Investigation of New Gait Representations for Improving Gait Recognition

Chirawat Wattanapanich, Hong Wei

Abstract—This study presents new gait representations for improving gait recognition accuracy on cross gait appearances, such as normal walking, wearing a coat and carrying a bag. Based on the Gait Energy Image (GEI), two ideas are implemented to generate new gait representations. One is to append lower knee regions to the original GEI, and the other is to apply convolutional operations to the GEI and its variants. A set of new gait representations are created and used for training multi-class Support Vector Machines (SVMs). Tests are conducted on the CASIA dataset B. Various combinations of the gait representations with different convolutional kernel size and different numbers of kernels used in the convolutional processes are examined. Both the entire images as features and reduced dimensional features by Principal Component Analysis (PCA) are tested in gait recognition. Interestingly, both new techniques, appending the lower knee regions to the original GEI and convolutional GEI, can significantly contribute to the performance improvement in the gait recognition. The experimental results have shown that the average recognition rate can be improved from 75.65% to 87.50%.

Keywords—Convolutional image, lower knee, gait.

I. INTRODUCTION

Gait recognition, which is non-intrusive in identifying individuals in distance, is still challenging in biometric research. The main advantage of gait recognition is that it allows low-resolution images to be used for long distance detection, and has non-interference to target activities. Moreover, gait information which is the personal walking characteristic is hardly to be forged.

There are two main stages in conventional gait recognition, gait feature extraction and classification. Gait features which represent the walking characteristic can be extracted from both gait model and gait image sequence. In a model free approach, gait features are usually extracted from gait representation called compact image which is generated from a complete gait cycle. The basic compact image, called GEI [1] or Average Silhouette [2], can be generated by averaging all binary silhouette gait images in a full cycle in a same view angle. GEI has been commonly used in the model free research because of its simplicity and low-time consuming. Nonetheless, other gait compact images have consequently been implemented to fulfill recognition efficiency such as Gait Entropy Image (GENI) [3], Gait Gaussian Image (GGI) [4], Flow Histogram Energy Image (FHEI) [5], Gradient Histogram Gauussian Image (GHGI) [6] and Gait Information Image (GII) [7]. This study has chosen GEI as the original gait representation. New gait representations are developed from the original in this study and presented in Section II.

Based on the compact images, there are various feature extraction processes available in gait recognition research such as PCA [8], [9], Linear Discriminant Analysis (LDA) [10], [11] and Convolutional Neural Network (CNN) [12]-[14]. In this study, a new type of compact image, called convolutional compact image, is developed, combined with PCA in gait feature extraction.

The second stage is classification. Existing classifiers in gait recognition include Nearest Neighbor (NN) [4], [15], SVM [16], [17], CNN [18], etc. This study has chosen one-against-all multi-class SVM as the classifier.

In summary, GEI and its variants are used as the gait representation, from which gait features are extracted by convolutional processes and the classification performance has been tested with multi-class SVM. CASIA dataset B [19] which has three appearances and 11 camera view angles is used in training and testing. The rest of the paper is organized as follows. Section II presents the methodology for the gait recognition system. Section III discusses experiments and results. The conclusion is given in Section IV.

II. METHODOLOGY

A gait recognition system, as shown in Fig. 1, usually has two phases: training and testing. The training phase is for model creation. In some cases, a gait representation can be directly used as the input for training a classifier. The testing or prediction phase compares the testing sample with all existing models to make a decision with the highest score in similarity.

A. Gait Representations

As indicated in Section I, there exist many gait representations. Nonetheless, one original and one newly developed representations are employed in this study. An example of the two gait representations is shown in Fig. 2.

1) GEI

GEI [1] is generated by averaging all binary images in walking sequence with the same view angle, as expressed in (1):

$$G(x, y) = \frac{1}{N} \sum_{t=1}^{N} B_t(x, y)$$  \hspace{1cm} (1)

where $N$ is the number of silhouette frames in a complete gait sequence, $t$ is the frame number in the gait sequence, $B_t(x, y)$
is the binary image at frame $t$, and $(x, y)$ is the pixel coordinate in a frame.

2) Convolutional Gait Energy Image (CGEI)

The new gait representation which is the average image of convolved images is generated from the original gait representation by applying convolution with normalization techniques to GEI. This gait representation is called CGEI. An example of the representation can be seen in Fig. 2 (b).

