FEATURE SELECTION ON GAIT ENERGY IMAGE FOR HUMAN IDENTIFICATION

Khalid Bashir, Tao Xiang, Shaogang Gong

Queen Mary, University of London

{khalid, txiang, sgg}@dcs.qmul.ac.uk

ABSTRACT

In this paper we address the problem of selecting the most relevant features for human identification by gait. Although gait as a behavioral biometric is concerned with how people walk rather than how people look, most existing gait recognition approaches employ both shape and dynamics information for recognition. This is because shape, as a static appearance feature also contains useful information for identification. However, the inclusion of shape information in the gait features can also introduce variations that will hinder the recognition performance. To address this problem, we develop both supervised and unsupervised feature selection methods to extract the most relevant and informative features from Gait Energy Image (GEI) for human identification. Extensive experiments are carried out which indicate that our feature selection methods significantly improve the performance of gait recognition.

Index Terms— Gait Recognition, Gait Energy Image, Feature Selection

1. INTRODUCTION

Among various image-based biometrics, face, iris, fingerprint, and gait are the most widely studied ones. It is well known that iris and fingerprint yield better and more reliable recognition performance compared to face and gait. However, both of them require cooperative subjects which may not be possible in situations such as covert video surveillance. For identifying non-cooperative people, gait is attractive when the distance between the camera and the subject is far away.

Gait recognition techniques mainly fall into two broad categories namely model based [3, 4, 5] and model free approaches [6, 8, 9]. The first step of most gait recognition algorithms is the extraction of silhouettes. This is because gait, as a behavioral biometric, is different from physical biometrics such as face in that gait mainly captures the dynamic aspect of human activity instead of the static physical appearance of human. In other words, gait is concerned with how people walk rather than how people look. By extracting silhouettes, a large part of physical appearance features have been removed from the image representation of human. Nevertheless, silhouette still contains information about the shape and stance of human body. Although body shape and stance seem to be irrelevant to gait, which is determined by the transitions of body stances, recent studies suggest that adding shape to gait dynamics help to improve the gait recognition performance since body shape can also contain useful information for distinguishing different people [2]. On the other hand, the inclusion of shape information in the gait features can also introduce variations that will hinder the recognition performance, as demonstrated later in this paper. For instance, a person wearing or without wearing a coat could be significantly different in shape while the ways he/she walks would be very similar. Based on these observations, we can conclude that including shape in the gait representation is a double-edged sword and a balance need to be struck in terms how much shape information should be included.

To this end, we propose to perform statistical learning based feature selection to select the most relevant shape and dynamic features for gait recognition. Gait Energy Image (GEI) is adopted as the gait representation. GEI is a spatio-temporal gait representation constructed using silhouettes [9]. Conventional silhouette based gait representations treat gait as a sequence of templates. In contrast, GEI represents gait using a single image which contains information about both body shape and human walking dynamics. GEI is thus a compact representation which makes it an ideal starting point for feature selection since it is computational expensive if the number of features to select is high. In spite of its compactness, it has been demonstrated that GEI is less sensitive to noise and able to achieve highly competitive results compared to alternative representations [9].

In this paper, both supervised and unsupervised feature selection methods are developed. Specifically, for supervised gait feature selection, a cross validation based approach is developed to search for the best features that lead to the optimal recognition performance. A key component of the approach is a hierarchical greedy search strategy that integrates a priori knowledge about the GEI feature characteristics to overcome the problem associated with searching through a high dimensional gait feature space exhaustively. For unsupervised feature selection, we propose a simple yet effective measurement of the relevance of GEI features without assuming that a labeled dataset is available. The effectiveness of both feature selection methods is evaluated through extensive experiments using the CASIA Gait database [7], which is one of the largest public gait databases consisting of 124 subjects, captured from 11 views. The results indicate that the proposed feature

selection methods significantly improve the performance of gait recognition.

2. FEATURE SELECTION FOR GEI

2.1. Gait representation

Given a human walking sequence, a human silhouette is extracted from each frame using the method in [10]. After applying size normalization and horizontal alignment to each extracted silhouette image, gait cycles are segmented by estimating gait frequency using a maximum entropy estimation technique presented in [6]. Gait Energy Image (GEI) is then computed as

$$G(x, y) = \frac{1}{N} \sum_{t=1}^{N} I(x, y, t) \tag{1}$$

Where N is the number of frames in a complete gait cycle, x and y are the image coordinates, and t is the frame number in the gait cycle.
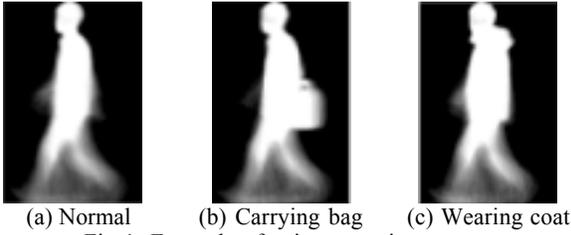
    (a) Normal   (b) Carrying bag   (c) Wearing coat
Fig 1: Example of gait energy images.

