relative distance comparison model to compute the distance of a pair of views, these authors considered a likely true match pair to have a smaller distance than that of a wrong match pair. These authors also used a new relative distance comparison model to measure the distance between pairs of person images and judge the pairs of true matches and wrong matches. Angela et al. proposed a new feature based on the definition of the probabilistic color histogram and trained fuzzy k-nearest neighbors (KNN) classifier based on an ad hoc dataset. The method is effective at discriminating and reidentifying people across two different video cameras regardless of viewpoint change. Metternich et al. [13] used a global color histogram and shape information to track people in real-life surveillance data, finding that the appearance of the subject impacted the tracking results. These authors also focused on the performance of matching techniques over cameras with different fields of view.

2.2. Metric Learning. Hirzer et al. [14] focused the matching method of metric learning on person reidentification. These authors accomplished metric learning from pairs of samples from different cameras. The method benefits from the advantages of metric learning and reduces the required computational effort. Good performance can be achieved even using less color and texture information. Khedher et al. [15] proposed a new automatic statistical method that could accept and reject SURF correspondence based on the
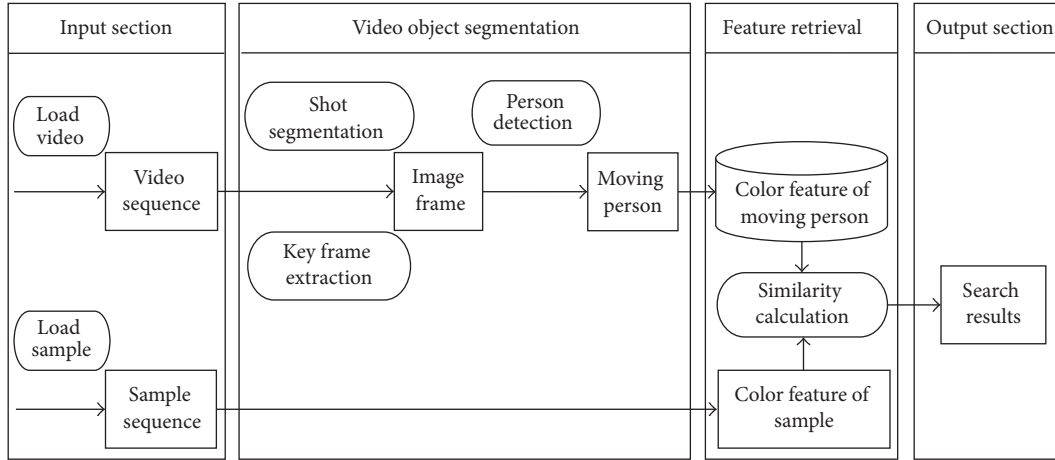
FIGURE 2: Overview of the system.

likelihood ratio of two Gaussian mixed models (GMMs) learned on a reference set. The method does not need to select the matching SURF pairs by empirical means. Instead, interest point matching over whole video sequences is used to judge the person identity. Matsukawa et al. [16] focused on the problem of overfitting and proposed a discriminative accumulation method of local histograms for person reidentification. The proposed method jointly learns pairs of a weight map for the accumulations and employs a distance metric that emphasizes discriminative histogram dimensions. This method can achieve better reidentification accuracy than other typical metric learning methods on various sizes of datasets.

3. System Description

3.1. An Overview of the Proposed System. The techniques of moving person retrieval information from a video database include shot segmentation, person detection, scene segmentation, feature extraction, and similarity calculation. As shown in Figure 2, shot segmentation refers to automatically segmenting video clips into shots as the basic unit for indexing. One second of video contains about 20–30 video frames, and neighboring frames are very similar to each other. There is no need to perform retrieval and matching for each frame, and frame differentiation is used to detect and extract the moving person. Frame differentiation relies on the change of pixel value between neighboring key frames. A change value greater than the established threshold value marks the pixel position of the moving person. This step is important in video parsing and directly affects the effectiveness of moving person retrieval.

