Enhancing Person Re-identification by Integrating Gait Biometric

Zheng Liu[1], Zhaoxiang Zhang[1], Qiang Wu[2], Yunhong Wang[1]

[1]Laboratory of Intelligence Recognition and Image Processing, Beijing Key Laboratory of Digital Media, School of Computer Science and Engineering, Beihang University, Beijing 100191, China
[2]School of Computing and Communications, University of Technology, Sydney, Australia

Abstract. This paper proposes a method to enhance person re-identification by integrating gait biometric. The framework consists of the hierarchical feature extraction and matching methods. Considering the appearance feature is not discriminative in some cases, the feature in this work composes of the appearance feature and the gait feature for shape and temporal information. In order to solve the view-angle change problem and measuring similarity, metric learning to rank is adopted. In this way, data are mapped into a metric space so that distances between people can be measured accurately. Then two fusion strategies are proposed. The score-level fusion computes distances of the appearance feature and the gait feature respectively and combine them as the final distance between samples. Besides, the feature-level fusion firstly installs two types of features in series and then computes distances by the fused feature. Finally, our method is tested on CASIA gait dataset. Experiments show that gait biometric is an effective feature integrated with appearance features to enhance person re-identification.

1 Introduction

With the growing demand for security surveillance, how to accurately identify people has drawn increasing attentions. Person re-identification is an important part of surveillance. The problem can be defined as recognizing an individual who has already appeared in another camera in a non-overlapped multi-camera system.

The research on person re-identification often concentrates on two main aspects: the extraction of features and matching methods, in which the former one focuses on how to describe individuals and matching methods try to measure distances between samples.

The selected feature should be robust to the variation of illumination, posture and view. Some excellent appearance-based feature extraction methods have been proposed. Farenzena *etal*. [4] used color histograms, MSCR [5] as well as RHSP with the symmetry property to describe a local patch. Gray and Tao [6] let a machine learning algorithm find the best representation. Bazzani *etal*. [1] condensed a set of frames of an individual into a highly informative signature, called

the Histogram Plus Epitome (HPE). Zhao *etal*. [17] found salience of people based on SIFT [12] and color histogram and then match patches with constraint. However, if there are several people having similar appearance, or appearance of the same person is quite different in different cameras, these appearance-based methods would have a poor performance.

Fig. 1. Example of people observed from different views with different appearance in CASIA dataset. Each row is the same individual that (a) is condition with bag, (b) is condition with coat and (c) is normal condition.

In those proposed approaches, biometric is seldom mentioned, which we think should also be taken into account. Biometric gradually developed into an important kind of features which has a strong individual discrimination. It would play a significant role in person re-identification problem if used properly. Even though many mature methods of traditional biometrics have been proposed, such as fingerprint and face, they have to be obtained by getting close to target people, or making contact with them. In recent years, gait features are proposed and gradually attracted the attention of many scholars. There are many methods have been proposed about Gait Recognition. Little and Boyd [10] developed the shape of motion which is a model-free description of instantaneous motion, and used it to recognize individuals by their gait. Sarkar *etal*. [16] measured the similarity between the probe sequence and the gallery sequence directly by computing the correlation of corresponding frame pairs.

The gait feature is an ideal way for application in security surveillance as a biometric, because getting gait features is just using a camera from a long distance away and then doing image processing. In addition, carrying temporal

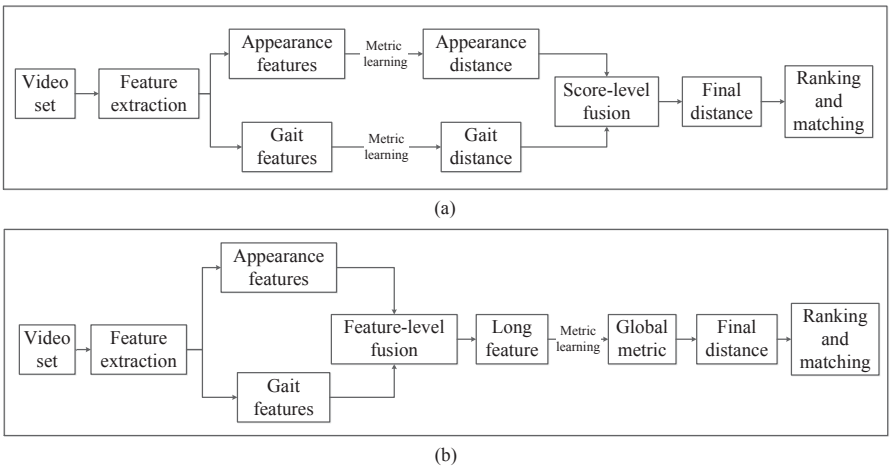


Fig. 2. Overview of proposed method. (a) is framework with score-level fusion and (b) is with feature-level fusion.

