# Gait Recognition by Ranking

Raúl Martín-Félez[1] and Tao Xiang[2]

[1] Institute of New Imaging Technologies, Universitat Jaume I, Castelló 12071, Spain
[2] School of EECS, Queen Mary, University of London, London E1 4NS, U.K.
martinr@uji.es, txiang@eecs.qmul.ac.uk

**Abstract.** The advantage of gait over other biometrics such as face or fingerprint is that it can operate from a distance and without subject cooperation. However, this also makes gait subject to changes in various covariate conditions including carrying, clothing, surface and view angle. Existing approaches attempt to address these condition changes by feature selection, feature transformation or discriminant subspace learning. However, they suffer from lack of training samples from each subject, can only cope with changes in a subset of conditions with limited success, and are based on the invalid assumption that the covariate conditions are known a priori. They are thus unable to perform gait recognition under a genuine uncooperative setting. We propose a novel approach which casts gait recognition as a bipartite ranking problem and leverages training samples from different classes/people and even from different datasets. This makes our approach suitable for recognition under a genuine uncooperative setting and robust against any covariate types, as demonstrated by our extensive experiments.

**Keywords:** Gait recognition, Learning to rank, Transfer learning.

## 1  Introduction

Gait can be used as a behavioral biometric. Compared to physiological biometrics such as fingerprint, iris and face, it has a number of distinctive pros and cons. The key advantage of gait for person identification is that it can operate from a distance and without subject cooperation. This makes gait ideal for situations where direct contact with or cooperation from a subject is not possible, *e.g.* surveillance in a public space. However, having uncooperative subjects also means that gait is susceptible to various covariate conditions, which are circumstantial and physical conditions that can affect either gait itself or the extracted gait features. Some examples are clothing, surface, load carrying, view angle, walking speed, and shoe-wear type. Figure 1 shows that due to significant changes in covariate conditions, in particular view angle and clothing, gait features of different people can be more alike than those of the same person.

As a classification problem (*i.e.* each person is a class), gait recognition is challenging. This is not only due to the variable covariate conditions mentioned above, but also because of a lack of training data to cope with the large overlap between classes in a feature space. Specifically, each subject may be captured
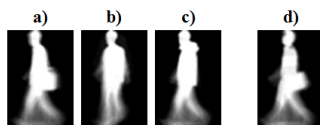
**Fig. 1.** Comparison of Gait Energy Images [9] of two people A (a:with a bag , b:a different view, and c:wearing a bulking coat) and B (d:with a bag). Among b, c, and d, d appears to be the best match to a because they share the same covariate conditions (view, carrying and clothing), which leads to a wrong match.

only in one sequence with a handful of gait cycles for feature extraction, resulting in an extremely under-sampled class distribution. Most existing approaches focus on extracting and selecting the best gait features that are invariable to different conditions [3,17,20]. However, they are based on human a priori knowledge (*e.g.* the most reliable features are in the most dynamic part of human body, *i.e.* legs) and select features in the highly overlapped original feature space, which only lead to very limited success. In addition, these methods are designed for addressing specific types of covariate conditions. Particularly, none of them can cope with large view angle changes, for which completely different approaches based on feature transformation [11,13] are developed. They in turn do not work on other covariate conditions. So far, none of the existing approaches can address all covariate conditions which typically co-exist under an uncooperative setting.

Different from those feature selection and transformation based methods, some learning based approaches are also proposed [9,12,18] attempting to maximise the inter-class distance whilst minimising intra class variations. These learning based approaches can be applied after feature selection/transformation. However, they assume that the same classes/people must be present in both the training and test sets and represented with enough samples. Both assumptions are often not valid in practice. More importantly, most existing works use a gallery set composed of gait sequences under similar covariate conditions and evaluate their performance on probe sets of possibly different but fixed covariate conditions. They therefore make the implicit assumption that the gallery data are collected in a cooperative manner so that the covariate conditions are known a priori. This essentially deprives gait of its most useful characteristic as an uncooperative and non-intrusive biometric.

In this paper, a novel approach is proposed which casts gait recognition as a ranking problem - a completely different perspective from previous approaches. More specifically, given a training and a test datasets consisting of gait features of different people who may even be captured from a completely different scene, we learn a bipartite ranking model based on the Primal RankSVM [6] algorithm. The model aims to learn a ranking function in a higher dimensional space where true matches and wrong matches become more separable than in the original space. The output of the model is a ranking function which gives a higher score if a pair of gait feature vectors belong to the same person than to different people. This new formulation has three distinctive advantages over the previous ones: 1) this model is data-driven and can address all covariate conditions

including view, *i.e.* one model for all; 2) critically, it does not make any assumption about the gallery and/or probe sets having the same covariate condition, either within each set or across the sets, what makes it particularly suitable for uncooperative settings, where gait should be used; 3) it does not suffer from the class under-sampling problem. As it is based on bipartite ranking, there are only two classes during training: true matches and wrong matches. This means that gait features from different people captured in different scenes/datasets can be used for training. In essence, it performs cross-class and cross-dataset transfer learning and is able to learn from an auxiliary dataset where plenty of data might be available. Extensive experiments have been conducted using the two largest public benchmarking datasets covering both indoor and outdoor environments. The results suggest that our approach drastically outperforms existing approaches, achieving up to 14-fold increase in recognition rate under the most difficult uncooperative settings.