The measurement method for similarity calculation influences the results ranking of object retrieval. Essentially, image similarity calculation computes the content of feature vectors from the objects. Each feature attribute selection can employ a different similarity computing method [17]. Frequently, image features are extracted in the form of feature vectors that can be regarded as points in multidimensional space.

The most common similarity measure method uses the distance between two spots in feature space. We also use distance measurement and correlativity calculation to scale the comparability between images.

Our proposed method is presented in Figure 3. We use traditional histogram and SPM histogram to retrieve the object. The traditional histogram method contains three parts, the color histogram feature extraction, color histogram distance computing, and outputting. The difference between SPM histogram and traditional histogram is the histogram distance computing part. The sample image and matching image are segmented into three parts, the upper, middle, and lower part. The three parts then separately computed the color histogram distance and use average distance to evaluate the results. Then the system uses GMM model to filter the top 20 results, extracts the GMM main color feature, and computes the similarity of them. Finally, the system outputs the rank of top 10 results.

3.2. Perception-Based Color Space Histogram Feature. Computations in the RGB and HSV color spaces cannot solve the problem of background illumination sensitivity. The color spaces always affect the computing accuracy of the color histogram [18]. We attempted to use perception-based color space, which exhibits good performance in image processing [19]. As the name suggests, the perception-based color space associated metric approximates perceived distances and color displacements, capturing relationships that are robust to spectral changes in illumination [20]. RGB color space can be transformed to perception-based color space through the following steps.

RGB color space can be transformed to perception-based color space through the following steps.

(1) Transform RGB to XYZ color space using the following formula (1):

$$
\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \frac{1}{0.177} \begin{pmatrix} 0.49 & 0.361 & 0.20 \\ 0.177 & 0.0812 & 0.011 \\ 0.00 & 0.01 & 0.99 \end{pmatrix} \begin{bmatrix} G(R) \\ G(G) \\ G(B) \end{bmatrix}, \quad (1)
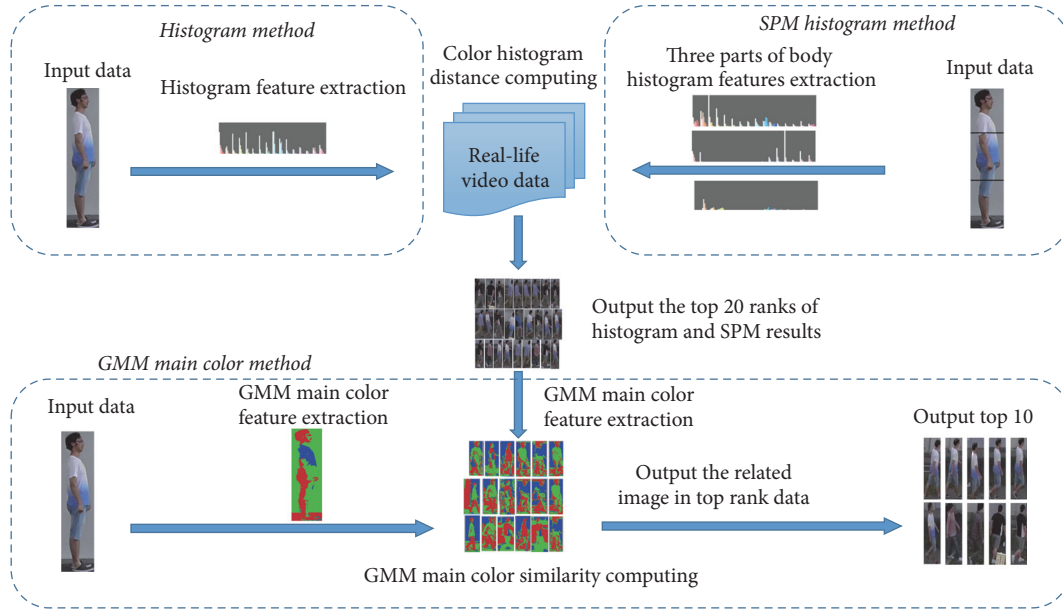$$

FIGURE 3: Overview of proposed method.

where $G()$ is the gamma correction function and equals 2.0. The gamma correction function addresses color distortion and rediscovers the real environment to a certain extent.

