# Gait Recognition Method for Arbitrary Straight Walking Paths Using Appearance Conversion Machine

Xiaohui Zhao[1,2], Yicheng Jiang[*,1], Tania Stathaki[*,2], and Huisheng Zhang[3]

1 Research Institute of Electronic Engineering Technology, Harbin Institute of Technology, Harbin 150001, China

2 Department of Electrical and Electronic Engineering, Imperial College, London SW7 2AZ, U.K.

3 Department of Mathematics, Dalian Maritime University, Dalian 116026, China

**Abstract**

We investigate the problem of multi-view human gait recognition along any straight walking paths. It is observed that the gait appearance changes as the view changes while certain amount of correlated information exists among different views. Taking advantage of that type of correlation, a multi-view gait recognition method is proposed in this paper. First, we estimate the viewing angle of the monitor equipment in terms of the probe subject. To this end, our method considers this as a classification problem, where the classification signals are the viewing angles, and the classification features are the elements of the transformation matrix that is estimated by the Transformation Invariant Low-Rank Texture (TILT) algorithm. Then, the gallery gait appearances are converted to the view of the probe subject using the proposed Appearance Conversion Machine (ACM), where the gait features of the spatially neighbouring pixels of the gait feature are considered as the correlated information of the two views. In the end, a similarity measurement is applied on the converted gait appearance and the testing gait appearance. Experiments on the CASIA-B multi-view gait database show that the proposed gait recognition method outperforms the state-of-the-art under most views.

**Index Terms**

multi-view gait recognition, human identification, appearance conversion machine, extreme learning machine, viewing angle estimation, transformation invariant low-rank texture.

## I. Introduction

Gait has been widely accepted as a biometric feature for human recognition. One of the proposed techniques for gait recognition is the Gait Energy Image (GEI) [21]. This technique has inspired many researchers to conduct human recognition based on the GEI, owing to its high discriminative power, simplicity, and time-efficient calculation [2, 7, 12]. In [25], the performance of the GEI is measured and it is shown that the best recognition result is achieved

*Corresponding author. Yicheng Jiang (email: jiangyc@hit.edu.cn), Tania Stathaki (email: t.stathaki@imperial.ac.uk).

Fig. 1. The appearances of the same subject under different viewing angles (from left to right $0^\circ$, $18^\circ$, $36^\circ$, $54^\circ$, $72^\circ$, $90^\circ$, $108^\circ$, $126^\circ$, $144^\circ$, $162^\circ$, $180^\circ$)

when recognitions are performed under the same view especially for profiles. However, in the real world, the non-cooperate individuals usually walk in random paths (in this paper, we assume subjects walk in random straight paths). It is observed in [25] that gait recognition is sensitive to varying views based on the fact that the visual features will change as the viewing angle changes. Therefore, it is not easy to achieve acceptable recognition accuracy when probe subjects are walking in random directions [22].

Several methods have been proposed for gait recognition from various different perspectives [1, 5, 15, 18–20, 23]. The state-of-the-art methods mainly fall into 3 categories, namely, view-invariant models estimation, viewing angle rectification, and view transformation model methods. One example that estimates the view-invariant model is the Joint Subspace Learning (JSL) method proposed in [20], where the view-invariant model is represented by a weighted sum of sufficiently small number of prototypes of the same view. Similarly, in [23], a distance metric is learned with a good discrimination ability based on the clustered and averaged GEI. However, this kind of methods can achieve good recognition results only for similar views. When the probe gait sequences are significantly different from the gallery sequences, they usually have poor performances. For viewing angle rectification methods, in [18], the Transform Invariant Low rank Textures (TILT) algorithm is used to rectify the deformed human silhouette into profile silhouette, and then a similarity measurement process is applied on the rectified profile silhouettes. Regarding the TILT, a generalized framework for low-rank recovery is proposed in [27] where the optimization problem is solved by adopting a proximal gradient based altering direction method. However, this kind of methods relies on the rectified result which is not stable in all circumstances. The methods proposed in [15–17] are recently published View Transformation Model (VTM) methods. All VTM-based methods construct the gait features of the target view using the information of its corresponding view, which require a technique that can find the best corresponding information among different views with stable performance. In [19], a technique is proposed for the extraction of the optimal correlation type of information. This method involves a motion co-clustering process that partitions gaits from different views into multiple groups. After that, the two most correlated gait features of two views are applied with a mapping operation to map the gait feature from one view into another view using the trained Canonical Correlation Analysis (CCA) subspaces. However, this method is quite complicated and difficult to implement. Furthermore, the performance of the VTM-based methods under distant views still requires improvement. Overall, most of present methods can achieve good recognition results for similar views but achieve poor recognition accuracy under distant views.

