Swiss-System Based Cascade Ranking for Gait-Based Person Re-identification

Lan Wei[1], Yonghong Tian[1*], Yaowei Wang[2], Tiejun Huang[1]

[1]School of EE & CS, Peking University, Beijing 100871, China

[2]Department of Electronic Engineering, Beijing Institute of Technology, Beijing 100081, China

Corresponding to: yhtian@pku.edu.cn

Abstract

Human gait has been shown to be an efficient biometric measure for person identification at a distance. However, it often needs different gait features to handle various covariate conditions including viewing angles, walking speed, carrying an object and wearing different types of shoes. In order to improve the robustness of gait-based person re-identification on such multi-covariate conditions, a novel Swiss-system based cascade ranking model is proposed in this paper. Since the ranking model is able to learn a subspace where the potential true match is given the highest ranking, we formulate the gait-based person re-identification as a bipartite ranking problem and utilize it as an effective way for multi-feature ensemble learning. Then a Swiss multi-round competition system is developed for the cascade ranking model to optimize its effectiveness and efficiency. Extensive experiments on three indoor and outdoor public datasets demonstrate that our model outperforms several state-of-the-art methods remarkably.

Introduction

In recent years, surveillance cameras have been widely deployed almost everywhere in the city. To automatically analyze the data captured from these cameras (e.g., to recognize or search a suspicious person), some biometric technologies (e.g., face recognition) are developed and have shown more and more important roles in public security applications and crime investigation. However, one major disadvantage of these technologies is that they can only be used in some cooperative conditions (e.g., the front-face is mostly available). In some uncooperative situations, these biometric features may be obscured, preventing the effective recognition of the suspicious person. To address this problem, human gait can be used as an alternative, which is able to ascertain the identity of a person at a distance from a camera using behavioral features. An important application of gait recognition is person re-identification, which is defined as the task of assigning the same identifier to all the instances of the same object (Vezzani, Baltieri, and Cucchiara 2013). However, gait-based person re-identification often needs different gait features to handle various covariate conditions,
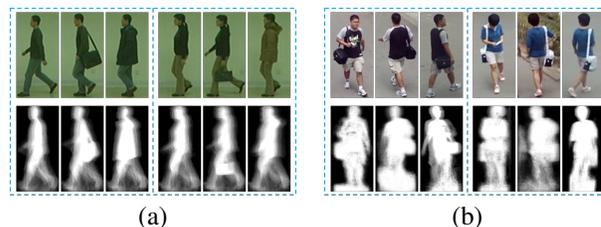
Figure 1: The changes in covariate conditions make the gait features less distinguishable. (a) The change of carrying condition and clothing, and (b) the change of viewing angles. The first row of each box represents different observations from the same subject with the second row representing the corresponding gait feature.

including viewing angles, walking speed, carrying an object and wearing different shoes. As shown in Fig. 1, due to the significant changes in covariate conditions, especially viewing angle, carrying condition and clothing, the resulting gait features of the same person differ greatly with each other.

Basically, the gait-based person re-identification problem tends to be formulated as a distance measurement problem, rather than a classification problem in which each person is treated as one class. This is not only due to the variable covariant conditions mentioned above, but also because of the difficulty of training extremely under-sampled class distribution (i.e., each person has only a few gallery sequences for training). Also, it is not reasonable that each newly-added subject must be re-trained every time for real world applications. Following the distance-measurement-based problem formulation, most existing approaches focus on extracting robust gait features. One of the most widely used gait features is Gait Energy Image (GEI) (Han and Bhanu 2006), which is obtained by averaging silhouettes across a gait cycle. However, it has been shown to be sensitive to various covariate conditions. To overcome this problem, a number of variations of GEI have been proposed. The basic idea is to select features from the most dynamic areas of human body. Yang et al. (Yang et al. 2008) propose to enhance those dynamic regions which are located by a variance analysis, and then extract enhanced GEI (EGEI) for recognition. Bashir et. al. (Bashir, Xiang, and Gong 2009) present a method to distinguish the dynamic and static areas of GEI by using Shannon entropy at each GEI pixel,