**Related Work** – Most existing gait recognition techniques extract information from silhouettes obtained from video sequences. One of the simplest yet effective representations is Gait Energy Image (GEI) [9], which is obtained by averaging silhouettes across a gait cycle. However, it has been shown to be sensitive to various covariate conditions [9,21]. 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, *i.e.* legs and arms, to cope with changes in carrying conditions, clothing, and surface. Yang *et al.* [20] propose to enhance those dynamic regions which are located by a variance analysis. Bashir *et al.* [3] present a method to distinguish the dynamic and static areas of GEI by using Shannon entropy at each GEI pixel, giving as a result a new gait representation called GEnI. Given a gait representation, classification by template matching (*i.e.* one-nearest-neighbour (1NN)) can be performed, but alternatively, a number of discriminative learning based methods can be adopted to maximise the inter-class distance whilst minimising the intra-class one. They range from Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) [9], to general tensor discriminant analysis (GTDA) [18].

None of the above methods can deal with large view changes, for which completely different models are developed. Most of these works aim to transform the gait features from one viewpoint to another one by learning a View Transformation Model (VTM), *i.e.* recognition is performed after different views have explicitly been transferred to the same [11,13]. A different method is proposed by Bashir *et al.* [2] 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.

All the aforementioned works use an experimental setting under which covariate conditions are a priori known and all subjects in the gallery set have the same covariate conditions, so do those in the probe set. This essentially requires cooperative subjects, with which it is well known that gait is much weaker than other alternative biometrics such as face. It is possible to apply some of the

existing techniques under an uncooperative setting provided that the covariate condition of each gait sequence can be estimated reliably and different covariate condition changes do not co-occur. However, in practice neither of the two assumptions are valid: different covariate condition changes typically co-exist and estimating them is far from being solved in an unconstrained environment even for the easier ones such as view [1] and carrying condition [7], let alone others such as clothing and shoe-wear type. A first attempt under a more realistic setting [3] discovered that when both the gallery and probe set contain a mix of different and unknown covariate conditions, the performance of existing approaches drops dramatically. This is also verified by our experiments (see Section 3). More importantly we also demonstrate that our approach performs equally well given either a cooperative or an uncooperative setting, and it is able to deal with any covariate condition changes including view (unlike [3]) without the need to explicitly estimate them. This makes our method more suitable for gait recognition under a genuine uncooperative setting.

To the best of our knowledge, no previous work has formulated gait recognition as a ranking problem. Our method is inspired by the success of using learning to rank in document retrieval [10] and computer vision [15]. There exist other ranking models such as RankBoost [8], but RankSVM is chosen because it is more suitable for a large scale learning problem with a severely overlapped feature space. Our model can be considered as a transfer learning model as it can be learned using other classes of either the same or different scene/site. Recently, cross-domain [19] or cross-dataset [4] transfer learning have received an increasing interest in computer vision. However, this is the first work on formulating transfer learning based on a ranking model for gait recognition.

## 2   A Ranking Model for Gait Recognition

Given a gallery set of gait sequences of people with known identities, the problem of gait recognition can be considered as a retrieval problem: given a probe gait sequence, the best matched sample in the gallery set is retrieved whose identity is then assigned to the probe one. It is within the context of content-based document retrieval where the concept of learning to rank was first introduced. We thus use similar terminology to formulate our gait ranking model.

A bipartite ranking model is learned using a training set $T$ comprising:

- a number of query documents $Q = \{q_1, q_2, \cdots, q_{|Q|}\}$, with $|Q|$ representing the number of elements in $Q$.
- for each query $q$, a set of retrieved documents $D(q) = \{d_1, d_2, \cdots, d_{|D(q)|}\}$, with $|D(q)|$ representing the number of elements in the set $D(q)$.
- each pair $(q_i, d_j) \in Q \times D(q)$ is represented by a feature vector $x(q_i, d_j) \in \mathbb{R}^{\mathbb{M}}$.
- each document $d_j$ is labelled by a relevance indicator/score $y(q_i, d_j)$ which represents its relationship to the query $q_i$. A higher relevance score means that the document is more relevant to the query and must be ranked at the top by the learned model. The score can also be binary indicating whether or not the document $d_j$ is relevant to the query $q_i$.