Examples of GEIs are shown in Fig. 1. Note that pixels with highest intensity values (white) in a GEI correspond to body parts that move little during a walking cycle (e.g. head, torso), while pixels with intensity values between the highest and lowest number (grey) correspond to body parts that move constantly (e.g. lower parts of legs and arms). The former mainly contain information about body shape and stance, whilst the latter tell us more about how people move during walking. We call the former *static areas* of a GEI and the latter *dynamic areas* of a GEI. There are also black areas in a GEI which contain no information about either body shape or gait dynamics. Intuitively, different areas of a GEI contain different types of information. The dynamic areas are invariant to the appearance of an individual; they seem to be the most informative part of the GEI representation for human identification. The static areas of a GEI also contain useful information for identification. However, as we mentioned earlier, since they mainly contain body shape information, they can introduce variations that hinder human identification through gait. For instance, in Fig. 1, three GEIs are computed from three sequences of the

input : GEI templates of size $h \times l$ divided into a training set and a validation set.
output: Features selected from GEI templates.

1   $CR_{max} = 0; h_s = \frac{h}{2}$;
2   **for** $i \leftarrow \frac{h}{2}$ **to** 1 **do**
3   use the bottom i rows as the selected features;
4   compute the classification rate CR_i;
5   **if** $CR_i > CR_{max}$ **then**
6         $h_s = i; CR_{max} = CR_i$;
7   **end**
8      $i = i - S_{coarse}$;
9   **end**
10   **for** $i \leftarrow h_s + \frac{S_{coarse}}{2}$ **to** $h_s - \frac{S_{coarse}}{2}$ **do**
11   use the bottom i rows as the selected feature;
12   compute the classification rate CR_i;
13   **if** $CR_i > CR_{max}$ **then**
14         $h_s = i; CR_{max} = CR_i$;
15   **end**
16      $i = i - S_{fine}$;
17   **end**
18   the bottom h_s rows are outputted as the selected features;

Algorithm 1: Supervised GEI feature selection. S_{coarse} and S_{fine} are the steps used for the coarse and fine search respectively.

same person walking under different conditions. The dynamic areas of the GEI suggest that they are the same person while the static areas suggest otherwise.

The above analysis suggests that parts of the information contained in a GEI are redundant and erroneous for human identification. It is well known that the inclusion of redundant and erroneous features in a pattern representation would hamper its recognition. To overcome this problem, automatic feature selection methods are developed.

2.2. Supervised Feature Selection

Supervised feature selection algorithms can be classified into two categories: filters and wrappers [1]. Filters weed out features without fitting model to the data. In contrast, wrappers select features that can be used by the subsequent learning algorithm to generate the best classification result. Wrapper algorithms are focused here as they usually lead to more accurate classification. One of the most widely adopted wrapper algorithms is cross validation. Dividing the data into a training and a validation set, the algorithm performs combinatorial search through the space of possible subsets of features so that the classifier learned using the selected features yield the highest classification rate on the validation set.

In our case, the original feature space has a dimensionality of $d = w \times h$ where w and h are the width and height of the GEI respectively in pixels. The total number of feature subsets to search through is 2^d. Therefore an exhaustive search is impossible and a greedy search strategy is needed to make the algorithm computational tractable. To this end, a hierarchical greedy search method is developed which incorporates our a priori knowledge about the characteristics of a GEI. Specifically, the search strategy drastically reduces the number of feature subsets to be evaluated by taking the following measures:

1. Each row of the GEI is treated as a feature unit.
2. The search area is focused on the bottom half of the GEI based on the observation that gait dynamics are mostly concentrated there.
3. We sequentially remove rows from the top of the search area based on the a priori knowledge that the lower a GEI row is located, the more dynamic information it contains.
4. A coarse scale search is conducted first, which is followed by a fine scale search centered at the optimal feature subset determined by the coarse scale search.

Our supervised feature selection algorithm is presented in Algorithm 1, where the classification method referred in lines 4 and 12 is described in Section 2.4.

2.3. Unsupervised Feature Selection

In spite of all the measures taken to speed up the search process in the proposed supervised feature selection algorithm, the algorithm is still computationally expensive especially if the size of the database is large. To overcome this problem an unsupervised feature selection algorithm is developed. The algorithm is based on the assumption that across all the normalized and aligned GEIs in a training set, the pixels whose intensity values exhibit larger variations are more likely to be useful for distinguishing different people. We therefore compute the standard deviations of the GEI intensity values at all pixel locations across GEI templates and use them as a score for the usefulness of each feature in a GEI. A threshold T is then used to select the features used for classification.