providing a new gait representation called GEnI. Lam et al. (Lam, Cheung, and Liu 2011) propose the Gait Flow Image (GFI) by using an optical flow field of the binary silhouette sequence. Instead of extracting from binary silhouette, Bashir et al. (Bashir et al. 2009) propose another optical flow fields based gait representation which is computed from the normalized and centred person images over a gait cycle. Their representation consists of a Motion Intensity Image (MII) and four Motion Direction Images (MDIs). None of the above methods can deal with large view changes, for which completely different models are developed. Most of these work aims to transform the gait features from one viewpoint to another by learning a View Transformation Model (VTM) (Worapan Kusakunniran and Zhang 2009). A different method is proposed by Bashir et al. (Bashir, Xiang, and Gong 2010a) which does not reconstruct gait features in different views, but models their correlation using Canonical Correlation Analysis (CCA) and uses the correlation strength as similarity measure. However, none of these view-specific models can address other covariate conditions. A first attempt under a more complex setting finds that when both the gallery and probe sets contain different and unknown covariate conditions, the performance of existing methods would drop dramatically (Bashir, Xiang, and Gong 2010b). Inspired by the success of using learning-to-rank in document retrieval (Joachims 2002) and computer vision (Han and Bhanu 2006), Martín-Félez and Xiang (Martín-Félez and Xiang 2012) formulate gait recognition as a bipartite ranking problem and use the PrRankSVM for recognition under a genuine uncooperative setting. Generally speaking, their work can achieve promising results in those scenes with one or two unknown covariate conditions settings. However, they may fail when the covariate conditions of scenes become extremely difficult and unpredictable, which can be seen from the following experiments in this paper. This is mainly because the over-simplified gait feature GEI used for learning, which contains very limited information of the walking patten.

To address the oversimplified gait feature problem, five such state-of-the-art gait features encoding both static, time-variant and motion information are extracted in this paper, including GEI, EGEI, GEnI, MII and MDIs. A problem raises up that, both distance-measurement-based method and learning-to-rank based method could not give good results with a simple mixture of different features, as instances with multi-features could not be learned well by one distance-measurement model or one linear rank model. Thus, a Swiss multi-round competition system of cascade ranking model is developed for multi-feature ensemble learning in this paper. In this system, a series of rankers are applied to every pair of matches between probe and gallery gait sequences. The initial ranker eliminates a large number of incorrect examples with very little processing, while the subsequent layers eliminate other incorrect ones by using more features at the cost of additional computation. Thus through such as multi-round competition, the effectiveness and efficiency of cascade ranking model can be improved remarkably. Extensive experiments are performed on three benchmarking datasets, including the in-
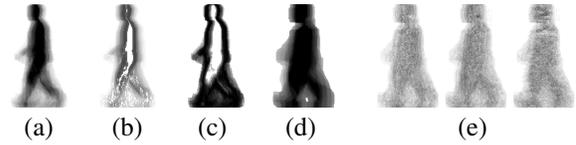

(a)　(b)　(c)　(d)　(e)

Figure 3: Examples of gait features. (a) GEI. (b) EGEI. (c) GEnI. (d) MII. (e) MDIs (three directions).

door CASIA database (CASIA 2005), the outdoor Soton Large dataset (Shutler et al. 2004), and the actual monitoring database (PKU 2014). The experimental results demonstrate that our method outperforms several state-of-the-art methods remarkably.

Gait-based Person Re-identification

Multiple Gait Features

The current trend in gait representation is appearance and period-based representation, such as the most popular gait feature GEI. GEI reflects a dynamic characteristic of a gait cycle. While, the intensity of specific subject's motion and the timevarying characteristics are hard to be obtained from GEI. As there's no one good feature meeting all demands existing, five state-of-the-art gait features are used in this study: GEI, EGEI, GEnI, MII and three MDIs. Some examples are shown in Fig. 3. GEI is obtained by averaging silhouettes over a gait cycle. EGEI constructs a dynamics weight mask to enhance the dynamic region and GEnI is computed by regarding the value of the GEI as the probability that the pixel takes the binary value. Unlike EGEI and GEnI both of which aim to select the intrinsic dynamic gait patterns, MII and MDIs are extracted over the optical flow fields. Among them, MII measures the intensity of the relative motion at each pixel location, and each MDI represents the likelihood along one specific motion direction during a complete gait cycle. The simple mixture of different features does not give good results with distance-measurement-based methods. Meanwhile, the ranking-based PrRankSVM tends to have difficulty in looking global convergence. Therefore, a cascade ranking method is future developed in our paper for multi-feature ensemble learning.