The bipartite ranking model aims to learn the ranking score function defined in Eq. 1 for each pair of query $q$ and document $d$:

$$\delta(q,d) = \omega^T x(q,d) \tag{1}$$

where $\omega$ refers to a weight vector which indicates the importance of each feature towards the ranking score returned by $\delta$. The ranking function should achieve the best agreement with the ranking induced by the relevance indicators of the documents.

In the test phase, given a query $q_t$, the learned ranking score function $\delta$ is used to assign a score to each document $d$ according to its relevance to $q_t$. Then the documents $D(q_t)$ are sorted in the descending order of their assigned scores to obtain a ranked list of documents.

The above ranking model for document retrieval can be extrapolated to our problem in the following way to reformulate gait recognition as a ranking problem. Given a query sequence $q$ of a walking subject $s$, which might be affected by some unknown covariate conditions, we wish to find those gait sequences $D(q)$ that are more relevant to $q$, regardless the type of covariate conditions that they might involve. Among them, we focus on the highest ranked sequence $d_j$, which hopefully would belong to the same subject $s$.

To this end, a training set $T$ composed of gait sequences from different subjects is defined, and each sequence is used as query against all the remaining samples. We define a binary relevance indicator $y$ which assumes a value 1 when a gait sequence $d$ belongs to the same subject as a query sequence $q$, and 0 otherwise. Thus, for each training sequence $q$, all the other training samples are assigned to one of the following two sets depending on its relevance indicator with $q$:

- $D(q)^+ = \left\{ d_1^+, d_2^+, \cdots, d_{|D(q)^+|} \right\}$, with $y(q, d_i^+) = 1$ for all $d_i^+ \in D(q)^+$
- $D(q)^- = \left\{ d_1^-, d_2^-, \cdots, d_{|D(q)^-|} \right\}$, with $y(q, d_i^-) = 0$ for all $d_i^- \in D(q)^-$

Let $|D(q)^+|$ be the number of true matches (relevant sequences) for the query sequence $q$. Likewise, $|D(q)^-|$ refers to the number of wrong matches for $q$. Note that $|D(q)^+| << |D(q)^-|$ because, given a sequence $q$ belonging to a subject $s$, only a few sequences of $s$ are available in the training set (as few as one), while all the sequences from other training subjects are considered as wrong matches.

For our problem, the feature vector $x(q,d)$ in Eq. 1 is the absolute difference vector between the feature vectors $x_q$ and $x_d$ representing the gait sequences $q$ and $d$ respectively. The goal is to learn a ranking model $\delta$ such that, for all pairs $(q, d_i^+)$ and $(q, d_j^-)$, we have $\delta(q, d_i^+) > \delta(q, d_i^-)$.

Now, after using every single sequence as query in turn, we obtain a set of preference pairs $P = \{(\hat{D}^+, \hat{D}^-)\}$, where $\hat{D}^+ = |x_{q_i}, x_{d_j^+}|$ and $\hat{D}^- = |x_{q_i}, x_{d_j^-}|$, by going through all queries $q_i$ as well as their corresponding $D(q_i)^+$ and $D(q_i)^-$. $P$ is then used to learn the ranking function in Eq. 1, which is used in the test phase to rank all the gallery samples with respect to a given probe query sample $q_t$. If the sample at the top of the rank belongs to the same person as the query $q_t$, it is considered a true match, otherwise, a wrong match.
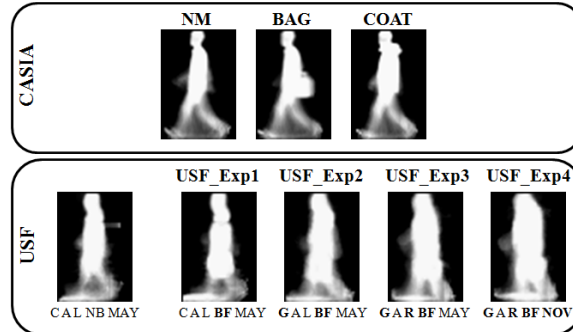
**Fig. 2.** Examples of GEIs from different people in CASIA (top) and USF (bottom) with different covariate conditions (see text)

Although any ranking method could be used under this approach, the primal-based pairwise RankSVM algorithm (PrRankSVM) proposed by Chapelle and Keerthi [6] is employed because it is able to cope with highly overlapped feature spaces with a large scale imbalanced number of positive and negative samples, exactly the problems faced by learning a gait ranking model. This method learns a ranking function in a higher dimensional feature space where true and wrong matches become more separable than in the original feature space. Specifically, it aims to solve the following optimisation problem:

$$\omega = argmin_\omega \frac{1}{2} \|\omega\|^2 + C \sum_{t=1}^{|P|} \ell \left( \omega^T \left( \hat{D}^+ - \hat{D}^- \right) \right) \tag{2}$$

where $t$ is the index of the preference pairs, $|P|$ is the total number of preference pairs used for training, $C$ is a positive importance weight on the ranking performance and $\ell$ is the hinge loss function such as $\ell(t) = max(0, 1 - t)^2$. The algorithm is computationally efficient, because it uses a Newton optimisation to solve the non-constraint model of Eq. 2. It allows to remove the explicit computation of the $(\hat{D}^+ - \hat{D}^-)$ pairs by using a sparse matrix. In this work, the $C$ parameter is automatically selected by cross validation on the training set.