### 2.4. Gait Recognition using Selected Features
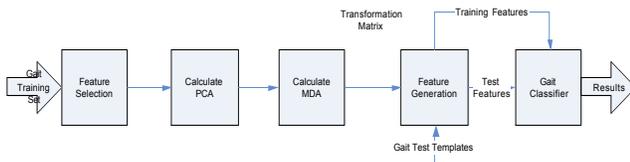
Fig. 2: The structure of the proposed gait recognition approach.

A byproduct of feature selection is the reduction of dimensionality of the feature space. However even after feature selection, the dimensionality of the feature space is still high and needs to be further reduced. This is done through combining Principal Component Analysis (PCA) and Multiple Discriminant Analysis (MDA) to the selected features (see [9]). In particular, after PCA, the dimensionality is reduced to $2c$ where c is the number of classes (human identities). The final dimensionality of the feature space after MDA is c. The system diagram of our gait recognition approach is shown in Fig. 2.

3. EXPERIMENTS

The CASIA Gait database [7] was used to evaluate the effectiveness of the proposed feature selection algorithms. It is an indoor gait database consisting of 124 subjects captured from 11 different views simultaneously starting

from 0° to 180° with an increment of 18°. The database has 10 walking sequences for each individual consisting of 6 normal walking sequences (Set A), 2 carrying-bag sequences (Set B) and 2 wearing-coat sequences (Set C). The total number of sequences in the database is 13640. We used the first 4 sequences of each individual in Set A as the training set (Set A1) and the rest as the test set including the rest sequences in Set A (Set A2), Set B and Set C. The original image size of the database is 320x240. After size nomalisation, the size of the GEIs became 128x88 (i.e. the original feature space has a dimensionality of 11264).
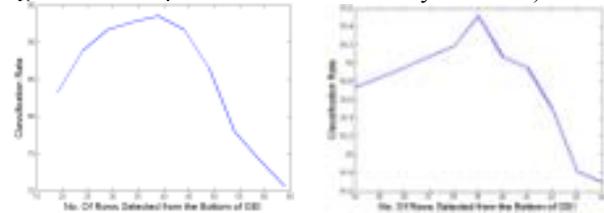

Fig 3: left: coarse search result for the optimal number of rows to keep; right: fine search result.


(a)   (b)
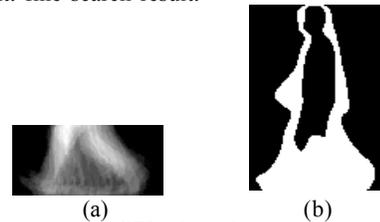
Fig 4: (a) Features in a GEI selected using supervised method. (b) The binary mask image highlighting the features selected by the unsupervised method.

We first conducted experiments using 90° view (side view) for both training and testing. Note that gait is most effective at side view. For supervised feature selection, S_{coarse} and S_{fine} (see Algorithm 1) were set to 5 and 1 respectively. Set A2 and half of Sets B and C were used as the validation set. Fig. 3 shows the results for coarse and fine searches in the feature space. 39 rows in the bottom of a GEI or 3432 features were selected after the fine scale search. For unsupervised learning, we set the threshold T to a value so that the number of feature selected by the unsupervised method is the same as that by the supervised method for fair comparison of the two methods. The features selected by the two methods are shown in Fig. 4. It can be seen from Fig. 4(b) that the unsupervised method selected mostly featured at the lower part of the legs (same as the supervised method) and the boundary of human body (not selected by the supervised method). The amount of times used for the two features selection methods were 1.25 hours and 5 seconds respectively, i.e. the later was 900 times faster than the former.

	Existing methods		Proposed methods	
	CAS	UCR	Supervised	Unsupervised
Set A2	97.6	99.4	98.6	**99.4**
Set B	32.7	60.2	**85.5**	79.9
Set C	52.0	22.0	**88.8**	31.3

Table 1: Comparison of the performance of the proposed methods with two existing methods without feature selection

(%). CAS–direct GEI shape match [7]; UCR–feature learning using PCA and MDA [9].

Table 1 compares the proposed methods with two existing methods. It can be seen that for normal walking sequences (Set A2), the performances are similar with or without feature selection. However for harder cases with people wearing a coat or carrying a bag, the results of our methods are significantly better than those without feature selection. Comparing the supervised and unsupervised feature selection methods, it is noted that the results of the unsupervised method were slightly better for normal-walking sequences (Set A2), slightly worse for carrying-bags sequences (Set B), and much worse for wearing-coat sequences (Set C) (but still better than the existing methods). The different performances of the unsupervised method on Sets B and C were due to the fact that most variations caused by wearing a bag were eliminated from the GEI via feature selection (e.g. applying the mask in Fig. 4(b) to Fig. 1(b)), whilst unsupervised feature selection helps much less in the case of wearing a coat (e.g. applying the mask in Fig. 4(b) to Fig. 1(c)).