This paper proposes a multi-view gait recognition method that aims at achieving improved recognition accuracy compared to existing methodologies. The widely researched GEI is employed as the gait feature and the following

processes will be applied based on the GEIs. Our method includes two parts, namely, the viewing angle estimation and the multi-view gait recognition.

For viewing angle estimation, one common sense shared among multi-view gait recognition researchers is that two distant views share much less correlated information than that of similar views [15–17, 19]. In that case, recognition performed among two distant views leads to poor results. Therefore, it is reasonable to perform similarity measurement on the probe subject and its nearest-neighbouring view in the gallery. To this end, we start with estimating the viewing angle of the monitor equipment in terms of the probe subject. This is considered as a classification problem in this paper, where the classification signals are the viewing angles. Since the Transformation Invariant Low-Rank Texture (TILT) algorithm can describe the degree of difficulty of transforming an image into a low-rank image [26], the low-rank image can be considered as a standard and the TILT algorithm can be used to describe this transformation. Therefore, we propose to consider the elements of the transformation matrices estimated by the TILT algorithm as the classification features, and then employ the Extreme Learning Machine (ELM) classification method to solve this classification problem. For multi-view gait recognition, it is already known that the appearances related to different views share certain amount of correlated information [15–17, 19]. Therefore, it is reasonable to assume that the gait appearances referring to two different views are able to convert between each other with tolerable error by using the correlated information that can be considered as the connection of two views. As can be seen from Fig. 1 that one subject exhibits similar appearance among different views, especially for similar views. Here, we propose the Appearance Conversion Machine (ACM) to convert the gait appearances across views (from one view to another view), where the view is estimated using the above mentioned estimation method. The proposed ACM attempts to find the correlation function across views and then convert the subject by exploring the information related to this type of correlation. In contrast with the correlation coefficient method used in [16, 17] or the complicated co-clustering method in [19], we assume that one pixel in the target view is only highly correlated with the spatial neighbouring pixels of the source view. This enables the correlation information to be extracted in a consistent fashion among different views, and thus leads to a stable conversion result. The ELM is employed as it achieves good generalization performance and solves the regression problem in the proposed ACM. The ELM was originally developed from feed forward networks and later extended to kernel learning [13]. This extension enables the ELM to achieve higher scalability with less computational complexity [12]. Having acquired the converted gallery (training) appearance, a similarity measurement is applied on the converted appearance and the probe appearance to identify the subject. The flowchart of the proposed gait recognition method is shown in Fig. 2. For scenarios with more than 1 camera, our method can be extended using the extended ACM by employing the gait information captured from two cameras. We explore the proposed method on the CASIA-B multi-view gait database (one of the most widely used multi-view gait databases with 124 subjects captured from 11 angles). As will be shown in the relative section, experimental results show the effectiveness of the proposed method. Furthermore, the recognition performance of the extended ACM with two cameras is complementary to the ACM with single camera. The encouraging experimental results assure the possibility of monitoring the entire scene (from $0^\circ$ to $180^\circ$) with relatively high accuracy using only two cameras.

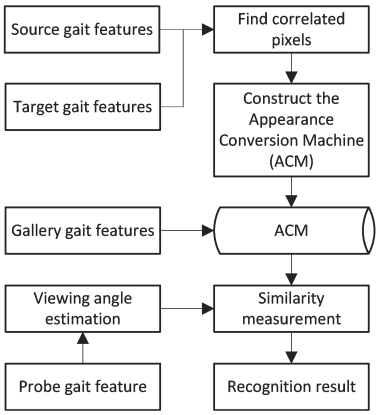



Fig. 2. The framework of the proposed recognition method.

The main contributions of this work are:

• We have proposed a robust method for estimating the viewing angle of the camera in terms of the probe subject.
• We have proposed to construct the Appearance Conversion Machine (ACM) for converting the gait appearance from one view to another view and thus increase the recognition performance of multi-view gait recognition by performing a similarity measurement based on the converted appearances.
• We have proposed a method for finding the correlated pixels between two views for the construction of the ACM.
• We have conducted a number of experiments on the widely used CASIA-B gait database and achieved significantly better recognition results than existing, widely used multi-view gait recognition methods.

