21st International Conference on Pattern Recognition (ICPR 2012)
November 11-15, 2012. Tsukuba, Japan

Person Re-identification using View-dependent Score-level Fusion of Gait and Color Features

Ryo Kawai, Yasushi Makihara, Chunsheng Hua, Haruyuki Iwama, Yasushi Yagi
ISIR, Osaka University, Japan
{*kawai, makihara, hua, iwama, yagi*}*@am.sanken.osaka-u.ac.jp*

Abstract

This paper describes a method for person re-identification across multiple non-overlapping cameras using both gait and color features. Because a single color feature is insufficient to distinguish persons with similar color clothes, a spatio-temporal histogram of oriented gradients is employed as a gradient-based shape and motion gait feature to discriminate such persons in conjunction with a background edge attenuation technique. However, since the gait feature is more sensitive to view differences than the color feature, a view-dependent score-level fusion framework adaptively controls the weights of the gait and color features. Experiments across seven non-overlapping cameras confirm the effectiveness of the proposed method.

1. Introduction

With the increased number of surveillance cameras in public areas for security and forensic purposes, requirements for automation functions such as pedestrian detection and tracking have also increased over the last decade. Person re-identification across multiple non-overlapping cameras is one such important function that can be applied to cross-camera tracking of suspicious persons and finding lost children.

Person re-identification can be achieved based on several cues such as the personfs face [6] and/or body (clothes) features. Because face recognition is difficult when the resolution of the captured images is low, more emphasis has been placed on body feature-based person re-identification.

Since it is reasonable to assume that the clothes of a pedestrian remain unchanged within the person re-identification problem setting, the majority of approaches for person re-identification exploit color information [8]. Some researchers have further incorporated color calibration techniques between cameras to overcome color differences due to differences in the camera white balance and/or illumination conditions [11]. Nevertheless, difficulties arise with these color-based approaches when there are multiple persons wearing similar color clothes.

Shape features are alternative cues for identifying persons and are jointly used with the color features in [17]. Shape-based methods, however, still suffer from large intra-subject variations derived from body pose changes.

On the other hand, it is well-known that *walking motion* information extracted from a video is a useful cue in identifying a person, a process commonly known as *gait recognition* [9]. In gait recognition, shape variations derived from the walking motion are regarded as useful cues in identifying a person instead of intra-subject variations.

In this paper, a gait feature containing both shape and motion information is incorporated into the person re-identification framework together with color information. This enables us to distinguish persons with similar color clothes. Note that this kind of walking motion information has not yet been used effectively even in a method incorporating multiple features collected from a short video [5].

The gait feature is, however, more sensitive to observation view differences than the color feature, as reported in cross-view gait recognition [18], which may be one of the major reasons why gait recognition approaches have not yet been employed in person re-identification methods. In other words, an observation view difference can be regarded as a *quality measure* [3] or confidence in each gait and color feature in the fusion framework.

We therefore, introduce view-dependent score-level fusion of the gait and color features. Intuitively speaking, a greater weight is given to the gait feature for similar-view matching, and to the color feature for different-view matching, thereby realizing better performance on the whole. Note that such adaptive weight control is automatically learnt using a joint distribution of the gait and color scores in a training set.

2. Assumptions and Problem Setting

The whole framework for video-based person re-identification is composed of several modules, including pedestrian detection, tracking, and matching. Be-

978-4-9906441-1-6

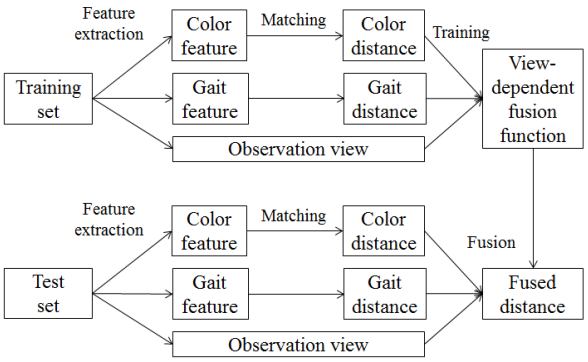

Figure 1. Overview

cause we focus on the matching module in this paper, pedestrian detection and tracking for each camera are assumed to be preprocessed by using a state-of-the-art technique such as that given in [13]. Moreover, for the purpose of view-dependent score-level fusion, an observation view for each pedestrian is also assumed to be given. Note that the observation view is relatively easily obtained based on the azimuth angle from the camera and tangential direction of the pedestrian's trajectory on the ground plane as proposed in [15] as long as each camera is geometrically calibrated in advance. Therefore, our problem setting is viewed as pedestrian matching across non-overlapping cameras using cropped pedestrian image sequences and observation views.