information is the other good quality of gait features. Instead of using the single-shot image, image sequences grabbed from videos are used to generate temporal feature, considering inputs of cameras are sequences of images. Thus, a spatio-temporal analysis for person re-identification can be performed by integrating the gait feature with appearance features.

For these reasons, the gait feature can be used to enhance sequence based person re-identification. The gait feature is robust even if people change their appearance, where appearance features are helpless. The gait feature is not such strong discriminative as traditional biometric. However, it is much more convenient to fuse gait with appearance features extracted from surveillance videos. From this perspective, the gait biometric is more suitable for person re-identification than recognizing problems. Since a fusion step has to be processed, the selection of features and fusion strategies are important and hard work.

In this paper, we try to solve the person re-identification problem by integrating the gait feature with appearance features. For appearance features, HSV histogram is widely used to describe the color information, and texture information can be well represented by Gabor feature [11]. Gait Energy image (GEI) [7] is a popular feature that used to represent the gait biometric. So our descriptor composed of HSV histogram and Gabor feature as the appearance feature and GEI as the gait feature. Then these two type features are fused by two strategies, namely score-level fusion and feature-level fusion. After the data are modeled by descriptive features, a metric learning method is adopted for similarity measurement. The idea of metric learning is comparing descriptors by a learned metric instead of Euclidean distance. Finally, we test our method on CASIA dataset which is a gait dataset created by Chinese Academy of Sciences.

2 Problem Description

In the real application scene, the framework for video-based person re-identification consists of many modules, such as individual detection, tracking and matching. Assuming that a background image has been given for each view, because the focus of this work is on feature extraction and matching method. On this basis, not only both detection and tracking step can be taken easily by background subtraction method, Gait Energy Images can also be obtained conveniently. In fact, getting background images in many surveillance scene is not an impossible work, so this assumption is not very strong. Therefore, our problem setting can be set forth as individual matching across non-overlapped cameras with pedestrian videos and background image of all views.

As an overview of the framework proposed in this paper, hierarchical feature extraction and matching are two main steps. The appearance feature and the gait feature are two parts of the descriptor. The HSV color histogram and Gabor filter are used to describe the appearance of people. Gait Energy Image is generated to obtain the spatio-temporal information and Principal Component Analysis (PCA) [15] is used to obtain the low-dimension GEI feature. The gait feature is integrated with the appearance feature by two different fusion strategies.In the matching step, there are different procedures for different fusion strategies. For score-level fusion, two metrics are trained for the appearance feature and GEI feature, respectively. After that, two distances are computed and fused to obtain the final distance between samples. For feature-level fusion, a global metric matrix is trained by metric learning to matching the fused feature and then similarities are measured by this global metric.

3 Proposed Framework

3.1 Hierarchical Feature Extraction

Gait Feature. Gait Energy Image (GEI) is the gait feature used in proposed method. In general, GEI is generated by silhouette images extracted from the view of 90°. Nevertheless, GEI is considered still contains some global information of individuals such as body shape and spatio-temporal changes of the human body in other views in addition to 90°. So in this work, GEI generated of all 11 views is adopted from 0° to 180° for cross-view analysis.

Firstly, background subtraction is used to extract the connected area of foreground. Then a series of foreground images are grabbed from each frame in the video, which form a sequence of silhouette images for each person of each view. Meanwhile, each image is normalized to the size of 64×32 for the convenience to be processed. Given the binary gait silhouette images $S_{ijt}(x, y)$ at time t in a sequence of person i from view j, Gait Energy Images then can be obtained as follows:

$$G_{ij}(x, y) = \frac{1}{N_{ij}} \sum_{t=1}^{N_{ij}} S_{ijt}(x, y) \tag{1}$$

where N_{ij} is the total number of frames in view j of person i. It may be different for each person and view because of the different length of videos. t is the frame index in the sequence and x and y are the coordinates of pixels in the image.