The rest of this paper is organized as follows. The construction of the GEI is introduced in Section II. The viewing angle estimation method and the ACM are discussed in Section III and Section IV, respectively. Experiments and performance evaluations are discussed in Section V. Discussions and conclusions are provided in Section VI.

## II. Feature Construction

In this paper, the GEI is employed as the gait feature. To construct a GEI for each subject, image patches of walking human are cropped from the sequential images of the CASIA-B gait database. Image registration as a pre-processing step is essential for the robust performance of gait recognition. To conduct registration, all gait image patches are resized to the same scale while keeping their aspect ratio (height/weight) unchanged. Then, these resized images are registered according to their horizontal first momentum. Finally, for the $k$-th subject with $N_k$ gait image patches captured from viewing angle $\theta_v$, its GEI is constructed as

$$G_{k,\theta_v}(x,y) = \frac{1}{N_k}\sum_{i=1}^{N_k} I_{k,\theta_v,i}(x,y) \tag{1}$$

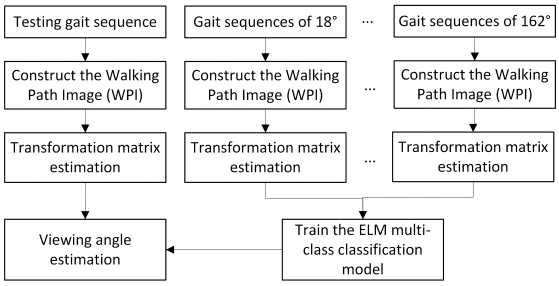



Fig. 3. Framework of the proposed viewing angle estimation method.

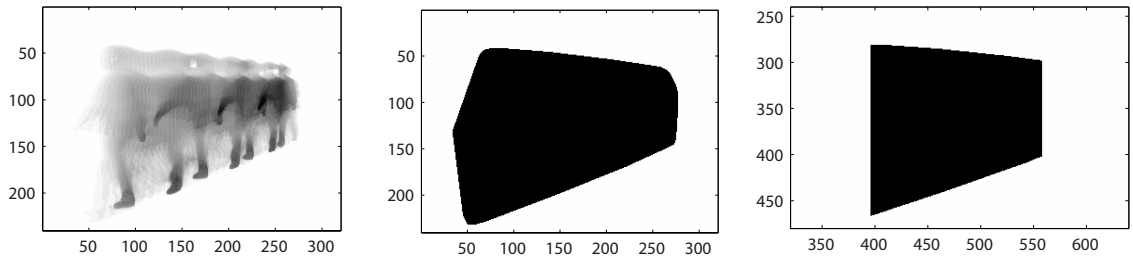
where $G_{k,\theta_v}(x, y)$ is the GEI of the $k$-th subject captured from viewing angle $\theta_v$, $I_{k,\theta_v,i}(x, y)$ is the $i$-th gait image patch of current subject $k$ captured from viewing angle $\theta_v$, and $x$ and $y$ are the horizontal and vertical coordinates of the 2-D image, respectively.

The constructed gait features are divided into two groups according to their appearances, namely the front-view group (0°, 18°, 36°, 144°, 162°, 180°) and the side-view group (54°, 72°, 90°, 108°, 126°). It is not difficult to find from Fig. 1 that two front-views with almost 180° variance are approximately the left-right overturned version of each other. Therefore, for recognition of two distant views, the front-view gait appearances of the gallery images are left-right overturned to enhance the recognition performance.

## III. Viewing Angle Estimation

It is known that performing gait recognition among similar views results in improved recognition accuracy [25]. In this section, we propose a viewing angle estimation method. Taking advantage of this estimation method, the most similar views in the gallery can be selected and the similarity measurement can be conducted between the selected gallery view and the probe view.

Our method considers the viewing angle estimation problem as a multi-class classification problem, where the classification signals are the viewing angles. Subsequent steps are used to extract proper features for the description of the viewing angles. Firstly, an image is built to record the walking path of the probe subject, which is denoted as the Walking Path Image (WPI). After that, the Transform Invariant Low-Rank Textures (TILT) technique is used to estimate the transformation matrix for transforming the WPI into a low rank image. This transformation matrix describes the degree of difficulty of this transformation. Moreover, subjects under the same view have similar transformation matrix, and subjects of different views have distinct transformation matrix. Therefore, transformation matrices are considered as the features of different viewing angles and are used to train the ELM classification model. In practical use, the transformation matrix of the WPI of the new coming subject is estimated and fed into the trained classifiers so that its viewing angle can be estimated.