## 3 Experiments

### 3.1 Experimental Settings

Extensive experiments have been carried out on the two largest benchmarking gait databases, CASIA [5] and USF HumanID [16], which cover both an indoor environment (CASIA) and an outdoor and less controlled scenario (USF).

**Gait Features.** For gait representation, Gait Energy Image (GEI) [9] is used because it has also been adopted by most existing works on covariate invariant

gait recognition. Each GEI is normalised to a fixed size of $64 \times 44$ pixels using the silhouettes provided by both databases. Example GEIs are shown in Figure 2, which clearly show the more challenging nature of the outdoor environment in USF indicated by much more noisy silhouettes.

**Uncooperative Setting.** All experiments are designed following an uncooperative setting as follows unless otherwise stated. Firstly, the whole set of sequences considered in a particular experiment is randomly and equally split into two subsets, one for training and the other for testing according to the subject, *i.e.* the subjects in the two sets are different. Secondly, the test set is further divided into a gallery set and a probe set in such a way that 1) each subject has at least a different covariate condition across the two subsets, and 2) both the gallery and probe have a mix of different covariate conditions. It is a challenging setting because for each probe sequence $q$ of a subject $s$ with a covariate type $k$, the gallery only contains sequences of the same subject $s$ with a different covariate condition type, and plenty of other subjects with the same covariate type $k$. For example, if Figure 1a was in the probe set, the gallery set would contain Figures 1c and 1d. All experiments are repeated five times with different training/testing splits to mitigate the effects of subset singularities.

**Competitors.** Three commonly used gait recognition methods are compared in all experiments. They are: 1) *1NN*. It uses the 1NN classifier with the original high dimensional GEI feature space (64x44 pixels = 2816 features); 2) *1NN PCA*. The well-known Principal Component Analysis (PCA) technique is used to only keep those principal components accounting for a 99% of the variance; and 3) *1NN PCA+LDA*. As in [9,3], PCA is applied along with the Linear Discriminant Analysis (LDA) technique to obtain both the best data representation and the best class separability respectively. After LDA, the number of features become $n = c - 1$, with $c$ being the number of classes (people identities). Note that all of them learn from the gallery set unlike our approach which uses a non-target auxiliary training set. Other methods are also compared in individual experiments whenever possible although a direct comparison is always difficult because previous works were mostly evaluated under a cooperative setting where all sequences in the gallery have the same covariate conditions, so do those in the probe.

**Performance Measures.** The averaged gait recognition performances from the different trials are displayed on *Cumulative Match Score (CMS)* curves [14]. A CMS curve shows the percentage of probe sequences whose identity has been correctly recognised in the gallery among the top $x$ matches.

### 3.2   Experimental Results on USF

The *USF HumanID Gait Database (USF)* [16] is composed of videos of 122 subjects captured in an outdoor uncontrolled environment, which comprises up to five covariate conditions: 1) *surface*: subjects walk in two different surfaces, concrete (C) and grass (G); 2) *footwear*: two different shoe types are considered

**Table 1.** Description of experiments carried out on USF and CASIA gait databases under uncooperative settings. Legend: B-Carrying conditions (Briefcase or Bag), C-Clothing changes, S-Surface, V-View, T-Time.

| Experiment | Covariate conditions | Subsets | #People | #Sequences |
|---|---|---|---|---|
| USF_Exp1 | B | $\{$C A L **NB** $M + N_1$, C A L **BF** $M + N_1\}$ | 121 | 242 |
| USF_Exp2 | B S | $\{$**C** A L **NB** $M + N_1$, **G** A L **BF** $M + N_1\}$ | 117 | 234 |
| USF_Exp3 | B S V | $\{$**C** A **L** **NB** $M + N_1$, **G** A **R** **BF** $M + N_1\}$ | 117 | 234 |
| USF_Exp4 | B S V T | $\{$**C** A **L** **NB** $M + N_1$, **G** A **R** **BF** $N_2\}$ | 34 | 68 |
| CASIA_Exp1 | B | $\{NM_{90°}, BAG_{90°}\}$ | 124 | 496 |
| CASIA_Exp2 | C | $\{NM_{90°}, COAT_{90°}\}$ | 124 | 496 |
| CASIA_Exp3 | B C | $\{NM_{90°}, BAG_{90°}, COAT_{90°}\}$ | 124 | 744 |
| CASIA_Exp4 | V | $\{NM_{90°}, NM_{\theta°}\} \; \theta° = 18° \cdot X$ | 124 | 1488 |
| | | with $0 \leq X \leq 5 \in \mathbb{Z}^+$ | | |