Gait Recognition by Ranking

Gait-based person re-identification is to solve the problem of retrieving the best matched sample of the given probe gait sequence from the gallery set. Given the gallery set $X = \{(x_g, y_g)\}_{g=1}^{|G|}$ and the probe set $X = \{(x_p, y_p)\}_{p=1}^{|P|}$ where x_g and x_p are the gait feature vectors, y_g is the known identity of x_g while y_p is the desired one. $|G|$ and $|P|$ are the numbers of samples in the gallery set and the probe set, respectively. Instead of using the gait features as the ranking features directly, we construct another feature space $X' = \left\{ \{(x_{p,g}, y_{p,g})\}_{g=1}^{|G|} \right\}_{p=1}^{|P|}$ to represent the relevant gait features of the pair of the gallery gait feature x_g and the probe gait feature x_p. Specifically ,

$$x_{p,g} = |x_p - x_g|, \quad y_{p,g} = \begin{cases} 1, y_p = y_g \\ 0, y_p \neq y_g \end{cases} \quad (1)$$
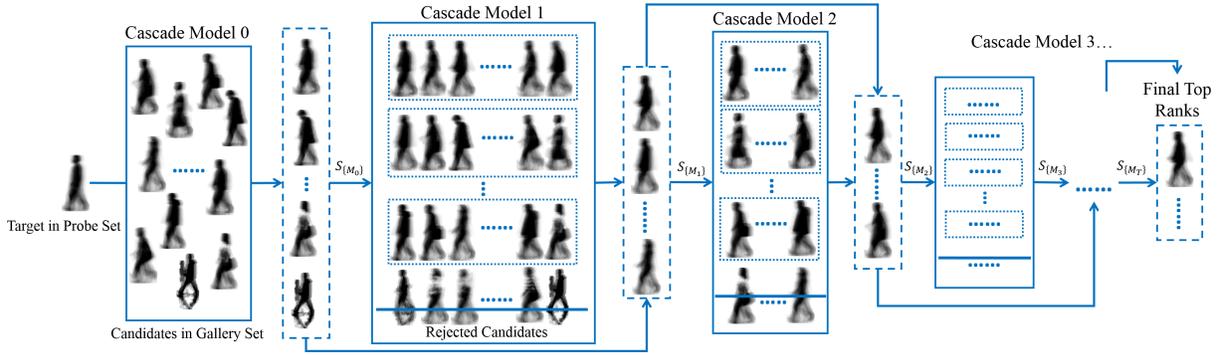
Figure 2: Schematic depiction of a the ranking cascade. A series of rankers are applied to every pair of match. The initial ranker eliminates a large number of incorrect examples with very little processing. The subsequent layers eliminate additional incorrect using more complex rankers. In this paper, the cascade model 0 is only an initialization of the ranking using NN, the subsequent cascade models use a SVM, and a winner-against-winner strategy is implemented.

Thus, the gait-based person re-identification problem can be seen as a ranking problem. For a specific probe gait feature, there exists a set of relevance ranks of the *possibly matched* gait sequences from the gallery set $G = \{x_{p,g_1}, x_{p,g_2}, \cdots x_{p,g_{|G|}}\}$, where $|G|$ is the number of ranks and $\succeq$ indicates the ordering. Note that there only exist two kinds of relevance ranks, namely, the correct and incorrect ranks/matches. Let $\hat{x}_{p,g}^{+} = x_{p,g}$ with $y_{p,g} = 1$ donates the correct ranks/matches, while $\hat{x}_{p,g}^{-} = x_{p,g}$ with $y_{p,g} = 0$ donates the incorrect ranks/matches. So our goal is to find a score model δ such that, for all ranks/matches $\hat{x}_{p,g}^{+}$ and $\hat{x}_{p,g}^{-}$, we have $\delta\left(\hat{x}_{p,g}^{+}\right) \succeq \delta\left(\hat{x}_{p,g}^{-}\right)$. In this case, the ranks between pairs that contain two matching gait features $\hat{x}_{p,g_i}^{+}$ and $\hat{x}_{p,g_j}^{+}$ (or two mismatching ones $\hat{x}_{p,g_i}^{-}$ and $\hat{x}_{p,g_j}^{-}$) do not matter.