(a) The averaged gait sequence. (b) The convex hull. (c) The Walking Path Image.

Fig. 4. The construction process of the WPI and its lowest rank image.

### *A. Walking Path Image*

Inspired by the Gait Texture Image (GTI) proposed in [18], the image representing the walking path of the probe subject can be constructed by averaging the gait sequence of the probe subject. In this section, we aim at constructing a relatively low-rank image with a sequence of gait images while preserving the walking path information of the subject. This is because the TILT algorithm is much more computation efficient with a low-rank image input [26]. Different with the averaging operation conducted in [18], a sequence of operations are applied here to produce an image with much lower rank.

Given a gait sequence $\{I_t(x,y)\}_{t=1}^{N_k}$ where $N_k$ is the number of images in this gait sequence and $I_t$ is the $t$-th gait image of the gait sequence, the averaged gait sequence is obtained as

$$I_{avg}(x,y) = \frac{1}{N_k}\sum_{t=1}^{N_k} I_t(x,y) \tag{2}$$

As shown in Fig. 4(a), according to the rank definition in [26], the $I_{avg}$ produced from the GTI with more varying pixels has higher rank. To achieve the rank-reduced version of $I_{avg}$, only its convex hull is kept, as shown in Fig. 4(b). We can find from the convex hull that, the upper and lower boundary reveals the walking path, and the left and right boundary reveals the body movement which is not necessary for our viewing angle estimation. Therefore, the border protruding parts are cropped in order to achieve a quadrangle-like shape, which further reduces the rank of $I_{avg}$, as shown in FIg.4(c). Here, we denote the final image as the Walking Path Image (WPI) of the probe subject. The WPI provides a relatively low rank version of $I_{avg}$ while reserving the walking path information of the probe subject.

### *B. Feature Extraction*

As described, the Transform Invariant Low-Rank Textures (TILT) algorithm is used to extract features for viewing angle estimation. According to the TILT algorithm, any image can be considered as a domain transformation (say affine or projective) of its low-rank version, and the low-rank image can be recovered from its deformed image, where a so called transformation function controls the transformation process [26]. We assume that the domain

transformation of the WPI is a projective transformation version of its low-rank image and denote the transformation matrix as $\tau$. Since the transformation matrix reveals the degree of transformation, it is able to be employed as the feature of the viewing angle. Therefore, for a given WPI of a low-rank image $I$, we have $WPI \circ \tau = I + E$. The aim is to recover the low-rank image $I$ and achieve the domain transformation matrix $\tau \in \mathbb{G}$, which can be considered as the optimization problem of finding the lowest rank of $I$ with the lowest error

$$\min_{I,E,\tau} rank(I) + \gamma \parallel E \parallel_0, s.t. WPI \circ \tau = I + E \tag{3}$$

where $\parallel E \parallel_0$ denotes the number of non-zero entries in $E$ and $\gamma$ denotes the trade off between the error sparsity and the rank of $I$.

The above problem includes the rank function and the $l^0$-norm which are NP-hard problems. However, according to [3, 6], they can be replaced by their convex surrogates (i.e. the matrix nuclear norm for the rank function and the $l^1$-norm for the $l^0$-norm, respectively). It is also noted that the optimization is not convex because the constraint $WPI \circ \tau = I + E$ is non-linear in $\tau \in \mathbb{G}$. A common technique to overcome this difficulty is to linearise the constraint around the current estimate and perform an iteration type of estimation. Here, the constraint for the linearised version of the above problem becomes

$$WPI \circ \tau + \nabla WPI \triangle \tau = I + E \tag{4}$$

where $\nabla WPI$ is the derivatives of the WPI with respect to the transformation parameters. Therefore, we end up with the optimization problem:

$$\begin{aligned} \min_{I,E,\triangle\tau} &= \parallel I \parallel_* + \lambda \parallel E \parallel_1, \\ s.t. &WPI \circ \tau + \nabla WPI \triangle \tau = I + E \end{aligned} \tag{5}$$

where $\parallel * \parallel_*$ denotes the nuclear norm and $\parallel * \parallel_1$ denotes the $l^1$-norm.

The above problem is a local approximation to the original non-linear problem, TILT solves it iteratively in order to converge to a (local) minimum of the original non-convex problem [26].