3. Proposed Framework

3.1. Overview

The proposed framework is mainly composed of training and test phases as shown in Fig. 1. In both phases, color and gait features as well as the observation view are extracted from an image sequence, and subsequent color and gait feature matching returns the individual distances. In the training phase, a view-dependent score-level fusion function is trained from a joint distribution of the color and gait distances and positive (the same person) and negative (different person) labels. In the test phase, the trained fusion function converts the color and gait distances using an observation view into a fused distance, which is used in the final decision. Details of each process are given in the following subsections.

3.2. Feature Extraction

Gait feature: The majority of gait recognition approaches exploit silhouette-based features (e.g., an averaged silhouette [7]). It may, however, be difficult to acquire a silhouette with sufficient quality for gait recognition under various background and lighting conditions across cameras. Moreover, the silhouette-based features discard useful motion information within the silhouette. Inspired by the great success of a gradient-based method for pedestrian detection [4], in this work a spatio-temporal histogram of oriented gradient (STHOG) features is employed as the gait feature containing both shape and motion information.

The pedestrian window is first size-normalized and is then divided into multiple spatio-temporal cells. The spatio-temporal gradient $\boldsymbol{G} = [G_x, G_y, G_t]^T$ is computed as the first-order derivative for each of the $x-$, $y-$, and t-axes from two successive frames. The orientation of spatial gradient ϕ and that of temporal gradient θ are defined as

$$\phi = \arctan\left(\frac{G_y}{G_x}\right),\ \theta = \arctan\left(\frac{G_t}{\sqrt{G_x^2 + G_y^2}}\right), \quad (1)$$

The orientation of the spatial and temporal gradients are separately voted according to a spatial gradient magnitude into individual 9-bin histograms for each cell, and are then normalized with regard to the L_1 norm. These two histograms are combined into a single histogram with 18 bins, and then the histograms from all the cells are further combined into a single histogram, thus constructing the STHOG feature.

In the voting process, suppressing the effect of background variations must be considered. A simple way of achieving this, which is actually used in many color-based approaches, is to mask out the background region using background subtraction. Background subtraction, however, is unsuitable for gradient-based features since these largely depend on the contour of the foreground region which is quite difficult to extract without any over-segmentation.

Therefore, we incorporate a gradient-based background attenuation technique [16] rather than a region-based mask. For this purpose, we adaptively attenuate the spatial gradient magnitude used as the voting weight. The basic idea is to attenuate an edge at a certain position if a background edge also exists at the same position and the edge patterns of the background and input images are similar. Consequently, the attenuated spatial gradient magnitude G_s is defined as

$$G_s = \frac{\sqrt{G_x^2 + G_y^2}}{\left(1 + K{G_x^B}^2 e^{-\frac{z_x^2}{\sigma^2}}\right)\left(1 + K{G_y^B}^2 e^{-\frac{z_y^2}{\sigma^2}}\right)}, \quad (2)$$

where K and σ are hyper parameters to control background attenuation, G_x^B and G_y^B are the horizontal and vertical gradients in the background image, and z_x and z_y denote the dissimilarity of the horizontal and vertical edge patterns in the background and input images, respectively, which are defined as

$$z_x = \max\{\|I^B_{x,y} - I_{x,y}\|, \|I^B_{x+1,y} - I_{x+1,y}\|\} \quad (3)$$

$$z_y = \max\{\|I^B_{x,y} - I_{x,y}\|, \|I^B_{x,y+1} - I_{x,y+1}\|\}, \quad (4)$$

where $I^B_{x,y}$ and $I_{x,y}$ are the pixel intensities at position (x, y) in the background and input images, respectively.
Color feature: The color histogram defined in a modified hue-saturation-value color system [10] is adopted

in this study. The histogram for each cell contains ten bins, which are composed of seven hue bins and three value bins (white, gray, and black). Unlike the STHOG feature, the color feature is a region-based feature, and hence the background region is simply masked out by background subtraction when voting. The color histograms are normalized for individual cells and are concatenated into a single histogram in the same way as the STHOG feature.

Note that since both the STHOG and color features are computed for individual frames in each image sequence, a series of features are extracted for each image sequence.

3.3. Matching

Because the series of features contain various gait stances, these need to be synchronized when matching. For this purpose, a variant of the baseline algorithm [14] is employed. We briefly describe the matching method, and refer the readers to [14] for more details.

A query sequence is segmented into subsequences with a certain length (e.g., gait period) and each of these is phase-synchronized with the other sequence by shifting the frame so as to minimize the total distance between them. Whereas the baseline algorithm utilizes the Tanimoto distance, we replace it with the $L_{0.5}$ norm to improve the robustness with respect to outliers. Subsequently, the distances of individual subsequences are computed, and the minimum value is chosen as the final distance.