(2) Transform XYZ to UVW color space. In UVW color space, the influence of lighting conditions is simulated by the tristimulus multiplication values and scale factor, as shown in the following formula (2):

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \longrightarrow \begin{bmatrix} U \\ V \\ W \end{bmatrix} = B^{-1}DB \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}, \tag{2}$$

where D is a diagonal matrix, accounting only for illumination, and independent of the material. B is the transfer matrix from the current color space coordinates to the base coordinates. The nonlinear transfer uses the following formula (3):

$$\begin{bmatrix} U \\ V \\ W \end{bmatrix} = A \left(\widehat{\ln} \left(B \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \right) \right), \tag{3}$$

where A and B are invertible 3×3 matrices and denote the component-wise natural logarithm. Matrix B transforms the color coordinates to the basis in which relighting best corresponds to multiplication by a diagonal matrix, while matrix A provides degrees of freedom that can be used to match perceptual distances. Based on similar color experiments in
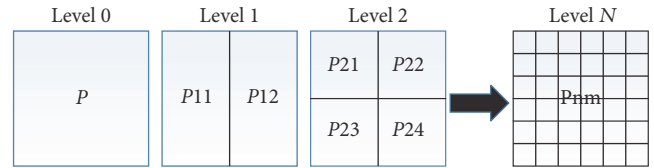

FIGURE 4: The method of SPM.

the database, A and B matrix-value formulas are shown as (4) and (5), respectively.

$$A = \begin{pmatrix} 27.07439 & -22.80783 & -1.806681 \\ -5.646736 & -7.722125 & 12.86503 \\ -4.163133 & -4.579428 & -4.576049 \end{pmatrix}, \tag{4}$$

$$B = \begin{pmatrix} 0.9465229 & 0.2946927 & -0.1313419 \\ -0.117917 & 0.9929960 & 0.007371554 \\ 0.0923046 & -0.046457 & 0.9946464 \end{pmatrix}. \tag{5}$$

3.3. SPM Model. Lazebnik et al. [21] proposed the Spatial Pyramid Matching (SPM) in 2006. SPM model contains broad space information, with which the color histogram information will be encoded orderly in space. The model divides the image into different levels, which can then be further refined. The SPM model space is shown in Figure 4. The level 0 image P is based on the original image feature information. But the image feature is based on the global unordered color information. Level 1 shows image separated as space geometry. $P11$ and $P12$ are expressed by a spatial order that contains simple space information.

P11 and P12, which also lack internal space information, are in level 1. If internal space information is necessary in P11 and P12, they must be separated using the same process. The level $i + 1$ feature is divided by level i. The levels of division are decided by the actual situation.

3.3.1. The SPM Histogram Feature.

Image similarity is computed by the levels corresponding to parts in SPM model. For two images P and Q, the formula is as follows:

$$d(P, Q) = \sum k_{ij} d\left(p_{ij}, q_{ij}\right), \qquad (6)$$

where P_{ij} is the image P histogram feature of the part j in level i; $d(p_{ij}, q_{ij})$ is the feature similarity degree images P and Q; and K_{ij} is the weight of the similarity calculation. In this case, we focus on part j of level i. The weight of calculation should be set high.

3.4. Gaussian Color Model.

Gaussian color model (GMM) is constantly used for color image segmentation according to the classification and clustering of image characteristics [22]. The image is divided into different parts based on pixel classification. We considered the main part of person identification to be based on minutia matching and ignored details. The retrieval of similar objects in a video system prioritizes the main part of similarity matching and does not emphasize accurate detail matching, so we considered the main colors as the features of the Gaussian color model.

3.4.1. Gaussian Distribution.

The Gaussian distribution is a parametric probability density function that is a mean value and variance continuous distribution maximum information entropy [23]. As shown in (7), when distributing a unit value that fits the normal distribution random variable, the frequency of the variable that follows the Gaussian distribution is entirely determined by the mean value μ and variance σ^2. As x approaches μ, probability increases. σ means the dispersion, and the value of σ is a much greater degree of dispersion.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}. \qquad (7)$$

For an image, the Gaussian distribution describes the distribution of specific pixel brightness that reflects the frequency of some gray numerical value [24]. A single-mode Gaussian distribution cannot represent a multicolored image. Therefore, we used a multiplicity of Gaussian models to show different pixel distributions that approximately simulate a multicolored image. Theoretically, we could increase the numbers of models to improve the descriptive ability.