(A) and (B); 3) *view angle*: subjects were captured by two cameras located in the left (L) and right (R) sides of the walking path yielding two view angles both close to side views, *i.e.* view change between L and R is small; 4) *carrying condition*: carrying a briefcase (BF) or not (NB); and 5) *time*: some sequences were recorded in May (M), while others in the next November (N) which implies clothing changes among others. A total of 32 possible subsets can be obtained based on the different combinations of these covariate conditions in the gallery and probe sets. Due to space limitation, here we only report results on four representative configurations, resulting in four experiments as shown in Table 1. Starting from the easiest one (USF_Exp1), which copes with only one covariate factor (carrying conditions), the experiments get harder, and the hardest one (USF_Exp4) deals with up to four covariate condition changes between the gallery and probe at the same time (carrying conditions, surface, view angle, and time). Note that these experiments are different from the common setup [16], under which all sequences in gallery or probe contain the same covariate conditions.

The results are shown in Figure 3. It can be observed that: 1) all the existing template matching (1NN and 1NN PCA) and learning based (1NN PCA+LDA) approaches yield very weak performances under these challenging settings. In addition, as expected, their performances become worse as the experiment gets harder. For instance, the best rank 1 matching rate (the correct classification rate) among the three drops from 15.8% in USF_Exp1 to 3.8% in USF_Exp4; 2) the learning based method (1NN PCA+LDA) suffers from the overfitting problem due to lack of training data. As a result, its performance is even worse than that of non-learning based ones; 3) our approach (PrRankSVM) significantly outperforms the compared ones. Its recognition rate also decreases as more covariate conditions are added but in a more graceful way (68.2% in USF_Exp1 and 14.4% in USF_Exp4). As a result, the performance gap between our method and the others gets bigger as the experiment becomes harder (4-folds increase in USF_Exp1 becomes 14-fold in USF_Exp4); and 4) even though our rank 1 result for USF_Exp4 is poor, the rank 10 result is almost 50%, which makes it of practical use for assisting a human operator in matching people by providing a small list of the likely candidates selected by the model.
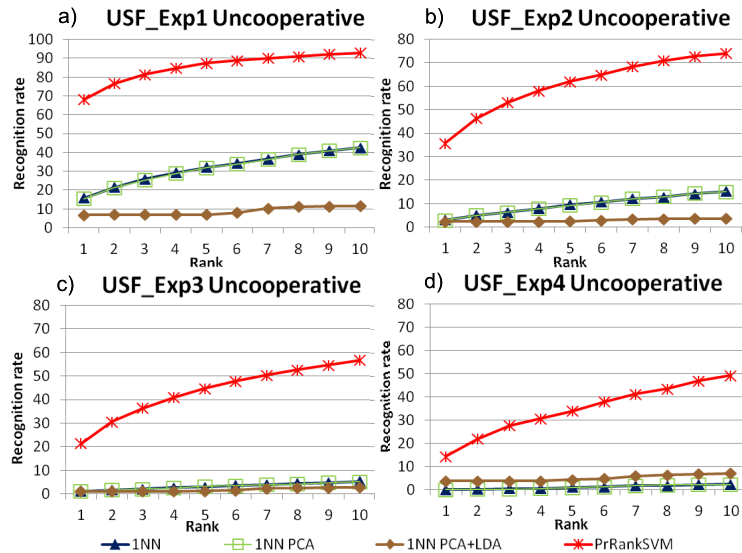
**Fig. 3.** CMS curves for the four experiments with USF

### 3.3 Experimental Results on CASIA

The *CASIA Gait Database* - Dataset B [5] contains 124 subjects captured under three different covariate condition changes: carrying, clothing, and view angle. Note that the view changes are much bigger in CASIA than in USF - each subject was captured from 11 different view angles from frontal view (0°) to back view (180°) including side-view (90°). For each view, each subject has 10 gait sequences: six normal (NM) where the person does not carry a bag or wears a coat, two carrying-bag (BAG) and two wearing-coat (COAT). All the videos were recorded indoor with a uniform background and controlled lighting.

**Carrying and Clothing Condition Changes.** Three experiments are first conducted to evaluate the different approaches under carrying and clothing condition changes. As shown in Table 1, CASIA_Exp1 focuses on carrying conditions alone, CASIA_Exp2 on clothing changes alone, and CASIA_Exp3 explores both covariate conditions together. For all the three experiments, only side view (90°) gait sequences are used because the effect of view is investigated in a separate experiment. From the 10 side-view sequences available for each subject, two normal sequences (NM) out of six are randomly selected along with the two in which the subject wears a coat (COAT), and the other two carrying a bag (BAG). It gives a total of six sequences per person, and 744 in total for CASIA_Exp3.