Swiss-System Based Cascade Ranking

By formulating the gait-based person re-identification as a bipartite ranking problem, the recognition process can be seen as a competition which outputs the ranks among participants. While the participants have changing various covariate conditions, it is fairer to obtain ranks through multi-round competitions rather than only one competition. In the chess tournament, this process is called as *Swiss-system*. We thus use the similar terminology to formulate our gait multi-round ranking model. A Swiss-system tournament runs with a major principle that participants compete with each other according to their scores in the last competition. More specifically, the first round is either drawn at random or seeded according to some prior orders, and then all participants proceed to the next round in which winners are pitted against winners while losers are pitted against losers, and so on in subsequent rounds. Participants competing with winers will obtain a higher score in each round. Note that our problem differs slightly from the chess competition in that there is no guarantee that a higher ranking score in the first random round is really the top rank candidate. Subjects with the same covariance conditions appear more likely in the top k matches to the target. So the ranking scores from the first ranker is more used as grounds of grouping. For the following ranking round, the participants are divided into different

---

Algorithm 1 Swiss-system based cascade ranking algorithm

1: Initialize the ranking score using the first cascade model $M_0 = \langle \delta_0 \rangle$;
2: **for** $t = 1 \cdots T$ **do**
3:     Select a cascade model $M_t = \langle J_t, R_t, \delta_t \rangle$ over the remaining instances;
4:     Prune the last ranked instances with the pruning function J_t, based on the ranking scores $S_{\{M_{t-1}\}}$;
5:     Divide the instances into N_t groups, using the grouping function R_t ;
6:     **for** $r = 1 \ldots N_t$ **do**
7:         Output the ranking score $S_{\{M_t^r\}}$ with ranker δ_t^r;
8:         Balance each set of scores: $S_{\{M_t^r\}}\left(x_{p,g_i}\right) = $

$$S_{\{M_t^r\}}\left(x_{p,g_i}\right) - \frac{1}{N_t^r} \sum_{i=1}^{N_t^r} S_{\{M_t^r\}}\left(x_{p,g_i}\right);$$

9:     **end for**
10:     Combine the current score for N_t groups with the score in the last round;
11: **end for**

---

groups to rank respectively, before which a pruning function is applied to speed up. Scores from each group are balanced for each round to narrow the scoring gaps from different groups, just like the chess competition. And the results from former round is combined with scores of current round to output the rank of the current round.

As described in Algorithm 1, our cascade ranking model consists of a sequence of ranking models $\{M_0 \cdots M_T\}$, where each model $M_t = \langle J_t, G_t, \delta_t \rangle$ contains a set of rankers $\{S_t\}$. Each ranker S_t is associated with a pruning function J_t, a grouping function G_t and a local ranking function δ_t. Each stage, S_t receives the set of the ranked candidate matches from the previous stage as its input and then performs the following operations: firstly, the pruning function J_t is used to remove a number of candidate matches from the input set, which reduces the involved pairs in this stage); Secondly, the grouping function G_t is applied to divide the remaining matches to G_t groups. The group-

ing method follows the principle of Swiss-system that pairs with a higher score from the last round are grouped together and matches with a lower score are grouped together; After grouping, the score contribution of the local ranking function δ_t is added to each group of the candidate matches so as to improve the quality of top k matches. As for some rankers, such as PrRankSVM, the ranking scores give no practical meaning except for ranks. For instance, there is no difference between the group of ranking scores:$\{1, 2, 3\}$ and $\{100, 200, 300\}$. Therefore, a simple normalization method is applied to balance the scores of each group. Finally, the scores in this round are combined with the last round to generate a new rank which will be forwarded to the next stage for further pruning and re-ranking. Notice that the first cascade model does not contain the pruning function J_0 and the grouping function δ_0, as there is no pre-ranked information available. In this case, none instances will be pruned and no groups will be divided. For the cascade ranking model, the remaining problem is how to define the pruning functions, grouping functions and local rankers.

Pruning Functions. In (Yu, Tan, and Tan 2006), three pruning methods J_t are defined, including rank-based, score-based and mean-max threshold pruning functions. In our problem, however, the scores provided by the model has no practical meaning but is only used for ranking. Meanwhile, the mean and variances scores with different ranking model of each round differs from each others. Thus the rank-based pruning function is used here. With a cutoff value $Jc\left(\beta_t\right)$, the pruning function is defined as:

$$J_t : \begin{cases} Keep, & if \quad S_{\{M_{t-1}\}}\left(p_i\right) \geq Jc\left(\beta_t\right); \\ Pruned, & otherwise. \end{cases} \quad (2)$$

A match is pruned if it ranks below the cutoff $Jc\left(\beta_t\right)$, where β_t is the the pruning parameter and the cutoff value is:

$$Jc\left(\beta_t\right) = \left(1 - \beta_t\right) \left|S_{\{M_{t-1}\}}\right|, \quad (3)$$

where $\left|S_{\{M_{t-1}\}}\right|$ is the number of the inputs from M_{t-1}. Large values of β_t lead to less aggressive pruning. That is, $\beta_t = 1$ means that no matches will be discarded.