Every pixel of the color image could be represented as a d dimensional vector x_i (color image $d = 3$ and gray image $d = 1$). The whole image could be represented as $X = (x_1^T, x_2^T, \ldots, x_N^T)$, where N is the sum of all pixels in a picture, X is represented as M states in GMM, and the value of M is usually restricted from 3 to 5. The linear stacking of the M Gaussian distributions could show the GMM of the

probability density function, as shown in (8): x is the pixel sampling of a picture.

$$P(x) = \sum_{k=1}^{M} p(k) p(x \mid k) = \sum_{k=1}^{M} \pi(k) N\left(x \mid \mu_k, \sum k\right). \qquad (8)$$

$N(x \mid \mu_k, \sum k)$ is the single Gaussian density function. As shown in (8), $k = 1, \ldots, M$ indicates the Gaussian density function of $No.k$. μ_k is the sample mean vector, $\sum k$ is sample covariance matrix, and π_k is the nonnegative coefficient of weight that describes the proportion of $No.k$ data in the total data.

3.5. Color Histogram Feature Extraction.

The histogram of an image is related to the probability distribution function of the images pixel density. When this concept is extended to a color image, it is necessary to obtain the joint probability distribution value for multiple channels [25]. In general, a color histogram is defined by the following equation (9):

$$h_{A,B,C} = N \bullet \text{Prob}(A = a, \ B = b, \ C = c), \qquad (9)$$

where A, B, and C indicate three color channels (R, G, and B or H, S, and V) and N is the sum of all pixels in the image. In terms of computing, the first step is to discretize the pixel values of the image, creating statistics for the number of pixels of each color for color histogram.

3.6. Histogram of Color Feature Similarity Measurement.

Several methods exist to calculate and weigh the similarity measurement of the histogram. The distance formula of the similarity measure between images is based on the color content. Euclidean distance, histogram intersection, and histogram quadratic distance are widely used in image retrieval.

The Euclidean distance of the histogram between two images is given by the following equation (10):

$$d^2(h, g) = \sum_A \sum_B \sum_C \left(h(a, b, c) - g(a, b, c)\right)^2, \qquad (10)$$

where h and g are two histograms and a, b, and c are the color channels. The formula subtracts the pixel value in the same bin of histograms h and g.

The formula for histogram intersection distance is as follows:

$$d(h, g) = \frac{\sum_A \sum_B \sum_C \min\left(h(a, b, c) - g(a, b, c)\right)}{\min\left(|h|, |g|\right)}, \qquad (11)$$

where $|h|$ and $|g|$ stand for the pixel values of image sampling in histograms h and g, respectively.

3.7. Evaluation Method.

(1) We focused on the degree of search result accuracy using evaluation parameters for precision. Precision reflects the capability of filtering irrelevant content. These video retrieval system performance criteria reference the evaluation method for information search systems. For a retrieval object, the retrieval system returns a

sort of search results. The precision rate expresses the number of correct relevant retrieval results divided by the number of total retrieval results.

$$\text{Precision (\%)} = \frac{A}{A + B} \times 100,$$

$$\text{AveragePrecision (\%)} = \frac{1}{n}\sum_{i=1}^{n}\text{Precision}(i).$$

$$(12)$$

In formula (12), A is the number of correct relevant retrieval examples, B is the number of irrelevant video retrieval examples, and C is the number of missing correct relevant retrieval examples.

(2) Cumulative Match Characteristic (CMC) curve is employed to evaluate the performance of the reidentification system. The CMC curve is used when the full gallery is available. It depicts the relationship between the accuracy and the threshold of rank. Most of the existing pedestrian reidentification algorithms use the CMC curve to evaluate the algorithm performance. Given a probe set and a pedestrian gallery set, the experimental result of CMC analysis describes what is the percentage of probe searches in the pedestrian dataset that returns the probes gallery mate within the top r rank-ordered results.