As the value of a pixel in the Gait Energy Image represents the possibility that the human body appears in the position, the feature can be described as a 1D vector that composed of values of all pixels. Because of the normalization step, the GEI feature is a 2048-dimension vector, so that dimensionality reduction has to be taken. Principal Component Analysis (PCA) is adopted since it is a classical approach that can reduce dimension of GEI feature significantly.

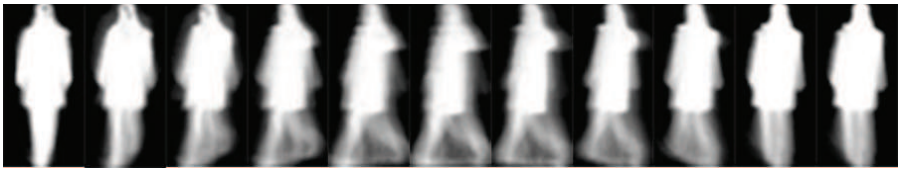


Fig. 3. Example of Gait Energy Images of all 11 views.

Appearance Features. Multiple frames help little for extraction of appearance features but increase calculation largely according to our experiments. So that a single frame for each person is randomly picked at this step to generate appearance features. The first part of the appearance feature is Hue-Saturation-Value (HSV) space color histogram. The reason why choosing HSV color space instead of RGB color space is that HSV color space is more adapting to the human perception of color and HSV model is not sensitive to illumination variation. Each person image is transformed from RGB color space to HSV color space in normalization step, then histogram equalization is carried out to reinforce the robustness to illumination further. Each dimension of HSV is divided into 128 bins to count the number of pixels whose value falls into the corresponding bin.

The other part is Gabor features [11]. The frequency and direction of Gabor filter are close to what in the human visual system so that the Gabor filter is often used to generate the texture feature in many fields, such as face and fingerprint recognizing. A 2-dimension Gabor filter is a Gaussian kernel function modulated by a complex sinusoidal plane wave in spatial domain. A Gabor filter can be defined as follows [3], [9], [13]:

$$\psi_{u,v}(z) = \frac{||k_{u,v}||^2}{\sigma^2} e^{-\frac{||k_{u,v}||^2 ||z||^2}{2\sigma^2}} [e^{ik_{u,v}z} - e^{-\frac{\sigma^2}{2}}] \tag{2}$$

where u is the orientation and v is the scale of the Gabor filters, $z = (x, y)$, and $k_{u,v}$ is defined as:

$$k_{u,v} = k_v e^{i\phi_\mu} \tag{3}$$

where $k_v = k_{max}/f^v$ and $\phi_\mu = \pi\mu/8$. k_{max} is the maximum frequency, and f is the spacing factor between kernels in the frequency domain [9].

Gabor filter of five different scales and eight orientations is used with the following parameters: $\sigma = \pi$, $k_{max} = \pi/2$ and $f = \sqrt{2}$.

3.2 Metric Learning to Rank

With the assumption that the view of observation is known, a view-independent method is used to measure similarities. A metric is trained by Metric Learning to Rank (MLR) [14] with data of all views. MLR is a general metric learning algorithm based on structural SVM [8] framework and view metric learning as problem of information retrieval. Its purpose is to learn a metric so that data can be well ranked by distances. In this paper, a metric is trained to map data to another space in which data from the same person locate closer than those from different people. By this metric, we try to solve the view-angle change problem. Data from different views of the same person can be clustered in metric space, so that similarities between people can be measured simply as distances in the metric space, regardless of view angles. In training step, the training set is built as a feature set $X = \{F_i\}_1^n$ where n is the total number of people in training set. F_i is also a set that contains individual features of all views, as $F_i = \{\boldsymbol{f_{ij}}\}_{j=1}^v$ where v is the number of views which is 11 in CASIA dataset, $\boldsymbol{f_{ij}} = (x_{1ij}, x_{2ij}, ..., x_{mij})^T$ is feature of person i view j and m is the length of the feature. Then a normalization step is taken. Data are normalized as follows:

$$Z(X) = \frac{X - \mu}{S} \tag{4}$$

where μ is the mean of features, S is the standard deviation and $Z(X)$ is normalized data matrix. After data are ready, a metric $W \in \mathbb{R}^{m \times m}$ is trained by MLR. Here as parameters of MLR, the area under the ROC curve(AUC) [2] is adopted to be the ranking measure and the slack trade-off parameter is 0.01 in MLR training.