Grouping Functions. After pruning the least probable candidates, the remaining matches will be divided into N_t groups with the grouping function R_t. Based on the fact that the number of correct matches is far larger than the number of wrong matches (that is $N_{\{\hat{x}_{p,g}^+\}} \gg N_{\{\hat{x}_{p,g}^-\}}$, where $N_{\{\hat{x}_{p,g}^+\}}$ and $N_{\{\hat{x}_{p,g}^-\}}$ are the numbers of correct and wrong matches), we use a novel greedy algorithm for grouping to avoid that there exists a group without any correct match. The grouping algorithm is described in Algorithm 2. Here N_t is the pre-defined parameter of how many groups to be divided. A large N_t is not advised as the too many groups with little data in each group is not good for training.

NN & SVM Ranker. We now turn to the problem of learning a well-chosen local ranker. Note that the ranker used in the first cascade model should be a no-training model. For simplicity, nearest neighbor (NN) is adopted in our study as δ_0. As mentioned before, the sample

---

Algorithm 2 The greedy algorithm for grouping
1: Based on the number of correct matches $N_{\{\hat{x}_{p,g}^+\}}$, divide the ranked instances into $N_{\{\hat{x}_{p,g}^+\}} + 1$ subsets:

$$\left\{ S'_1, S'_2, \cdots S'_{N_{\{\hat{x}_{p,g}^+\}}+1} \right\} =$$
$$\left\{ \left\{ \cdots S_{\hat{x}_1^+} \right\}, \left\{ S_{\hat{x}_1^+ +1} \cdots S_{\hat{x}_2^+} \right\} \cdots \left\{ S_{N_{\{\hat{x}_{p,g}^+\}}+1} \cdots \right\} \right\}$$

2: **for** $k = 1 \ldots \left(N_{\{\hat{x}_{p,g}^+\}} + 1 - N_t\right)$ **do**
3:    Select the subset with minimum instances, denoted by S'_i;
4:    Combine S'_i with the smaller adjacent subset:
$$S'_i \leftarrow \begin{cases} S'_{i-1} \cup S'_i, & if \quad N_{\{S'_{i-1}\}} \geq N_{\{S'_{i+1}\}}; \\ S'_i \cup S'_{i+1}, & otherwise. \end{cases}$$
5:    Update the size of each subset;
6: **end for**

---

space has been transferred from $X = \{(x_p, y_p)\}_{p=1}^{|P|}$ to $X' = \left\{ \{(x_{p,g}, y_{p,g})\}_{g=1}^{|G|} \right\}_{p=1}^{|P|}$. Thus, the matches should be ranked according to the similarity (or distance) between the feature and a zero vector, which means 0 gets the highest score. The NN ranker is formulated as:

$$\delta_0\left(x_{p,g}\right) = 0 - \|x_{p,g}\| \quad (4)$$

As for the rankers in the subsequent cascade model $\delta_t, t \neq 0$, rather than the simple distance-measurement-based ranker, a more complicated training strategy should be used to learn the internal feature relationship of each group. Meanwhile, using the NN ranker again will not change the ranks any more. Therefore, PrRankSVM is chosen in this study since it is suitable for a large-scale learning problem with a severely-overlapped feature space (Chapelle and Keerthi 2010). The scoring function is formulated as:

$$\delta_t\left(x_{p,g}\right) = w^T x_{p,g}, t \neq 0 \quad (5)$$

Then by going through all $\hat{x}_{p,g}^+$ and $\hat{x}_{p,g}^+$ in dataset X', we obtain a set of all pairwise relevant difference vectors which meet $w^T\left(\hat{x}_{p,g}^+ - \hat{x}_{p,g}^-\right) > 0$. A PrRankSVM model is then defined as the minimization of the objective function:

$$\frac{1}{2}\|w\|^2 + C \sum_{i=1}^{|X'|} \xi_i \quad (6)$$
$$s.t. \quad w^T\left(\hat{x}_{p,g}^+ - \hat{x}_{p,g}^-\right) \geq 1 - \xi_i, \xi_i > 0$$

where i is the index of the preferred match, $|X'|$ is the total number of the preferred pairs used for training, C is a positive importance weight on the ranking performance and ξ is the hinge loss function used in SVM. It will lose the global ranking information from previous rounds if just the ranking results from the last round are used. For example, a contestant with "poor strength" in previous rounds may give a good result in the current ranking round as it's ranked in the "loser-to-loser" group. So it's more reasonable to combine the ranking scores from the previous cascade and current cascade. Hence we combine the ranking scores as follow:

$$S_{\{M_t\}}\left(x_{p,g}\right) = \alpha_t S_{\{M_{t-1}\}}\left(x_{p,g}\right) + \left(1 - \alpha_t\right)\delta_t\left(x_{p,g}\right), \quad (7)$$
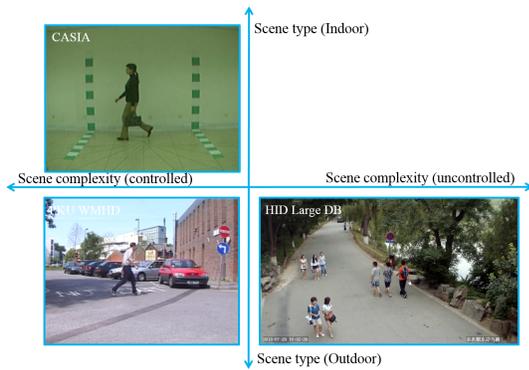
Figure 4: Visualizing the CASIA, Soton and PKU datasets.

where α_t is the parameter to weight scores from the two models. The smaller value of α_t, the less contribution the last cascade model made, i.e., $\alpha_t = 0$ means the re-ranking result fully depends on the current model.

Experiments

Experimental Settings

Extensive experiments have been carried out on the three gait databases: CASIA, Soton and PKU HumanID. As shown in Fig. 4, they cover an indoor environment (CASIA), outdoor (Soton) and no controlled scenario (PKU). All experiments are repeated ten times with different training/testing splits to mitigate the effects of subset singularities. The average gait recognition performance from these different trials are displayed on Cumulative Match Score (CMS) curve, which shows the percentage of the probe set whose identity has been correctly matched in the gallery among the top k matches. Our model is compared with six state-of-the-art approaches: Baseline (Yu, Tan, and Tan 2006), GEnI (Bashir, Xiang, and Gong 2009), EGEI+Gabor+DCV (Yang et al. 2008), fusion of MII and three MDIs (Bashir, Xiang, and Gong 2009), GFI (Lam, Cheung, and Liu 2011) and the newly proposed PrRankSVM (Martín-Félez and Xiang 2012).

Experimental Results on the CASIA Dataset

The CASIA Gait Database B contains 124 subjects captured under three different covariate condition changes: carrying, clothing, and view angle. Each subject ia captured from 11 different view angles. For each view, each subject has 10 gait sequences: six normal ones (*NM*), two carrying-bag (*BAG*) and two wearing-coat (*COAT*). All the videos are recorded indoor with a uniform background and controlled lighting. The 124 subjects are first randomly divided into training set and testing set for each experiment. Two sets of experiments are carried on the CASIA database. The first one (Experiment *BGCT*) focus on evaluating the different approaches under carrying and clothing condition changes. In these experiments, only side view (90°) gait sequences are used as the effect of view is investigated in another experiment. Among the 10 side-view sequences available for each subject, two *NM* ones out of six are randomly selected along with the two *COAT* and the other two *BAG*, for training or testing. The second set of experiments (Experiment
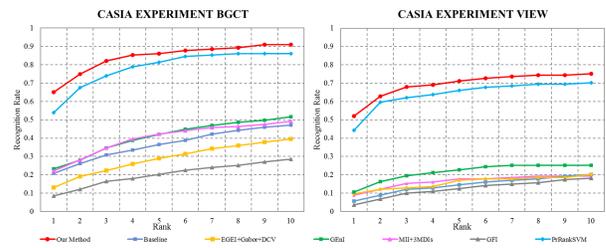

Figure 5: Experimental results on the CASIA dataset.

VIEW) is designed to evaluate our model under large view angle changes. For each possible pair, $(90°, \theta_i)$ are randomly picked for training or testing according to the subject. We use the data sequences under 0°, 36° and 72° for this set of experiments. Thus in the training set, there exist eight sequences of each subject. In the testing test, any of the two sequences from 90° and from θ_i is used as the gallery set, while the remaining other six sequences are used as the probe set.