![Fig. 1 Gait recognition system overview](image)

Fig. 1 Gait recognition system overview

![Fig. 2 Gait representations](image)

Fig. 2 Gait representations

The original gait representation $x$ has been convolved with $M$ multi-dimensional filter. Output feature map is formally given by

$$y_{i'j'd'} = \sum_{i=1}^{H} \sum_{j=1}^{W} \sum_{d'=1}^{D} f_{i'j'd'} * x_{i'j'j'+1,i'+1,j'+1,d'}$$  \tag{2}

where $x \in \mathbb{R}^{H \times W \times D}$, $f \in \mathbb{R}^{H' \times W' \times D'}$, $y \in \mathbb{R}^{H' \times W' \times D'}$, $H$ is the height, $W$ is the width and $D$ is the depth of $x$. In this experiment, $D$ is equal to 1 because $x$ is grey scale image.

The number of output images or feature maps has the same number of kernels or $D'$; however, each output may have very different value because of random filters applied. Each channel of the feature map $x$ has been normalized by:

$$y_{i'j'd'} = \frac{x_{i'j'd'} - \mu_k}{\sqrt{\sigma_k^2 + \epsilon_k}}$$

where $x, y \in \mathbb{R}^{H' \times W' \times K \times T}$ and $T$ is the number of images or feature maps. Finally, the new gait representation is generated from the average of normalizing the feature maps. The size of CGEI, which is smaller than the size of GEI, can easily be calculated by

$$S_{CGI} = S_O - S_K + 1$$ \tag{4}

where $S_{CGI}$ is the CGEI width/height, $S_O$ is the original gait representation image width/height, and $S_K$ is the kernel width/height.

B. Multiple Lower Knee Gait Image (MLKGI)

The main purpose of this study is to investigate the effect of cross appearances, such as wearing a coat and carrying a bag, on gait classification performance. One of our assumptions is that some parts of a human body may be more significant than other parts in contributing the variance, hence improving gait recognition rate. The lower knee region in the gait representation has been selected in this study because it is the body part with the largest movement in complete gait cycle.
Fig. 3 Average anthropometric measurement

The knee height has been calculated from the overall height based on average measurement of anthropometric as shown in Fig. 3 [20]. A lower knee image has been selected as all pixels below the knee (0.285H*W) from the whole body. Appending the low knee image to the original image produces a new representation, as shown in Fig. 4 (a).

\[ LW = H - 0.285 \]  \hspace{1cm} (5)

A multiple lower knee gait image (MLKGI) could be generated by appending more than one lower knee images to the original image. An example of this gait representation is shown in Fig. 4 (b).

(a) one duplicate lower knee image appended  
(b) two duplicate lower knee images appended

Fig. 4 Example of multiple lower knee image

III. EXPERIMENTS

Three main experiments were conducted in this study including convolutional gait image (A), multiple lower knee gait image (B) and feature reduction (C). Experiments A and B used the same gait classification system as presented in Fig. 1. Nevertheless, the gait feature extraction was bypassed because the entire gait representation image was used as the input for SVM training and prediction. Experiment 3.3 used the optimal feature map calculated from the gait representation image by PCA to generate reduced dimensional gait features for SVM classification. This gait classification system is demonstrated in Fig. 5.

This study focused on cross appearance classification. All experiments were conducted on the CASIA gait dataset B which has three appearances and eleven view angles from 0° - 180°. We chose 116 objects, which have the complete sequence silhouette images in each appearance and view angle from 124 objects in this dataset. Each object has 10 video sequences, six for normal walk, two for wearing a coat, and
two for carrying a bag. Four normal walk, images were used as gallery in training and the remaining, two from each appearance as probe for SVM prediction.

### Table I

**Classification Rate with Different Kernel Size**

<table>
<thead>
<tr>
<th>View Angle</th>
<th>GEI</th>
<th>CGEI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3x3</td>
<td>5x5</td>
</tr>
<tr>
<td>0</td>
<td>74.80%</td>
<td>61.90%</td>
</tr>
<tr>
<td>18</td>
<td>77.27%</td>
<td>81.26%</td>
</tr>
<tr>
<td>36</td>
<td>77.27%</td>
<td>83.79%</td>
</tr>
<tr>
<td>54</td>
<td>76.52%</td>
<td>85.86%</td>
</tr>
<tr>
<td>72</td>
<td>76.91%</td>
<td>84.99%</td>
</tr>
<tr>
<td>90</td>
<td>75.49%</td>
<td>89.91%</td>
</tr>
<tr>
<td>108</td>
<td>75.03%</td>
<td>86.68%</td>
</tr>
<tr>
<td>126</td>
<td>73.74%</td>
<td>84.14%</td>
</tr>
<tr>
<td>144</td>
<td>74.83%</td>
<td>82.33%</td>
</tr>
<tr>
<td>162</td>
<td>76.15%</td>
<td>79.66%</td>
</tr>
<tr>
<td>180</td>
<td>75.03%</td>
<td>76.58%</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td><strong>75.65%</strong></td>
<td><strong>83.35%</strong></td>
</tr>
</tbody>
</table>