When use metric W to solve person re-identification matching problem, the distance can be computed as follows:

$$D(\boldsymbol{f_1}, \boldsymbol{f_2}) = (\boldsymbol{f_1} - \boldsymbol{f_2})^T W (\boldsymbol{f_1} - \boldsymbol{f_2}) \tag{5}$$

where $\boldsymbol{f_1}$ and $\boldsymbol{f_2}$ are features of probe person and gallery person. $D(\boldsymbol{f_1}, \boldsymbol{f_2})$ is the distance between $\boldsymbol{f_1}$ and $\boldsymbol{f_2}$, which would be small if the probe image and the gallery image are similar. In test step, we have a gallery set A and a probe set B in general. Associating each person of set B to all people of A is the purpose of person re-identification. So for each probe person b_i, distances between b_i and all gallery people are computed, then ranking is processed.

3.3 Fusion Strategy

In order to effectively use the GEI feature and appearance features, two fusion methods in our framework will be introduced in the following. The first one is

the score-level fusion that fuse distances which are calculated by GEI feature and the appearance feature, respectively. The second one is the feature-level fusion that installs two type features in series.

Score-Level Fusion. This fusion strategy is a view-independent global function, formulated as follows:

$$D_{FIN}(S_1, S_2) = D_{APP}(S_1, S_2) + D_{GEI}(S_1, S_2) \tag{6}$$

where $D_{FIN}(S_1, S_2)$ is the final relative distance of two samples get from two different views. $D_{APP}(S_1, S_2)$ is the distance that derived from the appearance feature and $D_GEI(S_1, S_2)$ is distance derived from GEI feature. The reason why just simply adding two distances without a weighting factor is that distances computed by metric are already optimized. Therefore, putting a factor into the equation is unreasonable which is also proved by experiment. If a weighting factor is added into equation as follows:

$$D_{FIN}(S_1, S_2) = D_{APP}(S_1, S_2) + \theta D_{GEI}(S_1, S_2) \tag{7}$$

where θ is the weighting factor. Repeat experiments with the value of θ range from 0.1 to 10, and the best result is obtained when the value of θ is 1.

Feature-Level Fusion. This strategy is to fusion two type features before calculating distance. Suppose we have an appearance feature vector $\boldsymbol{A} \in \mathbb{R}^m$, and GEI feature vector $\boldsymbol{G} \in \mathbb{R}^n$. Then these two features can be combined as follows:

$$\boldsymbol{F} = [\boldsymbol{A}, \boldsymbol{G}] \tag{8}$$

where the operator [X,Y] is defined as installing X and Y in series. $\boldsymbol{F} \in \mathbb{R}^{m+n}$ is the final feature vector.

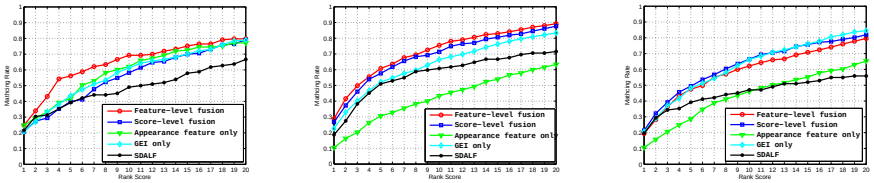
4 Experiment

4.1 Dataset

The proposed method is tested on the CASIA Gait Database B created by The Institute of Automation, Chinese Academy of Sciences (CASIA). This dataset contains eleven views of 124 individuals with three conditions. These three conditions are 'bag', which means the pedestrian appears with a bag in the video, 'clothes', which means the pedestrian appears with coat in the video and 'normal' means pedestrians appear without coat or bag. The view contains the angle of 0°, 18°, 36°, 54°, 72°, 90°, 108°, 126°, 144°, 162° and 180°. The CASIA dataset provides a background video for each view so that the foreground can be extracted conveniently.