The results are shown in Fig. 5. Due to the intrinsic difficulty of objects appearing in different wearing and viewing conditions, the recognition methods based on the covariate condition-invariant feature (e.g. EGEI, GEnI) perform much poorer than the ranking-based methods on average. As we have analysed before, the invariant gait features cannot handle the problem of tough uncooperative settings. On the other hand, our model performs better than PrRankSVM because the used multiple features extract more dynamic information that is invariant to covariate conditions and the cascade ranking model extracts the inner ranking relationship by dividing data into sets of groups. The performance of PrRankSVM drops noticeably compared with the results reported in (Martín-Félez and Xiang 2012) which only tests with one changing condition. This is unsurprising that with mixture of more different and unknown covariate conditions, the difficulty of problem raises.

Experimental Results on the Soton Large Dataset

The Soton Large dataset is a part of the Soton database (Shutler et al. 2004). It contains 116 subjects captured in both indoor and outdoor environment. This dataset has 6 subsets, *SotonSetsetA* to *SotonSetsetF*. In our experiments the outdoor sets *SotonSetsetE* and *SotonSetsetF* are used. Unlike other indoor datasets, there are no foreground walking silhouettes provided, which increases the difficulty of silhouettes-based gait recognition. Both the two sets contain two directions sequences: *L* means that the subject is walking towards the left side of the scene while *R* means that the subject is walking towards the right side of the scene. In our experiments, we randomly select one of the two direction sequences as the gallery set while the other as the test set. Moreover, in each set, half of the people are randomly selected for training and the rest for testing. For the testing, half sequences for each subject are used as the gallery set while the rest as the probe set. The results are shown in Fig. 6. For this set of experiments, walking direction poses the main changing covariate condition. Each subject's walking direction in the test sequences is always different from
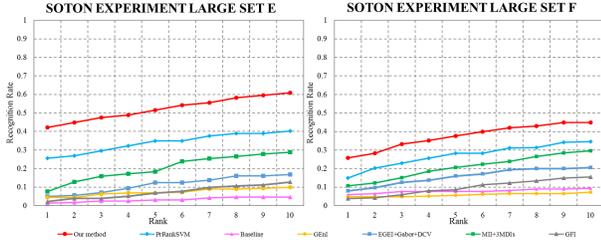
Figure 6: Experimental results on the Soton large dataset.

Figure 7: Experimental results with PKU dataset.

that in the gallery data. Baseline method based on GEI and direct L1-norm measurement gains the poorest performance since in the gallery data with the same walking direction always appears to be the best match. Other features such as EGEI and GEnI have a limited improvement in recognition accuracy compared with the baseline as it cannot deal with big difference of viewing angles. Meanwhile, the fusion of MII and MDIs gets a better result as it enhances the motion information in flow fields. Consistent with the results on CASIA dataset, ours and PrRankSVM performs better than others clearly. And the gaps between our method with PrRankSVM shows the significant improves of the usage of multi-features and cascade ranking model.

Experimental Results on the PKU Dataset

The PKU HumanID Gait Database (PKU) is an outdoor database captured in complete uncontrolled environment. It is composed of videos of 18 labeled subjects crossing 12 cameras in a campus. The subjects are all masked with bounding box by manually. Usually, one subject walks alone with a pre-designed walking route in other gait databases. However, the subjects in the PKU database walk totally freely and unpredictably. Moreover, other pedestrians except for the subjects are allowed to show up in the cameras. In this case, the occlusion of these unidentified pedestrians also increases the difficulty of the person re-identification. The major changing conditions include the free walking route of each subject, various camera setting, background, capture time and changing lighting. Two cameras *WMHD* and *YTX* are tested. In *WMHD*, every labeled subject has three sequences: two frontal views and one back view. For each subject, the temporal nodes that overlap the first frontal view sequences are used as the gallery data, and the other nodes are used as the probe data. As for *YTX*, walking sequences in two back views and one frontal view are captured. The temporal nodes covering the first back view sequence are taken as the gallery data, and the others are taken as probe data. Our experiments on the PKU dataset focus on the performance of cross-database recognition, for which the model is already well trained on the first dataset and then evaluated on another dataset. We use the data from the CASIA dataset for training and the data from the PKU dataset for testing.