The results of these experiments are depicted in Figure 4. They show similar trends as those in the USF experiments although with higher recognition rates for all the approaches mainly because of the cleaner silhouettes as compared to the USF ones (see Figure 2).
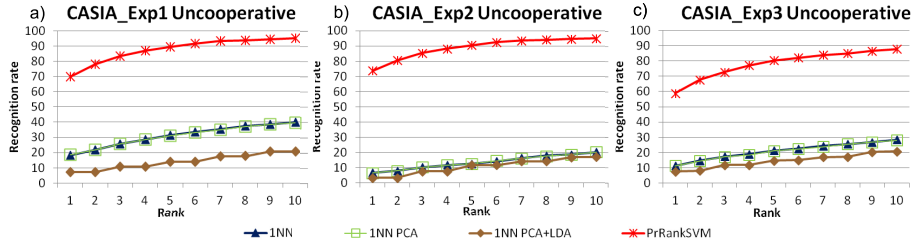
**Fig. 4.** CMS curves for the three first experiments with CASIA

More specifically, the main findings are: 1) consistent to the results reported in other works [3,21], clothing changes seem to affect gait more than the carrying conditions for the three compared baseline approaches, but they make little difference to our ranking-based method. Besides, two covariate condition changes co-existing (CASIA_Exp3) also represents a harder problem compared to only a single covariate alone (CASIA_Exp1 and CASIA_Exp2). Furthermore, the performance of the compared baseline approaches under the uncooperative setting is poor with less than 40% recognition rate even at rank 10 for CASIA_Exp3; 2) similarly to the results in USF, 1NN PCA+LDA suffers from overfitting and it is unable to deal with the large intra-class varariation; 3) again, our method (PrRankSVM) significantly outperforms the alternatives with at least a 3-fold increase in rank 1 recognition rate.

As mentioned before, due to the uncooperative setting we use, our results are not directly comparable with most results published in the literature, which were obtained under a cooperative setting. The only exception is [3], which uses a setting similar to our CASIA_Exp3 with a gallery set also containing a mix of NM, BAG and COAT sequences. Their result of 53% is comparable with our 58.9% in CASIA_Exp3. However, there is still a number of vital differences: 1) we use half of the 124 subjects for training whilst they use all for gallery and probe. Importantly their model is learned using the gallery set, thus using the same people as in the probe set; 2) they considered all the NM sequences instead of only two per person in the gallery set to make sure there were enough data in the gallery set to learn their model; and 3) they need to perform PCA+LDA for each pair of gallery and probe sequences, whilst our approach only learns the ranking model once and is able to very efficiently compute the matching score during testing by using Eq. 1. Overall, our method seems to be more generally applicable (*i.e.* it can deal with any other covariate condition changes such as view angle, and can work even with just a single sequence per subject in a gallery set) and suitable for real-time applications.

**Multi-view Gait Recognition.** The experiment CASIA_Exp4 is designed to evaluate our ranking approach under large view angle changes. It aims to match sequences of people seen in their side view (90°), which is considered the best angle for gait to be effective, with respect to sequences in some of the other

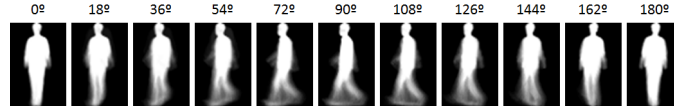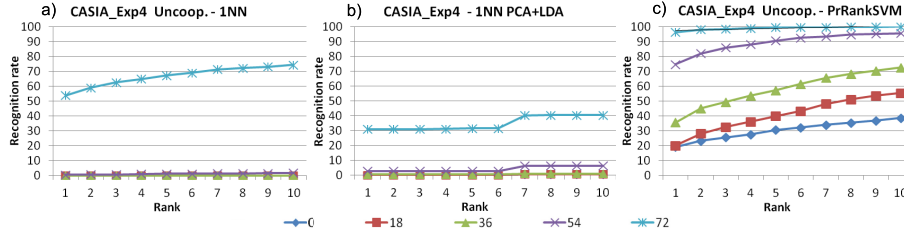**Fig. 5.** A subject from CASIA seen from different view angles



**Fig. 6.** CMS curves for the cross view experiment (CASIA_Exp4) in CASIA

view angles available in CASIA: $\theta = \{0°, 18°, 36°, 54°, 72°\}$. View angles greater than $90°$ are not chosen because they tend to achieve performances similar to those of their corresponding symmetrical angles [11,3], *i.e.*, $108°$ is similar to $72°$, $126°$ to $54°$ and so on. For each possible pair $(90°, \theta_i)$, an uncooperative setting is adopted as follows. Only the six NM sequences of each subject are considered, and all of them are assigned to either the training or test set. Thus, in the training set, each selected person is represented by six NM sequences from $90°$ and other six from the other view angle $\theta_i$. The test sequences are split into gallery and test following the procedure explained in Section 3.1, such that each test subject is represented in both the gallery and probe sets with only an unknown type of sequences in each, either $90°$ or $\theta_i$, and both view angles are equally represented in both sets (see Table 1 for detailed information).