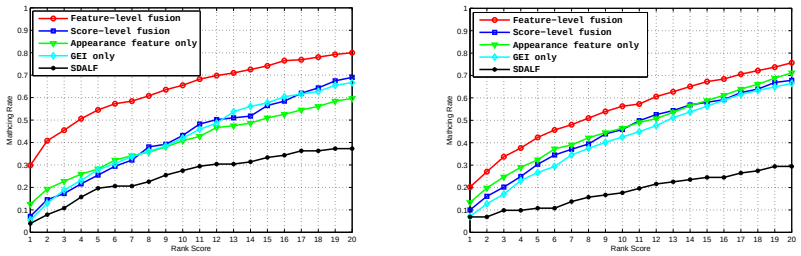
4.2 Results

In the first step of the experiment, only three original conditions are involved, in which people do not change their appearance. When learning the metric matrix, data of all views are used. For testing, the view of each probe image in probe set is randomly chosen, respectively, and the same random process is conducted to gallery set. Data are operated like that to fit the real application scenario of person re-identification most. Under these experiment constraints, all two type fusion strategies of features are implemented comparing to situations of only one type feature is used. SDALF method is also put into comparison. It is a very classical and effective appearance-based method for person re-identification problem. Those methods are tested on three conditions in CASIA dataset include 'bag', 'clothes' and 'normal'. The data are separated to training data and testing data, each contains half of randomly picked samples, and the same separation is adopted to all experiments.



(a) (b) (c)

Fig. 4. CMC curves on the single condition in CASIA dataset. (a) is on 'bag' condition, (b) is on 'clothes' condition and (c) is on 'normal' condition.



(a) (b)

Fig. 5. CMC curves on cross conditions. (a) is on 'bag-clothes' condition and (b) is on 'clothes-normal' condition.

The results shown in Fig. 4 and Fig. 5 primarily verify the availability of gait in this initial experiment. It shows that the GEI feature can well describe people integrating with appearance features. On 'bag' condition, the performance of our method with feature-level fusion is the best. On 'clothes' condition, our method with two fusion strategies are better than single-feature situations and on 'normal' condition, method with score-level performs a little better.

The next step of the experiment is implemented on two challenging cross conditions that include 'bag-clothes' and 'clothes-normal'. We try to challenge this harder task because in person re-identification problem, cameras are non-overlapped so that pedestrians have a certain probability to change their appearance. We would also like to indicate that gait can get rid of the limit of traditional appearance features and become a kind of robust features to enhance person re-identification. To this end, data that contain two appearance conditions are used to train a metric matrix. The probe set and the gallery set are built with different conditions. In other words, we try to match individuals from one view with one condition to another view in another condition. Testing views are also picked randomly and training set and testing set are separated randomly as well.

Table 1. Comparison of the average Matching Rate (%) on the single condition.

Rank	r = 1	r = 5	r = 10	r = 15	r = 20
Score-level	23.53	51.76	67.84	77.71	84.51
Feature-level	23.79	52.22	66.67	75.42	82.88
Appearance	21.41	48.01	64.67	74.15	82.25
GEI	15.27	33.68	50.50	60.25	68.58
SDALF	19.93	43.14	52.29	58.82	64.71

Table 2. Comparison of the average Matching Rate (%) on the cross condition.

Rank	r = 1	r = 5	r = 10	r = 15	r = 20
Score-level	8.53	27.94	44.51	57.25	68.43
Feature-level	25.00	48.43	60.88	70.69	77.84
Appearance	6.27	27.25	42.25	56.96	66.57
GEI	9.02	28.04	41.37	54.31	63.14
SDALF	5.39	15.20	22.55	28.92	33.33

As shown in Figure 6, our method with feature-level fusion is far beyond others. Results show that both in traditional person re-identification and cross-condition person re-identification, the fused feature has an acceptable perfor-

mance. Moreover, results show that the gait biometric can be an effective feature to enhance person re-identification.

