A NOVEL STUDY AND ANALYSIS ON SEGMENTAL GAIT SEQUENCE RECOGNITION

Nini Liu, and Yap-Peng Tan

School of Electrical and Electronic Engineering
Nanyang Technological University, Singapore

ABSTRACT

This paper presents a novel study and analysis on two important problems in gait recognition: one is how to perform gait recognition with only segment of a complete gait cycle for test and the other is how static and dynamic information affect the recognition results in such situation. In conventional gait recognition research, the gait sequence of at least one cycle is usually needed to ensure the completeness of gait dynamics. And in this paper we will show that this is not a necessary condition and considerable recognition results could still be achieved with only part of the cycle based on appropriate similarity measure algorithms. Moreover, we will also show the different roles that static and dynamic information play in such situation through experimental study. To our knowledge, there is little work on these important problems, results from which can make gait recognition more applicable in frequently happened scenarios such as occlusion.

Index Terms— Gait recognition, segment of gait cycle, feature set, static and dynamic gait.

1. INTRODUCTION

Gait recognition has attracted more attention in recent years as it enables the surveillance system to work at a distance without the notation of the target which fails other biometrics such as face, iris and so on. However, compared to these biometrics which work on still images, gait recognition needs a sequence of frames containing at least one gait cycle to work. Till now, the research in gait still regard this requirement as a necessary and thus most of the proposed methods can only work with complete gait cycle. Although the more information we know about the gait, the higher recognition rate can be achieved, this doesn’t mean that we can’t recognize a person with only a segment of a complete gait cycle. In this paper, we will study how to deal with this kind of recognition between part of a cycle and complete cycles, and then further analyze the role of different kinds of information in such situation. In fact, this is an important study as it will make the gait recognition research more applicable in real life, such as in some indoor environment (banks, market) or when occlusion happened, only part of a complete gait cycle or part of a complete silhouette (temporal or spacial segment of gait) are usually captured for recognition, and in situation of bad quality videos, we could obtain only a small number of frames with well aligned silhouettes.

To perform gait recognition with segment of a complete cycle for test, a new problem comes out as the frame number of the test sequence (part of a cycle) is different from that of the training sequences (complete cycle). In fact, there exists in gait recognition similar problem from the beginning as different subjects have different number of frames in a complete gait cycle. And Dynamic Time Warping (DTW) [1] is proposed to solve this problem for similarity measure. However, it works only for small variance in frame number and most importantly, it needs the frames between sequences well aligned according to gait pose. Another popular solution is to shrink or stretch the sequence to have the same number of frames using the binary morphological transform technique described in [11]. The problem is not well solved until the popular feature representation Gait Energy Image (GEI) [2] was proposed in 2006. GEI represents a gait sequence by the mean of all frames and thus avoid the different frame number problem, besides, it achieves significant recognition results which makes it popular in gait recognition and most of the latest proposed methods are based on the GEI feature. But it still needs the gait sequence to be a complete cycle. Therefore, we need to look into this problem from another point of view and in this paper, we propose to treat each frame as a feature vector again and regard the gait sequence as a set of features, thus efficient set-to-set matching algorithms could be applied to calculate the similarity between different sequences.

In fact many computer vision tasks can be cast as learning problems over feature sets and thus the set-to-set matching problem has been studied a lot. Relevant previous approaches can be broadly divided into model-based methods and model-free methods. In model-based approaches, each feature set is modeled by a distribution function, typically Gaussian. In [3], the Kullback-Leibler Divergence (KLD) is used to measure the closeness of two distributions, while in [4], intrapersonal and interpersonal subspaces are built for training image sets and MAP Bayesian analysis is introduced to measure the similarity between different feature sets. However, it is difficult for parameter estimation with only a small number of feature samples, thus modelling errors may be significant. For model-free approaches, the Earth Mover’s Distance [9] is introduced.
to measure the similarity of two video sequences for face. Subsequently, the idea of subspace learning was reported to achieve the promising results in which each set is presented by a linear subspace and the set matching problem becomes the similarity measure among different subspaces. In [10], the inner product of basis vectors from different subspaces is used to calculate the distance. In [5], the smallest principle angle between two subspaces is defined as the similarity. More recently, researchers have paid more attention to canonical correlations for image set matching [7] which also show advantages over previous methods. And in our task, we apply this canonical correlations, which maps two feature sets into the same dimensional subspace with maximum correlations, to measure the similarity of two gait sequences with different number of frames. Experimental results show that considerable recognition results (over 80%) can also be obtained with only several (3-5) frames from a gait cycle especially by using the canonical correlation method.