Figure 6 shows a comparison of the results of two non-ranking methods (1NN and 1NN PCA+LDA) and our approach. Each plot depicts the CMS curves for all possible pairs $(90°, \theta_i)$. It is clear that, under an uncooperative setting, both the two non-ranking techniques fail miserably when the view angle difference is beyond $18°$. This is unsurprising because, as can be seen in Figure 5, the GEIs of a same subject under large view angle changes look completely different. In addition, given a probe GEI $q_t$ of a subject $s$ in a specific view angle $\theta_i$, the gallery contains samples of $s$ but from different view angles to $\theta_i$, while it also comprises plenty of other samples from other subjects in the same view angle $\theta_i$. Under this setting, the recognition rate will probably be worse than random guess because it is almost certain that the probe sample $q_t$ will be matched with a wrong subject with the same view angle $\theta_i$. In comparison, our ranking based approach gives much better results under this challenging setting as the view angle difference becomes larger, because it learns transferable information to cope with the view change. The performance of our method is slightly worse but comparable to those reported in [2,11]. Nevertheless, a direct comparison
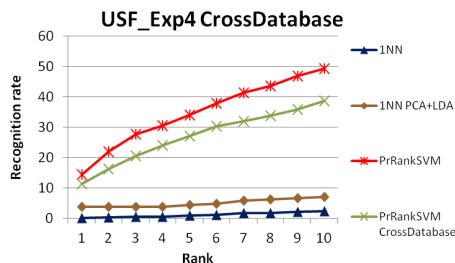
**Fig. 7.** Comparison of CMS curves for USF_Exp4 including the cross-database approach

cannot be made because: 1) their results were obtained under a cooperative setting, *i.e.* it was assumed that the view angle of each sequence can be either known a priori or estimated reliably; 2) the number of subjects used for training and test is different because we distribute 50% of them in each set whilst they use 24 subjects for training and 100 for test; and 3) their methods are specifically designed for cross-view gait recognition whilst our method can cope with any covariate condition changes either present alone or together.

### 3.4   Further Evaluation

**Cross-Database Gait Recognition.** So far our ranking model has been learned using different subjects from the same database. In this experiment, our model is learned using a particular database and applied to another one for recognition. More specifically, the gallery and probe sets are identical to those in USF_Exp4 (see Section 3.2), which has proved to be the most challenging experiment (see Figure 3d). The difference is that the model is now trained using the same training set as in CASIA_Exp2 (see Section 3.3), which contains changes in clothing. Figure 7 compares the result of this model (PrRankSVM Cross-Dataset) with those in Figure 3d. As expected, the performance of PrRankSVM CrossDatabase is weaker than that of PrRankSVM because the model is learned from a completely different dataset (indoor vs. outdoor) which features only one out of the four covariate condition changes of USF_Exp4. However, it continues outperforming the non-ranking based methods. We can thus conclude that useful information about what features are more reliable under different covariate conditions can be transferred between the two datasets using our model.

**Cooperative Setting.** All the experiments presented in the previous sections follow an uncooperative setting. In this section, we present some results under a cooperative setup similar to those used in previous works. As in the uncooperative experiments, we use 50% of people with all their sequences for training, and all the remaining ones for test. However, in this case, the type of sequences in gallery and probe are different and a priori known.
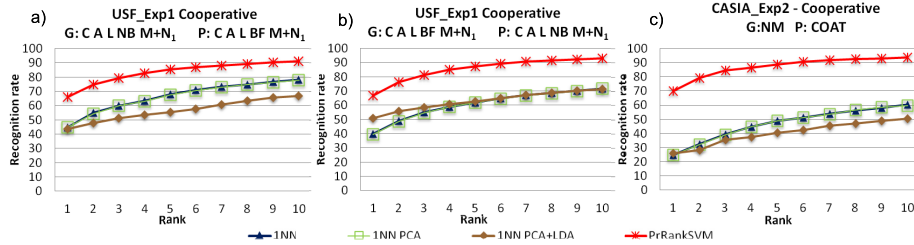
**Fig. 8.** CMS curves for cooperative experiments: a and b in USF dealing with briefcase covariate, and c in CASIA coping with clothing changes

Figures 8a and b show the results for the USF_Exp1 following a cooperative setting. This experiment involves two kinds of sequences (see Table 1): those in which people carry a briefcase (C A L BF MAY) and those in which they do not (C A L NB MAY). Thus, two different cooperative settings can be considered where both the gallery and probe sets must be composed of only a single type of sequences. The results in Figures 8a and b show that our rank 1 recognition rates almost double those of the non-ranking methods in both cases.