### Table II

**Classification Rate with Different Number of Kernels**

<table>
<thead>
<tr>
<th>View Angle</th>
<th>Number of kernels 8 16 32</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>77.01%</td>
</tr>
<tr>
<td>18</td>
<td>81.64%</td>
</tr>
<tr>
<td>36</td>
<td>81.61%</td>
</tr>
<tr>
<td>54</td>
<td>82.30%</td>
</tr>
<tr>
<td>72</td>
<td>84.05%</td>
</tr>
<tr>
<td>90</td>
<td>84.63%</td>
</tr>
<tr>
<td>108</td>
<td>84.08%</td>
</tr>
<tr>
<td>126</td>
<td>81.21%</td>
</tr>
<tr>
<td>144</td>
<td>80.17%</td>
</tr>
<tr>
<td>162</td>
<td>78.97%</td>
</tr>
<tr>
<td>180</td>
<td>76.47%</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td><strong>81.10%</strong></td>
</tr>
</tbody>
</table>

### A. CGEI Testing

This experiment examines the effect of number of kernels and kernel size used to generate CGEI on gait classification performance. The entire gait representation image is used as gait features. However, the size of CGEI depends on the kernel size since there was no padding in the convolutional process. For example, GEI has size of 120x120 pixels or 14400 pixels. The average correct classification rate (CCR) of cross appearance is shown in Tables I and II for the fixed convolution number of kernels and the fixed kernel size, respectively.

Table I shows that the number of kernels affects the classification rate. CGEI generated from the average of 32 convolved images had the highest CCR. From Table II, the CGEI generated from the kernel size of 3x3 had better CCR than the other kernel size. All CCR results from CGEI are significantly better than those from GEI. Especially, CGEI has a better result than CGEI under each view angle.

### B. Multiple Lower Knee Gait Image Testing

This experiment investigates the effect of number of duplicated lower knee regions appending to the original image, from one to five on the gait recognition performance. Both of GEI and CGEI are used as the input to generate MLKGI. The average CCR of each MLKGI has been shown in Tables III (GEI) and IV (CGEI).

Table III indicates that MLKGI can continuously improve the CCR of GEI with increasing the number of duplicated lower knee regions appending to the original image, whilst Table III shows that the result of MLKGI matches the result of CGEI only when the number of duplicated lower knee regions is greater than or equal to 3.

### Table III

**Classification Rate with Different Number of Lower Knee Region**

<table>
<thead>
<tr>
<th>View Angle</th>
<th>Number of duplicated lower knee regions 1 2 3 4 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>77.61%</td>
</tr>
<tr>
<td>18</td>
<td>78.10%</td>
</tr>
<tr>
<td>36</td>
<td>78.22%</td>
</tr>
<tr>
<td>54</td>
<td>78.48%</td>
</tr>
<tr>
<td>72</td>
<td>78.39%</td>
</tr>
<tr>
<td>90</td>
<td>78.59%</td>
</tr>
<tr>
<td>108</td>
<td>77.41%</td>
</tr>
<tr>
<td>126</td>
<td>76.87%</td>
</tr>
<tr>
<td>144</td>
<td>77.21%</td>
</tr>
<tr>
<td>162</td>
<td>78.10%</td>
</tr>
<tr>
<td>180</td>
<td>77.36%</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td><strong>77.85%</strong></td>
</tr>
</tbody>
</table>

### Table IV

**Classification Rate with Different Number of Lower Knee Region in Case of CGEI**

<table>
<thead>
<tr>
<th>View Angle</th>
<th>Number of duplicate lower knee regions 1 2 3 4 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>80.00%</td>
</tr>
<tr>
<td>18</td>
<td>82.70%</td>
</tr>
<tr>
<td>36</td>
<td>84.02%</td>
</tr>
<tr>
<td>54</td>
<td>84.57%</td>
</tr>
<tr>
<td>72</td>
<td>85.72%</td>
</tr>
<tr>
<td>90</td>
<td>86.52%</td>
</tr>
<tr>
<td>108</td>
<td>85.75%</td>
</tr>
<tr>
<td>126</td>
<td>82.50%</td>
</tr>
<tr>
<td>144</td>
<td>81.78%</td>
</tr>
<tr>
<td>162</td>
<td>81.38%</td>
</tr>
<tr>
<td>180</td>
<td>80.83%</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td><strong>83.25%</strong></td>
</tr>
</tbody>
</table>

### C. Feature reduction with PCA

Experiments A and B used whole GEI and CGEI images as gait features in the classification process. Since the large number of features is used, it is time consuming in the SVM training phase. In this experiment, PCA was added to the gait recognition system to reduce the dimensionality of the feature space. 464 selected components are applied in the optimal feature map as gait representation before SVM training and prediction processes. CGEI generated with the kernel size of 3x3...
3x3 and 32 kernels from the original MLKGI was chosen for this test because this kernel setting gave the best result from previous experiments. The CCR of MLKGI-CGEI with PCA is shown in Table V.