Moreover, the static and dynamic information based gait recognition has been studied a lot recently. It shows that the dynamic information mainly from the movement of the limbs is very discriminative for recognition [8]. However, they are still based on the GEI feature which are representation of complete gait cycle. In this paper, we will also study the role of static shape information and dynamic information under the particular condition as before. We divide the whole silhouette into upper and lower body parts which are known to contain mainly static and dynamic information separately, and then perform the recognition to see their contribution. Detailed analysis and conclusions based on different experimental results are provided in the experimental part.

2. CANONICAL CORRELATIONS

Given a segment of walking video, preprocessing steps containing walking person detection and tracking, silhouette extraction and silhouette alignment and normalization are first applied to obtain binary silhouettes for each frame in the sequence. The methodologies we used are described in [12].

In real application, it usually happens that only part of a complete gait cycle is available for recognition, and this brings a problem as the frame number is different from the training sequences which are complete gait cycles. To solve this problem, we propose to treat each gait sequence as a feature set and adopt the efficient canonical correlations to measure the similarity between different sets.

2.1. Subspace extraction

After applying the preprocessing steps as mentioned above, we regard each frame as a feature vector and thus each gait sequence can be represented as a two dimensional matrix (∈ R^d+p), where d is the feature dimension and p is the number of frames in the sequence). Then we obtain the corresponding subspace for each of the gait sequences by SVD:

\[ A = UDV^T \] (1)

where A is the original representation of the gait sequence in R^d+p. U (V) is the left (right) singular vector matrix of A in R^d*d (R^p*p) and D is the singular value matrix in R^d+p. Obviously, the subspace defined by U = (u_1, u_2, ..., u_p) is the corresponding feature subspace (eigenvector matrix of A * A^T) and we extract part of the basis vectors from U to represent the subspace because it is not full ranked (p ≪ d). As singular values are ranged decrease in D, we choose the first m (= p - 1) vectors U' = (u_1, u_2, ..., u_m) from U to represent the corresponding subspace for A.

2.2. Subspace similarity

After subspace extraction for each gait sequence, the similarity measure between gait sequences becomes the similarity between different subspaces. Usually, the principle angles between corresponding basis vectors reflect the similarity of two subspaces. Here, we also apply SVD to calculate the solutions for CCA [7].

Given two basis vector matrices U'' = (u''_1, u''_2, ..., u''_n) and U''' = (u'''_1, u'''_2, ..., u'''_n) obtained from training and testing sequences separately through equation (1), the principle angles between basis vectors of U'' and U''' can be calculated first by SVD as

\[ (U''^T U''') = U_1 D_1 V_1^T \] (2)

where U_1 and V_1 are basis vectors that can project U'' and U''' onto the same subspace in which the corresponding projections achieve the maximum correlation. The singular values in D_1 = diag(λ_1, λ_2, ..., λ_{min{m,n}}) is just the cosines of principle angles between basis vectors of U'' and U'''. Then the similarity between U'' and U''' is defined as

\[ S(U'', U''') = \lambda_{max} \] (3)

where λ_{max} is the largest singular values in D_1. In fact, different similarity definitions have been proposed with different number and form of principle angles, and we choose this simplest one as it performs the best in our case. Because the subspace extraction can predict the "missing" information in original sequence, the use of more singular values for similarity measure would instead bring negative affect due to wrong prediction. At last, we use nearest neighbor classifier for identity assignment.

3. EXPERIMENTS & ANALYSIS

In this section, we will provide extensive experiments as well as intensive analysis based on the experimental results to conclude the important study.
3.1. Database

We choose CASIA gait database (Dataset B) [6] in our experiments. This database is a large database contain gait sequences from 124 subjects, among whom 93 were men and 31 were women. It is also a multi-view and multi-variants database, and we only use the 6 normal walking sequences captured in profile view (90°) from each subject for experiments, among which 4 are treated as training set and the left 2 are treated as testing set. And for each training sequence, we extract a complete cycle to learn the corresponding subspace. As for the testing sequences, we extract different part from the complete cycle for recognition according to different experimental conditions.

3.2. Gait cycle detection

We use a simple and common approach [12] to detect the gait cycle for all sequences. It regards the number of foreground pixels from the bottom half of each silhouette as a periodical vector and uses the autocorrelation of the number vector to calculate the period.