3.4. View-dependent Score-level Fusion

For effective use of the STHOG and color features, we introduce a score-level fusion function, which maps the two-dimensional distance vectors derived from the STHOG and color features into a single distance. Since the reliability of STHOG features is highly dependent on the observation view difference, this fusion function is trained separately for each discrete observation view difference.

Of the vast array of approaches to score-level fusion, linear logistic regression (LLR) [1] of the likelihood ratio between positive and negative samples is chosen for its simplicity and high generalization capability. Moreover, inspired by the recent success of ranking methods in person re-identification [12], the relative distance of each negative sample with respect to that of a corresponding positive sample is input into the fusion framework.

Once the fusion function of the LLR has been trained, the input two-dimensional distance vector in a test set is converted into a single fused distance. Finally, the person in one camera image is re-identified to the person with the minimum fused distance in another camera image.

4. Experiment

4.1. Dataset and Setup

Because there are no publicly available datasets for person re-identification that include both successive image sequences and observation view information, we have constructed our own dataset. We captured walking image sequences of 27 subjects (14 subjects for training and 13 subjects for testing) from seven non-overlapping views ranging from the front view to the rear-oblique view. Some of the subjects wore clothes with fairly similar colors (Fig. 2), and hence the dataset is challenging. Of the seven views, the near side view was used as the query view, and a person in the query view was matched with all persons in another view.

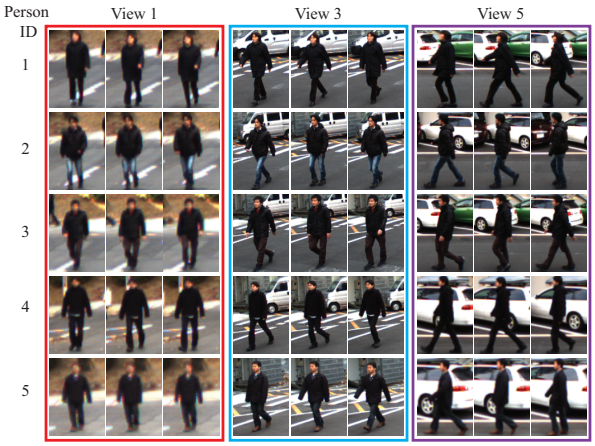

Figure 2. Sample cropped image sequences from our dataset

The size normalized window and spatio-temporal cell sizes were set to 100 $\times$ 140 pixels, and 100 pixels $\times$ 20 pixels $\times$ 3 frames, respectively, giving seven cells per window. Hyper parameters K and σ for background attenuation were set to 5.0 and 10.0, respectively, which are the default values defined in [16]. The length of the subsequence for matching was set experimentally to 15 frames.

Experiments on person re-identification were repeated 20 times by randomly choosing the training and test subjects. Averaged performances for the 20 trials were evaluated by cumulative matching characteristics.

4.2. Results

Results of person re-identification experiments are shown in Fig. 3. We can see that the performance using the STHOG feature is higher than that using the color feature for similar observation views (e.g., view IDs 4, 5, and 6), whereas it is lower than that using color features for different observation views (e.g., view IDs 1, 2, and 3). The proposed fusion method successfully controls the weight of the STHOG and color features depending on the observation view differences, and achieves the best performance on average.

5 Conclusion

This paper described a method for person re-identification across multiple non-overlapping cameras using both gait and color features. The STHOG feature

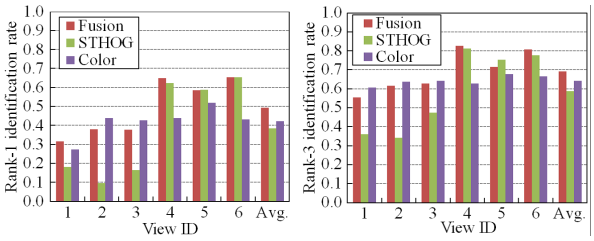

Figure 3. Rank-1 (left) and rank-3 (right) identification rates. View IDs 1 to 6 correspond to gradual view transitions from the front view to the rear-oblique view. The query view is the closest to ID 5 (side view).

is employed as a gradient-based shape and motion gait feature to discriminate persons with similar clothes in conjunction with a background edge attenuation technique. Moreover, view-dependent score-level fusion of the gait and color features by linear logistic regression is incorporated to overcome the sensitivity of gait features to view changes. Experiments on person re-identification across seven views confirmed the effectiveness of the proposed method.