4. Experiment

We evaluate our reidentification method on three datasets, that is, the multicamera video data, the VIPeR data, and the SARC3D data. We examine our proposed SPM histogram + GMM main color method, the SPM histogram method, and the traditional histogram method on three datasets and further compare our method with the Symmetry-Driven Accumulation of Local Features (SDALF) method on the public VIPeR and SARC3D datasets. The code of SDALF could be downloaded on https://github.com/lorisbaz/sdalf. All the experiments are run on a desktop computer with an i7-3.4 GHz CPU.

4.1. Experiment on Multicamera Videos. We evaluated the performance of different color spaces for real-life video data. Uneven illumination distribution should affect person reidentification results in color images. Therefore, we created a video data set to test the validity and robustness of our method. We recorded the video data on a school campus. Six pedestrians walked from left to right in order under a surveillance camera, as shown in Figures 5 and 6. Our real-life video data consists of two videos that were recorded simultaneously at different locations. Location 1 was bright and location 2 was dark. The videos were recorded at 25 frames per second. Pictures of the side viewpoints of the six pedestrians were used as the retrieval samples, as shown in Figure 7. The six pedestrians were without a hat, bag, or other accessories. The RGB results are based on machine vision, while the HSV results are closer to human visual perception. As shown in Table 1, our proposed method outperforms the traditional histogram method and the SPM histogram method. We find that although the RGB color space reflects all sorts of colors from the images, the background color which is mixed in

Table 1: The average precision for persons retrieval in location 1.

Method	RGB	HSV	UVWS
Histogram	75	73.33	80
SPM histogram	71.66	75	70
GMM	86.66	85	88.33

Table 2: The average precision for persons retrieval in location 2.

Method	RGB	HSV	UVWS
Histogram	73.33	75	81.66
SPM histogram	83.33	85	80
GMM	81.66	83.33	85

these channels has affected the reidentification result. This problem is even severe in the SPM method, in which the lower part of the separated image contains a greater part of the background color than the body color. As shown in Table 2, the performance of UVW is better than HSV and RGB. The reason is that the results were affected mostly by the color transfer. In different illumination, the color histogram of one's clothes would be transferred to another color. For example, the red color in a dark environment seems like a black or gray color. The UVW color space is aimed at this problem. In the GMM color modeling, to solve the color transfer problem in low resolution images, we employ the primary colors of red, blue, and green as the dominant colors. However, for the dark background images, the GMM method generates a poor result.

4.2. Experiment on VIPeR Dataset. We examine the appearance model for person reidentification based on the VIPeR dataset, which consists of 632 pedestrian image pairs taken from arbitrary viewpoints under varying illumination conditions. Each image is scaled to 128×48 pixels.

As shown in Figure 8, our proposed method outperforms the histogram-based methods in the RGB color space, and the traditional histogram and the SPM histogram methods generate very similar results. We also observe that the proposed method in the HSV space performs better than in the RGB space, as shown in Figure 9. This is because that the image illumination in the VIPeR dataset varies significantly. The SDALF method renders a slightly better result than our proposed method, while our method has a great advantage on the calculation cost. Specifically, the SDALF takes about 3850 seconds to extract its features from 1264 images in the VIPeR dataset, while our proposed method takes only 40 seconds to extract and calculate the color histogram features. In addition, the SDALF method needs about 4260 seconds to compare all 399424 pairs of images, while our method needs only 610 seconds to calculate the GMM similarity for comparison in 1264 images. This result suggests that in terms of computational cost our approach significantly outperforms the SDALF method.