![Width vector and autocorrelation](image)

Suppose we obtain the number vector \( W = (w_1, w_2, ..., w_s) \) from a gait sequence, we calculate its autocorrelation \( C = (c_1, c_2, ..., c_r) \) as follows:

\[
c_r = \frac{1}{s} \sum_{k=1}^{s-r} (w_k - \mu)(w_{r+k} - \mu)
\]

where \( \mu \) is the mean of \( W \), and then assign twice the width between two nearest maximum values as the gait cycle for the sequence. Noise in the binary silhouette can easily affect the number vector, therefore, we choose the autocorrelation of the vector because the curve of autocorrelation is more smooth than the original one, see Fig.1.

3.3. Experimental results

![Frames in a complete gait cycle](image)

We show in Fig.2 the frames in a complete gait cycle. Obviously, different gait poses (frames) contain different discriminative information. Therefore, the recognition rates using different part of a cycle should also be different, see Fig.3. Because of this, we conduct each experiment 15 times with different starting frames and regard the mean as the final result.

![Recognition rates](image)

Table 1 shows the performance of different methods we mentioned before for this particular problem. Clearly, the conventional methods (GEI and DTW) fail when the gait sequence is less than half a cycle. However, the methods for set-to-set matching can still work, the recognition rates are over 80% even with \( \frac{1}{4} \) gait cycle (3-5 frames). Nevertheless, CCA performs the best and much faster than the other two. This demonstrates that the subspace representation is more robust to missing data as it can 'fill-in' the missing information.

It is very promising that the recognition rate can reach above 80% with only a few number of frames. As gait dynamic would reduce greatly in such situation, the results prove that the static shape information contained in the gait pose is also very discriminative for recognition. Besides, as gait is a symmetry movement especially in the profile view (see in Fig.2), the recognition results from conventional methods is also promising with at least \( \frac{1}{2} \) gait cycle.

<table>
<thead>
<tr>
<th>Method</th>
<th>1 cycle</th>
<th>( \frac{1}{4} ) cycle</th>
<th>( \frac{1}{2} ) cycle</th>
<th>( \frac{3}{4} ) cycle</th>
<th>( \frac{1}{2} ) cycle</th>
</tr>
</thead>
<tbody>
<tr>
<td>GEI</td>
<td>97.18</td>
<td>96.47</td>
<td>94.21</td>
<td>76.26</td>
<td>50.62</td>
</tr>
<tr>
<td>DTW</td>
<td>89.52</td>
<td>85.19</td>
<td>82.17</td>
<td>69.86</td>
<td>63.6</td>
</tr>
<tr>
<td>KLD</td>
<td>91.13</td>
<td>89.63</td>
<td>88.84</td>
<td>85.9</td>
<td>81.21</td>
</tr>
<tr>
<td>EMD</td>
<td>96.37</td>
<td>96.21</td>
<td>93.85</td>
<td>89.3</td>
<td>82.09</td>
</tr>
<tr>
<td>CCA</td>
<td>95.2</td>
<td>94.97</td>
<td>94.39</td>
<td>91.21</td>
<td>85.7</td>
</tr>
</tbody>
</table>

It is widely accepted in gait recognition that the lower part of the body contains most of the dynamic information during walking, while the upper body part mainly contain the static body shape information. To further study the role of this two kind of information, we equally divide the silhouette into upper and lower parts and conduct the experiments as before and the results are shown in Fig.4. From the figure we can see that when the frame number is greater than half period, the information from the lower dynamic part plays dominant role, and when the frames are less than half a cycle, the upper static part is more important for recognition. And the combination
of both always performs better. Moreover, another conclusion from the figure is that the recognition rate using upper body part changes a little with increase of frame number while that using lower body part increases significantly until the frame number reaches half a period.

4. CONCLUSIONS

This paper provides a novel study on two important problems in gait recognition. The first one is how to conduct recognition in condition of only segment of a complete gait cycle is available. We introduce the algorithms in image set matching to solve it and promising recognition results (over 80%) are achieved with only a few number (3-5) of frames available. Secondly, we study the importance of static and dynamic information under such particular condition based on experimental results, in which we divide the whole body into upper and lower body parts to perform recognition separately. Experimental results show their importance changes with the increase of information known about the gait cycle. This is an important study because it makes the gait recognition more applicable in some indoor environment or when occlusion happens, in which only a few frames or part of the body is available for recognition.

5. REFERENCES