## IV. Multi-view Gait Recognition

In this section, we introduce a technique for converting the appearance of the probe subject from various views to the target view. After conversion, the new appearance should be approximately similar to the appearance of the same subject captured from the target view while keeping its discriminative information for distinguishing from other subjects.

### A. Appearance Conversion Machine

Since GEI is employed as the gait feature, $G_{k,\theta_v}$ is used here to denote the gait appearance of the $k$-th subject captured from viewing angle $\theta_v$. We denote $M_{\theta_i,\theta_j}$ as the conversion machine for converting the subject from viewing angle $\theta_i$ to $\theta_j$. Thus, the aim is mathematically described as

$$G_{k,\theta_j}(x,y) = M_{\theta_i,\theta_j}(G_{k,\theta_i}(x,y)) \tag{6}$$

where $M_{\theta_i,\theta_j}(*)$ indicates the conversion conducted by the mapping function.

For any appearance that corresponds to the view from a specific angle, it is clear that perfect conversion to the appearance that corresponds to another angle's view, without sufficient information provided, is not a simple task. More specifically, in most cases, the information associated with the correlation between the appearances of the two views is not sufficient enough to yield perfect construction of the appearance of one view based on the appearance of another view. However, if certain amount of errors can be tolerated, it is possible to estimate one appearance based on the correlation information provided by another view[18]. Therefore, Equ. 6 can be approximated as

$$\min_{M_{\theta_i,\theta_j}} ||G_{k,\theta_j}(x,y) - M_{\theta_i,\theta_j}(G_{k,\theta_i}(x,y))||_F^2 \tag{7}$$

Equ. 7 can be considered as the estimation of the mapping function $y = f(x)$ where $x = G_{k,\theta_i}(x,y)$, $y = G_{k,\theta_j}(x,y)$ and $f(*) = M_{\theta_i,\theta_j}(*)$, which can be solved by regarding this as a regression problem. The above mapping function is denoted as the Appearance Conversion Machine (ACM) in this paper. The problem of converting the appearance from one view into another view with the achieved mapping function $M_{\theta_i,\theta_j}$ is solved by using the Extreme Learning Machine (ELM), which will be introduced in Section IV-A1.

It is not difficult to know that pixels of $G_{k,\theta_j}$ are not related to all the pixels of $G_{k,\theta_i}$. For the estimation of every single pixel in $G_{k,\theta_j}$, only the correlated pixels in $G_{k,\theta_i}$ are used. That is, given the GEI with $R$ rows and $C$ columns, the above problem is divided into $R \times C$ regression problems,

$$M_{\theta_i,\theta_j} = \begin{pmatrix} m^{1,1}_{\theta_i,\theta_j} & \cdots & m^{1,C}_{\theta_i,\theta_j} \\ \vdots & m^{x,y}_{\theta_i,\theta_j} & \vdots \\ m^{R,1}_{\theta_i,\theta_j} & \cdots & m^{R,C}_{\theta_i,\theta_j} \end{pmatrix} \tag{8}$$

$$m^{x,y}_{\theta_i,\theta_j} = \min_{m^{x,y}_{\theta_i,\theta_j}} [p_{k,\theta_j}(x,y) - m^{x,y}_{\theta_i,\theta_j}(q_{k,\theta_i}(x,y))]^2 \tag{9}$$

where $p_{k,\theta_j}(x,y)$ is the target pixel in $G_{k,\theta_j}$ and $q_{k,\theta_i}(x,y)$ is the group of correlated pixels of $G_{k,\theta_i}$ in terms of $p_{k,\theta_j}(x,y)$.

*1) Implementation Details:* The Extreme Learning Machine (ELM) is employed in this paper to solve the proposed regression problem. Compared to other types of Single-hidden-Layer Feed forward Networks (SLFNs),

also called support vector network [4], ELM achieves better generalization performance with lower computation complexity. This is because instead of tuning the hidden layers of normal SLFNs or using the bias term $b$ as SVM, random kernels are used in ELM where fewer optimization constraints result in simpler implementation and thus lead to higher scalability [13]. The output function of the ELM for generalized SLFNs (one output node case) is

$$m_L(x) = \sum_{i=1}^{L} \beta_i h_i(x) = h(x)\beta \tag{10}$$

where $h(x) = [h_1(x), \cdots, h_L(x)]$ is a feature mapping, $\beta = [\beta_1, \cdots, \beta_L]^T$ is the vector of the output weights, and the input data $x$ from the $d$-dimensional input space are mapped to the $L$-dimensional ELM feature space.