4.3. Experiment on SARC3D Dataset. The SARC3D dataset consists of short video clips of 50 people which have been

FIGURE 5: Location 1.
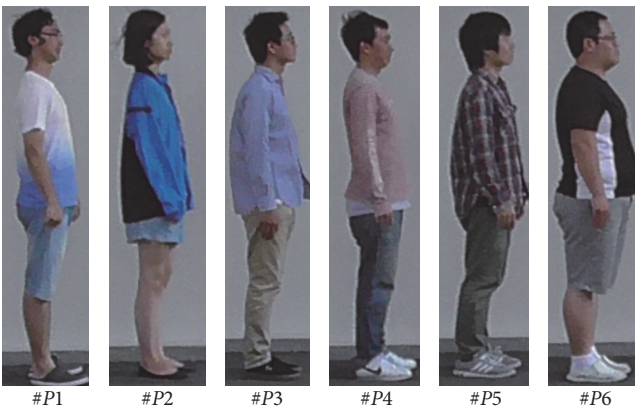

FIGURE 6: Location 2.

#P1   #P2   #P3   #P4   #P5   #P6

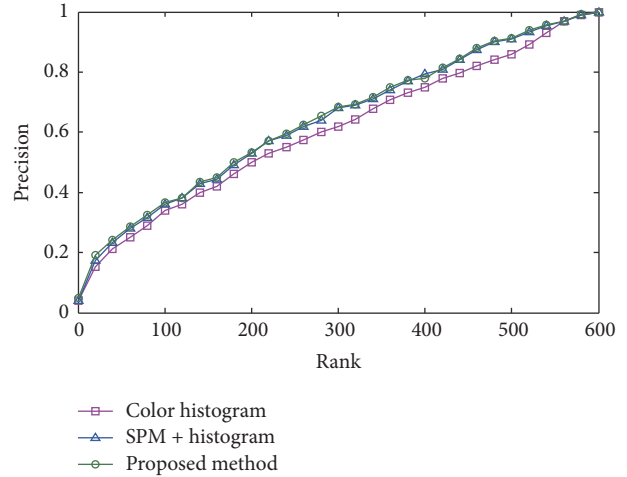FIGURE 7: Example of placing a figure with experimental results.

FIGURE 8: CMC curves on the VIPeR dataset for the proposed method and histogram methods in RGB space.
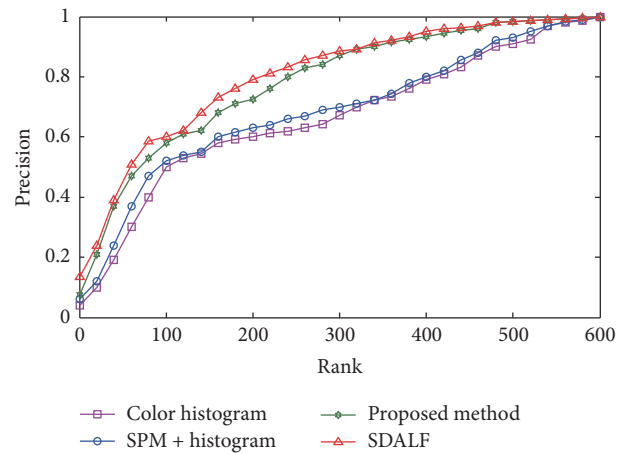
FIGURE 9: CMC curves on the VIPeR dataset for the proposed method and the other methods in HSV space.

captured with a calibrated camera. We employ the SARC3D dataset to effectively evaluate different person reidentification methods. To simplify the image alignment process, we manually select four frames for each clip which correspond to the predefined positions and postures, that is, back, front, left, and right, of these people. The selected dataset consists of 200 snapshots with four views for each person. For person reidentification, we randomly choose one of the four views for each person, calculate the similarity scores with all other images, and find the most similar images by sorting their similarities with the chosen image. The images of the same person