Since the PKU databse has a limited number of subjects, each with some long walking sequences, the multi-object tracking measurement rather than CMS is adopted for the performance evaluation. This is because some methods may fail to extract a gait cycle, but yields a higher accuracy with
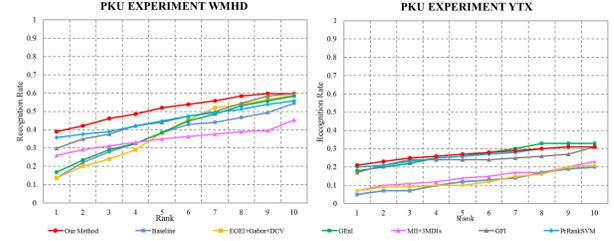
smaller evaluation base. For each camera $c\,(c = 1, \cdots C)$, C_n nodes are set in the time domain. Each node n_j^c has a temporal range (such as 200 frames from the beginning to end). For a specified target i in camera c during time t_j to t_{j+1}, if the recognition result coincides with the ground-truth, the result will be recognized to match the node. A set of gallery nodes and a set of probe nodes are predetermined for each target. With $R_{t,c}^i$ and $N_{t,c}^i$ representing the correctly matched notes and all labeled probe node for subject i under camera c, the precision metrics is formulated as:

$$P_n = \sum_{i=1}^{M} \sum_{t=1}^{R_c} R_{t,c}^i / \sum_{i=1}^{M} \sum_{t=1}^{N_c} N_{t,c}^i \qquad (8)$$

where M is the number of subjects, R_c is the number of correctly matched nodes and N_c is the number of all labeled probe nodes. As the model is learned on another dataset, the performance of the cross-database experiments drops than that of the the same-database. This is unsurprising because the covariate condition changes of CASIA and PKU are obviously different. In comparison, our ranking still gives better results under this challenging dataset as it learns transferable information to cope with the different camera. The performance of all methods with *YTX* are poor because the occlusion from other pedestrians increases the difficulty of gait period analysis, leading to no gait feature extracted during the corresponding temporal nodes. Nevertheless, with our rank 1 result of 21%, this does give an indication that our model is superior even with cross-database recognition.

Conclusions

In this work, we present a novel cascade ranking model based on Swiss-system for gait-based person re-identification. We have shown that with cascade grouping ranking and multi-features, the proposed model can improve the accuracy remarkably on indoor, outdoor and real monitoring gait databases. In future work, we will investigate how to build the multi-class ranking model rather than one ranker for all so as to improve the robustness to the circumstances.

Acknowledgments

References

Bashir, K.; Xiang, T.; Gong, S.; and Mary, Q. 2009. Gait representation using flow fields. In *BMVC*, 1–11.

Bashir, K.; Xiang, T.; and Gong, S. 2009. Gait recognition using gait entropy image. In *ICDP*, 1–6. IET.

Bashir, K.; Xiang, T.; and Gong, S. 2010a. Cross view gait recognition using correlation strength. In *BMVC*, 1–11.

Bashir, K.; Xiang, T.; and Gong, S. 2010b. Gait recognition without subject cooperation. *Patt. Recogn. Lett.* 31(13):2052–2060.

CASIA. 2005. Casia gait database. http://www.sinobiometrics.com.

Chapelle, O., and Keerthi, S. S. 2010. Efficient algorithms for ranking with svms. *Inform. Retrieval* 13(3):201–215.

Han, J., and Bhanu, B. 2006. Individual recognition using gait energy image. *PAMI* 28(2):316–322.

Joachims, T. 2002. Optimizing search engines using click through data. In *KDD*, 133–142.

Lam, T. H.; Cheung, K. H.; and Liu, J. N. 2011. Gait flow image: A silhouette-based gait representation for human identification. *Patt. Recogn.* 44(4):973–987.

Martín-Félez, R., and Xiang, T. 2012. Gait recognition by ranking. In *ECCV*. 328–341.

PKU. 2014. Pku human id database. http://mlg.idm.pku.edu.cn/resources/dataset.html.

Shutler, J. D.; Grant, M. G.; Nixon, M. S.; and Carter, J. N. 2004. On a large sequence-based human gait database. In *Appl. & Sci. in Soft Comput.* 339–346.

Vezzani, R.; Baltieri, D.; and Cucchiara, R. 2013. People reidentification in surveillance and forensics: A survey. *AC-SUR* 46(2):29.

Worapan Kusakunniran, Qiang Wu, H. L., and Zhang, J. 2009. Multiple views gait recognition using view transformation model based on optimized gait energy image. In *ICCV Workshops*, 1058–1064.

Yang, X.; Zhou, Y.; Zhang, T.; Shu, G.; and Yang, J. 2008. Gait recognition based on dynamic region analysis. *Sign. Proc.* 88(9):2350–2356.

Yu, S.; Tan, D.; and Tan, T. 2006. A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition. In *ICPR*, volume 4, 441–444. IEEE.