The results of an experiment in CASIA (CASIA_Exp2) using a cooperative setting are depicted in Figure 8c. Again, our approach gets about 3-fold improvement over the non-ranking approaches. Under a similar setting, a rank 1 result of 32.7% and 44% are reported by [3,21] respectively, although their experimental setting is still slightly different from ours with larger gallery and probe sets (our learning based method needs to use part of the data for training whilst they do not). Nevertheless, compared with our rank 1 result of 70%, this does give an indication that our model is superior even under cooperative settings.

## 4   Conclusions

In essence, the proposed gait recognition approach is novel in that it introduces transfer learning to gait recognition for the first time. The original multi-class classification or identification problem is reformulated into a bipartite ranking problem which learns transferable information independent of the identity of people; it thus becomes a simple verification problem (genuine or imposter). This provides a number of advantages including: 1) only a single model is needed to cope with all possible covariate conditions; and 2) the model can be learned from different class/subjects as well as from a different dataset making it more generally applicable with limited data per person in a gallery set. Extensive experiments have validated the effectiveness of our approach particularly under challenging uncooperative settings.

# References

1. Aghajanian, J., Warrell, J., Prince, S.J., Li, P., Rohn, J., Baum, B.: Patch-based within-object classification. In: ICCV, pp. 1125–1132 (2009)
2. Bashir, K., Xiang, T., Gong, S.: Cross view gait recognition using correlation strength. In: BMVC, pp. 1–11 (2010)
3. Bashir, K., Xiang, T., Gong, S.: Gait recognition without subject cooperation. Pattern Recognition Letters 31(13), 2052–2060 (2010)
4. Cao, T.S.L., Liu, Z., Huang: Cross-dataset action detection. In: CVPR, pp. 1998–2005 (2010)
5. CASIA: CASIA Gait Database (2005), http://www.sinobiometrics.com
6. Chapelle, O., Keerthi, S.: Efficient algorithms for ranking with SVMs. Information Retrieval 13, 201–215 (2010)
7. Damen, D., Hogg, D.C.: Detecting Carried Objects in Short Video Sequences. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part III. LNCS, vol. 5304, pp. 154–167. Springer, Heidelberg (2008)
8. Freund, Y., Iyer, R., Schapire, R.E., Singer, Y.: An efficient boosting algorithm for combining preferences. Journal of Machine Learning Research 4, 933–969 (2003)
9. Han, J., Bhanu, B.: Individual recognition using Gait Energy Image. PAMI 28(2), 316–322 (2006)
10. Joachims, T.: Optimizing search engines using clickthrough data. In: ACM SIGKDD, pp. 133–142 (2002)
11. Kusakunniran, W., Wu, Q., Zhang, J., Li, H.: Support Vector Regression for multi-view gait recognition based on local motion feature selection. In: CVPR, pp. 974–981 (2010)
12. Lee, H., Hong, S., Kim, E.: An efficient gait recognition with backpack removal. EURASIP J. Adv. Signal Process 2009, 46:1–46:7 (2009)
13. Makihara, Y., Sagawa, R., Mukaigawa, Y., Echigo, T., Yagi, Y.: Gait Recognition Using a View Transformation Model in the Frequency Domain. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006, Part III. LNCS, vol. 3953, pp. 151–163. Springer, Heidelberg (2006)
14. Phillips, P.J., Moon, H., Rizvi, S.A., Rauss, P.J.: The feret evaluation methodology for face-recognition algorithms. PAMI 22(10), 1090–1104 (2000)
15. Prosser, B., Zheng, W.-S., Gong, S., Xiang, T.: Person re-identification by support vector ranking. In: BMVC, pp. 1–11 (2010)
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. PAMI 27(2), 162–177 (2005)
17. Singh, S., Biswas, K.K.: Biometric Gait Recognition with Carrying and Clothing Variants. In: Chaudhury, S., Mitra, S., Murthy, C.A., Sastry, P.S., Pal, S.K. (eds.) PReMI 2009. LNCS, vol. 5909, pp. 446–451. Springer, Heidelberg (2009)
18. Tao, S.J.D., Li, X., Wu, X., Maybank: General tensor discriminant analysis and gabor features for gait recognition. PAMI 29(10), 1700–1715 (2007)
19. Wang, H., Nie, F., Huang, H., Ding, C.: Dyadic transfer learning for cross-domain image classification. In: ICCV, pp. 551–556 (2011)
20. Yang, X., Zhou, Y., Zhang, T., Shu, G., Yang, J.: Gait recognition based on dynamic region analysis. Signal Processing 88(9), 2350–2356 (2008)
21. Yu, S., Tan, D., Tan, T.: A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition. In: ICPR, vol. 4, pp. 441–444 (2006)