Future work is to evaluate the proposed method with more large-scale database for person re-identification (e.g., 3DPes [2]).

Acknowledgement

This work was partially supported by Grant-in-Aid for Scientific Research (S) 21220003 and the hR&D Program for Implementation of Anti-Crime and Anti-Terrorism Technologies for a Safe and Secure Societyh, Strategic Funds for the Promotion of Science and Technology of the Ministry of Education, Culture, Sports, Science and Technology in the Japanese Government.

References

[1] F. Alonso-Fernandez, J. Fierrez, D. Ramos, and J. Ortega-Garcia. Dealing with sensor interoperability in multi-biometrics: the upm experience at the biosecure multimodal evaluation 2007. In *Proc. of SPIE 6994, Biometric Technologies for Human Identification IV*, Orlando, FL, USA, Mar. 2008.

[2] D. Baltieri, R. Vezzani, and R. Cucchiara. 3dpes: 3d people dataset for surveillance and forensics. In *Proc. of the 1st Int. ACM Workshop on Multimedia access to 3D Human Objects*, pages 59–64, Scottsdale, Arizona, USA, Nov 2011.

[3] S. Bengio, C. Marcel, S. Marcel, and J. Mariethoz. Confidence measures for multimodal identity verification. *Information Fusion*, 3(4):267–276, 2002.

[4] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *CVPR2005*, II, pages 886–893, 2005.

[5] O. Hamdoum, F. Moutarde, B. Stanciulescu, and B. Steux. Persion re-identification in multi-camera system by sinature based interest point descriptors collected on short video sequences. In *Proc. of the 2nd ACM/IEEE Int. Conf. on Distributed Smart Cameras*, pages 1–6, 2008.

[6] Y. Ijiri, S. Lao, T. Han, and H.Murase. Efficient facial attribute recognition with a spatial codebook. In *Proc. of the 20th Int. Conf. on Pattern Recognition*, pages 1461–1464, Aug. 2010.

[7] Z. Liu and S. Sarkar. Simplest representation yet for gait recognition: Averaged silhouette. In *Proc. of the 17th Int. Conf. on Pattern Recognition*, volume 1, pages 211–214, Aug. 2004.

[8] C. Nakajima, M. Pontil, B. Heisele, and T. Poggio. Full-body person recognition system. *Pattern Recognition*, 36(9):1997–2006, 2003.

[9] M. S. Nixon, T. N. Tan, and R. Chellappa. *Human Identification Based on Gait*. International Series on Biometrics. Springer-Verlag, Dec. 2005.

[10] U. Park, A. Jain, I. Kitahara, K. Kogure, and N. Hagita. Vise: Visual seach engine using multiple networked cameras. In *Proc. of the 18th Int. Conf. on Pattern Recognition*, pages 1204–1207, 2006.

[11] F. Porikli. Inter-camera color calibration by correlation model function. In *2003 IEEE Int. Conf. on Image Processing*, volume 3, pages II–133–6, 2003.

[12] B. Prosser, W.-S. Zheng, S. Gong, and T. Xiang. Person re-identification by support vector ranking. In *Proc. of the 21th British Machine Vision Conf.*, pages 21.1–11, 2010.

[13] M. Rodriguez, J. Sivic, I. Laptev, and J.-Y. Audibert. Density-aware person detection and tracking in crowds. In *Proc. of the 13th Int. Conf. on Computer Vision*, 2011.

[14] S. Sarkar, J. Phillips, Z. Liu, I. Vega, P. G. ther, and K. Bowyer. The humanid gait challenge problem: Data sets, performance, and analysis. *Trans. of Pattern Analysis and Machine Intelligence*, 27(2):162–177, 2005.

[15] K. Sugiura, Y. Makihara, and Y. Yagi. Gait identification based on multi-view observations using omnidirectional camera. In *Proc. 8th Asian Conference on Computer Vision*, pages 452–461, Tokyo, Japan, Nov. 18-22 2007. LNCS 4843.

[16] J. Sun, W. Zhang, X. Tang, and H.-Y. Shum. Background cut. In *Proc. of the 9-th European Conf. on Computer Vision*, volume 2, pages 628–641, 2006.

[17] X. Wang, G. Doretto, T. Sebastian, J. Rittshcer, and P. Tu. Shape and appearance context modeling. In *Proc. of the 11th Int. Conf. on Computer Vision*, pages pp.1–8, 2007.

[18] S. Yu, D. Tan, and T. Tan. Modelling the effect of view angle variation on appearance-based gait recognition. In *Proc. of 7th Asian Conf. on Computer Vision*, volume 1, pages 807–816, Jan. 2006.