The input $x$ consists of the pixels of the grey GEI and lie in the space of $x \in [0, 255]$ which is a bounded measurable compact subset of the Euclidean space $\mathbf{R}^d$. As described in the ELM learning theory [8], a widespread type of feature mapping $h(x)$ can be used in the ELM so that ELM can approximate any continuous target functions. That is, given any continuous target function $m(x)$, there exists a series of $\beta_i$ such that

$$\lim_{L\to\infty} \parallel m_L(x) - m(x) \parallel = \lim_{L\to\infty} \parallel \sum_{i=1}^{L} \beta_i h_i(x) - m(x) \parallel = 0 \tag{11}$$

Then, the algorithm is described as,

**Input**: Formulate a training set $\aleph = \{(q^i_{k,\theta_i}, p^i_{k,\theta_j}) | q^i_{k,\theta_i} \in \mathbf{R}^{N_s}, p^i_{k,\theta_j} \in \mathbf{R}^1, i = 1, \cdots, N\}$, an activation function $h(x)$, and decide on the hidden node number $N$ .

**Step 1**: Randomly assign input weights $w_i$ and biases $b_i, i = 1, \cdots, \tilde{N}$.

**Step 2**: Calculate the hidden layer output matrix $\mathbf{H}$.

**Step 3**: Calculate the output weight $\beta = \mathbf{H}^{\dagger}\mathbf{P}$, where $\mathbf{P} = [p^1_{k,\theta_j}, \cdots, p^N_{k,\theta_j}]^T$.

where $N_s$ is the number of variables in $q^i_{k,\theta_i}$ and $\mathbf{H}$ is the hidden-layer output matrix

$$\mathbf{H} = \begin{bmatrix} h(q^1_{k,\theta_i}) \\ \vdots \\ h(q^N_{k,\theta_i}) \end{bmatrix} = \begin{bmatrix} h_1(q^1_{k,\theta_i}) & \cdots & h_L(q^1_{k,\theta_i}) \\ \vdots & \cdots & \vdots \\ h_1(q^N_{k,\theta_i}) & \cdots & h_L(q^N_{k,\theta_i})) \end{bmatrix} \tag{12}$$

As introduced in [11], almost all non-linear piecewise continuous functions can be used in the feature mapping process. For any given non-linear piecewise continuous function $G(a, b, q_{k,\theta_i})$ satisfying the ELM universal approximation capability theorems [8–11] and $\{(a_i, b_i)\}_{i=1}^{L}$ randomly generated from any continuous probability function, we can have

$$h(x) = [G(a_1, b_1, q_{k,\theta_i}), \cdots, G(a_L, b_L, q_{k,\theta_i})] \tag{13}$$

Readers may refer to [14] for more information about the ELM.

Fig. 5. The target GEI ($54^\circ$) is on the left side with the target pixel labelled in red, the GEI of $90^\circ$ is on the right side with the correlation pixels confined in the red square.

*2) Region of Interest Selection:* To construct the conversion machine $M_{\theta_i,\theta_j}$ for converting gait appearances from $\theta_i$ to $\theta_j$, the sub conversion machine $m^{(x,y)}_{\theta_i,\theta_j}$ is constructed for every single target pixel in $G_{k,\theta_j}(x,y)$. Therefore, each pixel in target GEI is related to a regression process where the input is a group of pixels selected from its source GEI $G_{k,\theta_i}(x,y)$. Several methods have been proposed for finding the most correlated pixels. In [15], a Linear Discriminant Analysis is applied to get the rank-reduced image, which is considered to contain the information associated with correlation. In [16, 17], images are separated into 6 parts and the most correlated pixels are found in these 6 parts using the correlation coefficient method. In [19], images are partitioned into several parts and co-clustered into several motion co-clustering groups using the bipartite graph multi partitioning method. Their exploration of gait appearances across views have yield to the conclusion that a particular body part from one view possesses stronger correlation relationship with the same part from another view than with other parts [19]. In contrast with these methods, we assume that the most related pixels lie in the spatial neighbourhood around of the position of the target pixel. A much simpler correlation based pixel selection method is proposed here. The pixels in the square patch within the square patch surrounding the position of the target pixel are considered as the most related pixels and are taken as the input of the regression process, as shown in Fig. 5. Let $(p_{x_t}, p_{y_t})$ be the coordinate of the target pixel. The Region of Interest (ROI) is used here to denote the square patch $q_{k,\theta_i}(x,y)$ that contains the majority of correlated pixels. It is described as,

$$\begin{aligned} &q_{k,\theta_i}(x,y) = \{t(x_1,y_1),\cdots,t(x_n,y_n),\cdots,t(x_{N_s},y_{N_s})\}, \\ &p_{x_t}-s<x_n<p_{x_t}+s, p_{y_t}-s<y_n<p_{y_t}+s \end{aligned} \tag{14}$$

where $t(x_n,y_n)$ is the correlated pixel in $G_{k,\theta_i}(x,y)$, $N_s=(2s+1)^2$ is the number of pixels in the ROI, and $s$ is the parameter determining the scale of the ROI.