with different positions and postures should be ranked higher than the other images. In the dataset, 6 people are not fully visible in their images and 2 people are observed with the same dressing, that is, colors and combinations, except for the waling postures. We remove images of these people to avoid the different size of their masks form in the original images. All methods in the experiment are based on the RGB color space. Figure 10 shows the average CMC curves for the person reidentification under different methods. Our method significantly outperforms the SDALF method in recognition rate because the backward information in GMM matching has been filtered out given the people annotation template in the dataset. In the meantime, our method significantly outperforms the SDALF method in calculation cost, with only 30 seconds for color histogram feature extraction and image matching in 126 images, while the latter takes about 440 seconds for feature extraction and 70 more seconds for image matching.
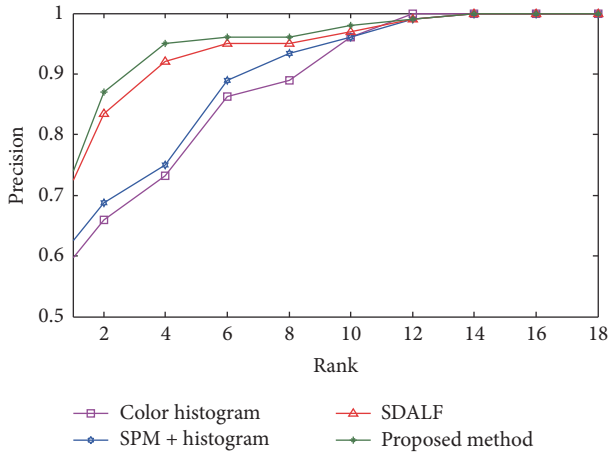
FIGURE 10: CMC curves on the VIPeR dataset for the proposed method and the other methods in RGB space.

5. Conclusion

Person reidentification in multicamera videos often has some problems that contain person pose variation, varying illumination, and low image resolution. We propose to solve two common problems in person reidentification, which are the varying illumination and low image resolution. Varying illumination conditions usually occur because of the difference between camera views. For example, the same people in different camera video have a color transfer. The low resolution image often contains high noise. It is difficult to extract the robust feature from the low resolution image. In order to improve the illumination problem in histogram methods, we introduce the perception-based color space which has been successfully employed in the image segmentation research into the person identification method. Secondly, for the low resolution images we incorporate spatial pyramid matching (SPM) method into the main color extraction method, which has shown great improvement in our experiment. In addition, our method has shown significant advantage in the computation cost compared with the traditional methods. In this paper we just extract the main color feature by the GMM model. We did not analyse the feature information from the mean value parameter and variance in the GMM. The main color feature also used the global object color; we could combine the SPM model with GMM main color local feature to retrieve the object from the video data.

Competing Interests

The authors declare that they have no competing interests.

Acknowledgments

References

[1] R. Satta, "Appearance descriptors for person reidentification: a comprehensive review," https://arxiv.org/abs/1307.5748.

[2] A. Dangelo and J.-L. Dugelay, "People re-identification in camera networks based on probabilistic color histograms," in *Visual Information Processing and Communication II*, vol. 7882 of *Proceedings of SPIE*, January 2011.

[3] G. Doretto, T. Sebastian, P. Tu, and J. Rittscher, "Appearance-based person reidentification in camera networks: problem overview and current approaches," *Journal of Ambient Intelligence and Humanized Computing*, vol. 2, no. 2, pp. 127–151, 2011.

[4] M. A. Saghafi, A. Hussain, H. B. Zaman, and M. H. Md Saad, "Review of person re-identification techniques," *IET Computer Vision*, vol. 8, no. 6, pp. 455–474, 2014.

[5] S. Lee, N. Kim, K. Jeong, I. Paek, H. Hong, and J. Paik, "Multiple moving object segmentation using motion orientation histogram in adaptively partitioned blocks for high-resolution video surveillance systems," *Optik—International Journal for Light and Electron Optics*, vol. 126, no. 19, pp. 2063–2069, 2015.

[6] A. Bedagkar-Gala and S. K. Shah, "A survey of approaches and trends in person re-identification," *Image and Vision Computing*, vol. 32, no. 4, pp. 270–286, 2014.

[7] R. Vezzani, D. Baltieri, and R. Cucchiara, "People reidentification in surveillance and forensics: a survey," *ACM Computing Surveys*, vol. 46, no. 2, article no. 29, 2013.

[8] X. Wang and R. Zhao, "Person re-identification: system design and evaluation overview," in *Person Re-Identification*, pp. 351–370, Springer, 2014.

[9] B. Ma, Q. Li, and H. Chang, "Gaussian descriptor based on local features for person re-identification," in *Proceedings of the Asian Conference on Computer Vision (ACCV '14)*, pp. 505–518, Springer, Singapore, November 2014.