Probe angle	Set A2		Set B		Set C	
	CAS	Sup	CAS	Sup	CAS	Sup
0°	0.4	0.8	0.4	0.6	1.2	0.4
18°	2.4	1.6	2.8	0.8	2.4	1.5
36°	4.8	4	5.2	2.9	4	5.2
54°	17.7	**18.2**	8.5	**9.7**	6	**15**
72°	82.3	**90.3**	42.3	**79.4**	20.6	**77.5**
90°	97.6	**98.6**	52	**85.5**	32.7	**88.7**
108°	82.3	78.5	31.9	**60.6**	16.5	**62.3**
126°	15.3	**16.2**	9.7	**11.3**	6	**12.5**
144°	5.2	3.6	6	2.6	3.6	6.2
162°	3.6	2.1	3	1.2	3.2	2
90°	1.2	0.8	2	0.4	0.8	1.2

Table 2: Comparison of recognition performance of the proposed method with an existing method taking into account the probe view angle change (%). Sup–the proposed method with supervised feature selection.

We also used the classifiers learned using 90° view sequences to recognise people captured at different view angles. This experiment aims to test the performance of different methods when the probe view angle is different from a fixed gallery angle (90° in this case). The results are presented in Table 2. Due to the limitation of space, we only include the results on the proposed method with supervised feature selection and the template matching without feature selection method in [7]. As expected, the performance of all algorithms deteriorated as the difference between the probe and gallery angles was getting bigger. However, our results indicate that our feature selection based recognition method outperforms those without feature selection when the probe angle is close to the gallery angle.

4. DISCUSSIONS AND CONCLUSION

Our experiments demonstrated that feature selection on a silhouette based gait representation greatly improves the recognition performance when variations are introduced in

human body shape and stance by carrying a bag or wearing a coat. It is clear from the experiments that the assumption we made on the relevance of each feature in the unsupervised feature selection method is valid although the method is less successful given body shape variations caused by wearing a coat compared to our supervised feature selection method. In spite of the inferior performance, an unsupervised feature selection is advantageous over a supervised one in terms of the computational cost. More sophisticated unsupervised feature selection methods such as mutual information or maximum information compression index based methods [1] will be investigated. Although our experiments suggest that feature selection makes gait recognition algorithms more robust to different probe view angle, techniques such as canonical view synthesis [2] need to be incorporated into our approach to better cope with view angle change.

5. REFERENCES
[1] M. Law, M Figueiredo, and A. Jain, "Simultaneous feature selection and clustering using mixture models," IEEE Tran. PAMI, 29(9), pp 1-13, 2004.
[2] Z. Liu and S. Sarkar, "Outdoor recognition at a distance by fusing gait and face," Image and Vision Computing, Vol. 25, issue 6, pp 817-832, June 2007.
[3] H. Lu, Plataniotis, and Venetsanopoulos, "A Layered Deformable Model for Gait Analysis," 7th Intl Conf. Automatic Face and Gesture Recognition, pp. 249-254, Apr. 2006.
[4] G. Zhao, G. Liu, H. Li, and Pietikainen,"3D Gait Recognition Using Multiple Cameras," 7th Intl Conf, Automatic Face and Gesture Recognition, pp. 529-534, Apr. 2006.
[5] R. Zhang, C. Vogler, and D. Metaxas, "Human Gait Recognition," Computer Vision and Pattern Recognition Workshop, pp 18, June 2004.
[6] S. Sarkar, P. Phillips, Z. Liu, I. Vega, P. Grother, K. Bowyer, "The HumanID Gait Challenge Problem: Data Sets, Performance, and Analysis," IEEE Tran. PAMI, 27(2), pp.162-177, 2005
[7] S. Yu, D. Tan, T. Tan, "A Framework for Evaluating the Effect of View Angle, Clothing and Carrying Condition on Gait Recognition," 18th International Conference on Pattern Recognition, pp. 441-444, 2006.
[8] S. Yu, L. Wang, W. Hu and T. Tan, "Gait Analysis for Human Identification in Frequency Domain," Proceedings on Image and Graphics 2004, pp 282-285, Dec 2004.
[9] J. Han and B. Bhanu, "Individual Recognition using Gait Energy Image," IEEE Trans. PAMI, 28(2), pp 316-322, 2006.
[10] L. Wang, T. Tan, H. Ning, and W. Hu. Silhouette analysis-based gait recognition for human identification. IEEE Tran. PAMI, 25(12):1505-1518, 2003.