### *B. Extended Appearance Conversion Machine*

As will be shown in the experimental results section, conducting the appearance conversion between two similar views (within $36^\circ$) is enable us to achieve an acceptable recognition rate (mostly higher than $80\%$). However, a conversion between two distant views results in a poor recognition rate (lower than $70\%$). As mentioned in SectionIV-A, the reason is that stronger correlation exists between two similar views but less between two distant views. In fact, the lack of the correlation can be compensated by constructing the ACM with more than 1 view.

To do this, two views' ROIs are concatenated together as the input of the ELM. The extended ROI is denoted by $ROI_e$

$$ROI_e = \{q_{k,\theta_1}(x,y), \cdots, q_{k,\theta_E}(x,y)\} \tag{15}$$

where $E$ is the number of the involved ROI.

## V. Experimental Results

For experimental evaluation, we test the proposed method on the CASIA-B multi-view gait database. This database contains the gait data of 124 subjects captured from 11 views ($0^\circ$, $18^\circ$, $36^\circ$, $54^\circ$, $72^\circ$, $90^\circ$, $108^\circ$, $126^\circ$, $144^\circ$, $162^\circ$, $180^\circ$) with frame size $320 \times 240$, where each is represented by a 6-gait sequence. Therefore, there are totally $124 \times 11 \times 6 = 8184$ gait appearances sequences, and $124 \times 6 = 744$ gait appearances for each view. The height of the subjects within frames vary in the range of $(60, 240)$ pixels. For the sake of computational efficiency and avoidance of the introduction of useless information in the process of image enlargement (interpolation), all template patches are resized to $60 \times 60$ for feature construction (i.e. the construction of the GEIs). Furthermore, the proposed method is implemented using Matlab R2013a and tested on a computer with 3.4GHz CPU and 8GB RAM.

Since the methods for comparison don't involve the process of view-angle estimation, to test the performance of the proposed method, we evaluate the performance of the view-angle estimation method in Section 5.2 and then evaluate the ACM method in Section 5.3, respectively. In this way, we can focus on the performance comparison between the proposed ACM and other methods. ACMs are constructed for every combination of two views using the ELM. Since the main purpose of this paper is to introduce the multi-view gait recognition method, the popular default kernel, namely, the Radial Basis Function (RBF) kernel is used in the ELM for performance evaluation. The gallery GEIs of the source view $\theta_i$ are converted to the target view $\theta_j$ using its ACM $M_{\theta_i,\theta_j}$. The database is divided into 2 parts: the first part contains 24 subjects for training and the second part contains 100 subjects for evaluation. Specifically, training stage includes 24 subjects and 1 of the 6 sequences of each subject is used (i.e. 24 gait appearance sequences are used) for each ACM. In the similarity measurement stage, the similarity is measured using the Euclidean distance between the gallery-converted GEI and the probe GEI. The GEIs with the smallest Euclidean distance will be identified.

Regarding the computational complexity, the training process requires less than 0.3 seconds in order to construct one ACM model, and 0.1 to 0.5 seconds to achieve the converted appearance depending on the selected parameters. It is worth mentioning that both the ACM construction process and the appearance conversion process can be completed off-line beforehand. More specifically, each gallery appearance can be converted to the target view using its ACM in advance. Therefore, only the similarity measurement process is required for recognition, which requires less than 0.01 seconds in order to achieve the final result.

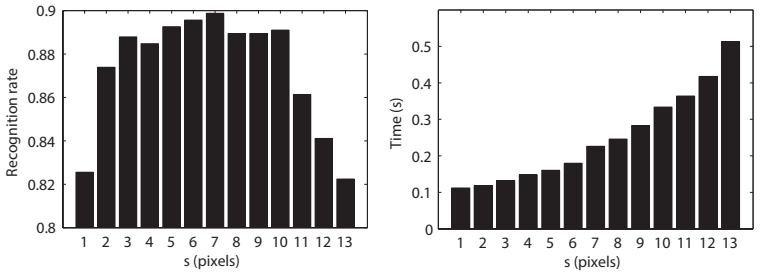



Fig. 6. The impact of the ROI scale. The probe view is $90^\circ$ and the gallery view is $54^\circ$.