[10] D. Gray and H. Tao, "Viewpoint invariant pedestrian recognition with an ensemble of localized features," in *Computer Vision—ECCV 2008: 10th European Conference on Computer Vision, Marseille, France, October 12–18, 2008, Proceedings, Part I*, vol. 5302 of *Lecture Notes in Computer Science*, pp. 262–275, Springer, Berlin, Germany, 2008.

[11] M. Farenzena, L. Bazzani, A. Perina, V. Murino, and M. Cristani, "Person re-identification by symmetry-driven accumulation of local features," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '10)*, pp. 2360–2367, IEEE, San Francisco, Calif, USA, June 2010.

[12] R. Zhao, W. Ouyang, and X. Wang, "Unsupervised salience learning for person re-identification," in *Proceedings of the 26th IEEE Conference on Computer Vision and Pattern Recognition (CVPR '13)*, pp. 3586–3593, IEEE, Portland, Ore, USA, June 2013.

[13] M. J. Metternich, M. Worring, and A. W. M. Smeulders, "Color based tracing in real-life surveillance data," in *Transactions on Data Hiding and Multimedia Security V*, vol. 6010 of *Lecture Notes in Computer Science*, pp. 18–33, Springer, Berlin, Germany, 2010.

[14] M. Hirzer, P. M. Roth, M. Köstinger, and H. Bischof, "Relaxed pairwise learned metric for person re-identification," in *Proceedings of the European Conference on Computer Vision*, pp. 780–793, Springer, Florence, Italy, October 2012.

[15] M. I. Khedher, M. A. El-Yacoubi, and B. Dorizzi, "Probabilistic matching pair selection for SURF-based person re-identification," in *Proceedings of the International Conference of the Biometrics Special Interest Group (BIOSIG '12)*, pp. 1–6, Darmstadt, Germany, September 2012.

[16] T. Matsukawa, T. Okabe, and Y. Sato, "Person re-identification via discriminative accumulation of local features," in *Proceedings of the 22nd International Conference on Pattern Recognition*

(ICPR '14), pp. 3975–3980, IEEE, Stockholm, Sweden, August 2014.

[17] W.-S. Zheng, S. Gong, and T. Xiang, "Reidentification by relative distance comparison," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 3, pp. 653–668, 2013.

[18] H. Y. Chong, S. J. Gortler, and T. Zickler, "A perception-based color space for illumination-invariant image processing," *ACM Transactions on Graphics*, vol. 27, no. 3, article 61, 2008.

[19] K.-J. Yoon and I.-S. Kweon, "Human perception based color image quantization," in *Proceedings of the 17th International Conference on Pattern Recognition (ICPR '04)*, pp. 664–667, August 2004.

[20] L. Shamir, "Human perception-based color segmentation using fuzzy logic," *IPCV*, vol. 2, pp. 96–502, 2006.

[21] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: spatial pyramid matching for recognizing natural scene categories," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '06)*, vol. 2, pp. 2169–2178, New York, NY, USA, June 2006.

[22] D. A. Reynolds, T. F. Quatieri, and R. B. Dunn, "Speaker verification using adapted Gaussian mixture models," *Digital Signal Processing*, vol. 10, no. 1, pp. 19–41, 2000.

[23] C. E. Rasmussen, "The infinite gaussian mixture model," *NIPS*, vol. 12, pp. 554–560, 1999.

[24] Z. Zivkovic, "Improved adaptive gaussian mixture model for background subtraction," in *Proceedings of the 17th International Conference on Pattern Recognition (ICPR '04)*, vol. 2, pp. 28–31, Cambridge, UK, August 2004.

[25] C. Liu, S. Gong, C. C. Loy, and X. Lin, "Person re-identification: what features are important?" in *Computer Vision—ECCV 2012. Workshops and Demonstrations: Florence, Italy, October 7–13, 2012, Proceedings, Part I*, vol. 7583 of *Lecture Notes in Computer Science*, pp. 391–401, Springer, Berlin, Germany, 2012.