From Table V, almost MLKGI with PCA had slightly higher CCR than CGEI. The five duplicate lower knees had the best CCR which is 87.50% in these experiments. The detail of each appearance CCR which has compared all best CCRs from all experiments has been show in Fig. 6.

### TABLE V

**The CCR of MLKGI with PCA in Case of CGEI**

<table>
<thead>
<tr>
<th>View Angle</th>
<th>Number of duplicate lower knee</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>80.34%</td>
<td>80.09%</td>
<td>81.47%</td>
<td>82.16%</td>
<td>85.29%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>84.05%</td>
<td>82.82%</td>
<td>83.76%</td>
<td>84.20%</td>
<td>86.44%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>85.49%</td>
<td>83.97%</td>
<td>85.37%</td>
<td>85.75%</td>
<td>87.13%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>54</td>
<td>87.30%</td>
<td>86.81%</td>
<td>87.33%</td>
<td>87.99%</td>
<td>89.28%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>72</td>
<td>89.28%</td>
<td>89.05%</td>
<td>88.76%</td>
<td>89.14%</td>
<td>90.90%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>90</td>
<td>89.83%</td>
<td>89.60%</td>
<td>88.42%</td>
<td>88.56%</td>
<td>90.66%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>108</td>
<td>89.17%</td>
<td>89.05%</td>
<td>88.05%</td>
<td>88.25%</td>
<td>89.57%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>126</td>
<td>85.40%</td>
<td>85.60%</td>
<td>85.43%</td>
<td>85.75%</td>
<td>87.33%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>144</td>
<td>83.48%</td>
<td>83.62%</td>
<td>84.02%</td>
<td>84.40%</td>
<td>85.80%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>162</td>
<td>82.07%</td>
<td>81.47%</td>
<td>82.33%</td>
<td>82.93%</td>
<td>85.63%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>180</td>
<td>80.37%</td>
<td>80.55%</td>
<td>82.16%</td>
<td>82.96%</td>
<td>85.29%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average</td>
<td>85.16%</td>
<td>84.78%</td>
<td>85.19%</td>
<td>85.62%</td>
<td>87.50%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The comparison CCR of various methods which were recently published has been shown in Tables VI and VII.

### TABLE VI

**Average CCR Comparison for Recently Research on CASIA Dataset B Using All View Angles**

<table>
<thead>
<tr>
<th>Method</th>
<th>NM</th>
<th>BG</th>
<th>CL</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>[21]</td>
<td>97.39</td>
<td>75.08</td>
<td>86.28</td>
<td>86.25</td>
</tr>
<tr>
<td>[22]</td>
<td>98.00</td>
<td>90.00</td>
<td>64.00</td>
<td>84.00</td>
</tr>
<tr>
<td>[23]</td>
<td>99.00</td>
<td>79.00</td>
<td>60.00</td>
<td>79.33</td>
</tr>
<tr>
<td>[24]</td>
<td>94.10</td>
<td>84.20</td>
<td>87.60</td>
<td>88.60</td>
</tr>
<tr>
<td>Our proposed method</td>
<td>98.96</td>
<td>83.35</td>
<td>80.19</td>
<td>87.50</td>
</tr>
</tbody>
</table>

### IV. CONCLUSION

In this paper, novel gait representations have been proposed and developed in gait recognition tests on the CASIA dataset B. Based on the original GEI, a set of gait representations were generated by applying convolutional operation to the GEI and its variants, which were obtained by appending the lower knee regions to the original GEI. The results from testing experiments of cross appearance gait recognition have shown that the newly developed gait representations can better represent individual gait characteristic, and the recognition rate has dramatically been improved from 75.65% (GEI) to 87.50% (CGEI generated by appending five lower knee regions to the GEI and 32 convolutional kernels with 3*3 in size applied to the appended GEI). Interestingly, it has been approved on the CASIA dataset B that adding information of lower knee in GEI is a positive approach, and convolutional GEIs are a superior gait representation to GEI in cross appearance gait recognition.

### REFERENCES