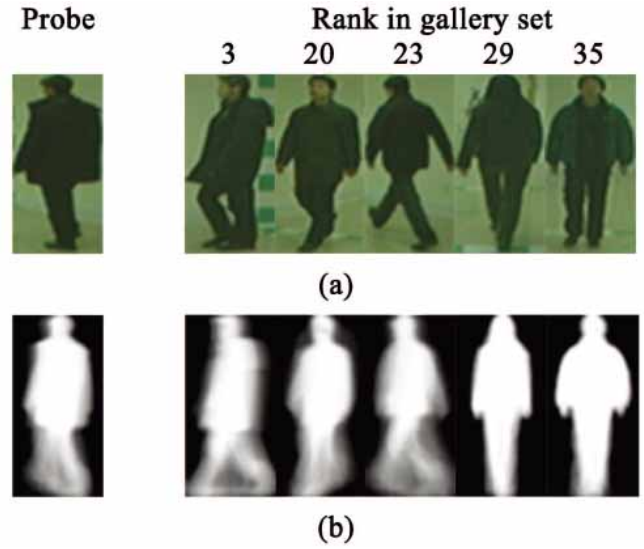


Fig. 6. Example of the effect of GEI feature. The appearance of Gallery images shown in (a) are similar to the probe image. They are all dressed in the black coat and cannot be distinguished by appearance features. Instead, Gait Energy Images shown in (b) are quite discriminative and help to re-identify the person in the gallery set.

The result in Table 1 shows a holistic perspective. In traditional application scene, fused features are as good as the appearance feature. On cross conditions, only the feature-level fusion method maintains the good performance. As shown from Table 1, neither the appearance feature nor the gait feature can handle the cross condition alone, which shows our method that integrating gait biometric with appearance features is effective.

5 Conclusion

The work of this paper is attempting to enhance person re-identification by integrating gait biometric. The proposed method contains hierarchical feature extraction and similarity measurement. Our experiments indicate that appearance features are not discriminative in some cases and other information should be integrated. Experiments show that gait biometric can be applied well in person re-identification. Even though GEI is not such discriminative, the attempt to enhance person re-identification by integrating gait biometric is successful in this paper. Our method is also verified as an effective way to combine gait bio-

metric to appearance features, and can enhance person re-identification in many conditions.

References

1. Bazzani, L., Cristani, M., Perina, A., Murino, V.: Multiple-shot person re-identification by chromatic and epitomic analyses. Pattern Recognition Letters **33** (2012) 898–903
2. Bradley, A.P.: The use of the area under the roc curve in the evaluation of machine learning algorithms. Pattern recognition **30** (1997) 1145–1159
3. Daugman, J.G.: Two-dimensional spectral analysis of cortical receptive field profiles. Vision research **20** (1980) 847–856
4. Farenzena, M., Bazzani, L., Perina, A., Murino, V., Cristani, M.: Person re-identification by symmetry-driven accumulation of local features. In: Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2010), San Francisco, CA, USA, IEEE Computer Society (2010)
5. Forssén, P.E.: Maximally stable colour regions for recognition and matching. In: Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on, IEEE (2007) 1–8
6. Gray, D., Tao, H.: Viewpoint invariant pedestrian recognition with an ensemble of localized features. In: Computer Vision–ECCV 2008. Springer (2008) 262–275
7. Han, J., Bhanu, B.: Individual recognition using gait energy image. Pattern Analysis and Machine Intelligence, IEEE Transactions on **28** (2006) 316–322
8. Joachims, T.: A support vector method for multivariate performance measures. In: Proceedings of the 22nd international conference on Machine learning, ACM (2005) 377–384
9. Lades, M., Vorbruggen, J.C., Buhmann, J., Lange, J., von der Malsburg, C., Wurtz, R.P., Konen, W.: Distortion invariant object recognition in the dynamic link architecture. Computers, IEEE Transactions on **42** (1993) 300–311
10. Little, J., Boyd, J.: Recognizing people by their gait: the shape of motion. Videre: Journal of Computer Vision Research **1** (1998) 1–32
11. Liu, C., Wechsler, H.: Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition. Image processing, IEEE Transactions on **11** (2002) 467–476
12. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. International journal of computer vision **60** (2004) 91–110
13. Marčelja, S.: Mathematical description of the responses of simple cortical cells*. JOSA **70** (1980) 1297–1300
14. McFee, B., Lanckriet, G.R.: Metric learning to rank. In: Proceedings of the 27th International Conference on Machine Learning (ICML-10). (2010) 775–782
15. Preisendorfer, R., Mobley, C.: Principal component analysis in meteorology and oceanography. Elsevier Science Ltd (1988)
16. Sarkar, S., Phillips, P.J., Liu, Z., Vega, I.R., Grother, P., Bowyer, K.W.: The humanid gait challenge problem: Data sets, performance, and analysis. Pattern Analysis and Machine Intelligence, IEEE Transactions on **27** (2005) 162–177
17. Zhao, R., Ouyang, W., Wang, X.: Unsupervised salience learning for person re-identification. In: Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on, IEEE (2013) 3586–3593