### A. Parameters Analysis

A number of factors affect the performance of the proposed method, e.g., the scale of the ROI, the number of the training subjects, the length of the gait sequence for gait-appearance construction, and the parameter settings of the ELM.

*1) The ROI Scale:* The ROI scale is a factor of crucial importance for the construction of the ACM. This is because the ACM aims at converting the gait appearance across views, which relies on the correlated information between views. The scale of the ROI determines the quality and the quantity of the effective pixels for appearance conversion.

The recognition accuracy and the elapsed time are measured to test the impact of the ROI scale with a scale range from 1 to 13. 24 subjects are involved to construct the ACM. The probe view is $90^\circ$ and the gallery view is $54^\circ$. As shown in Fig. 6, the recognition performance of the ACM appears as a bell shaped curve as the ROI scale changes, and the best result is achieved when the scale is 7, which implies that $(2 \times 7 + 1)^2 = 225$ pixels (i.e. the size of the patch is $15 \times 15 = 225$) are involved in the regression process to construct the ACM for one target pixel. Furthermore, the scale is also a function of the size of the template images. It is not difficult to conclude that the best recognition result is achieved when the size of the patch divided by the size of the template images is equal to $225/3600 = 9/144$. The results shown in Fig. 6 meet our expectation that a proper scale introduces sufficient effective pixels with enough information to construct an ACM. Smaller quantity of pixels provides insufficient information while larger quantity decreases the ratio of efficient information. More specifically, the ratio of pixels that has the optimal type of correlation information affects the performance of the ACM. In terms of the running time, it is quite straightforward to find that a larger scale introduces more pixels and thus consumes more time than a smaller scale. In consideration of the computation time and recognition accuracy, a smaller ROI scale $s = 3$ is used in the following experiments.

*2) The Number of Training Subjects:* To construct an effective ACM, sufficiently high quality information should be provided as mentioned in Section V-A1. In fact, the quality of the provided information is not only influenced by the ROI scale but also the number of subjects for ACM construction. The number of training subjects is increased from 1 to 51 to test its impact. In other words, a maximum of 51 subjects are used for ACM construction and the

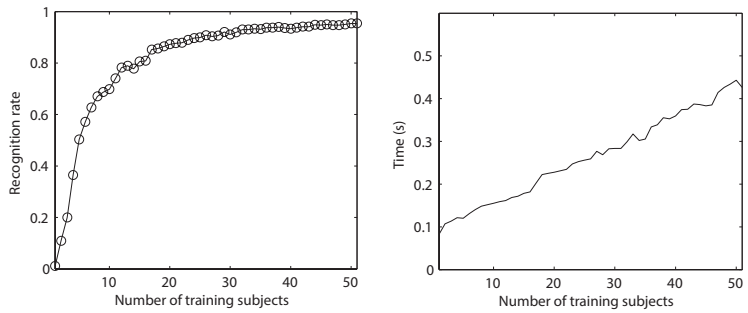



Fig. 7. The impact of the number of training subjects. The probe view is $90^\circ$ and the gallery view is $54^\circ$.

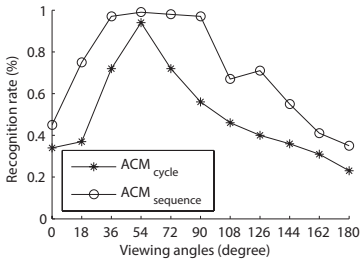



Fig. 8. The impact of the length of the involved gait sequence.

rest of the subjects are not part of the training set. The probe view is $90^\circ$ and the gallery view is $54^\circ$. It can be seen from Fig. 7 that as the number of training subjects increases, a more complete ACM is built and thus results in a better recognition result. Moreover, too many training subjects also results in an increase of the training time.

*3) The Length of Involved Gait Sequence:* For all captured gait sequences, we assume that their gait information is evenly and distributed over the entire gait sequence. This is because the gait sequential images are actually obtained by discrete sampling of the continuous walking movement. Therefore, a complete gait feature requires sufficiently enough gait images of a gait sequence. To test the impact of the number of involved gait images selected from a gait sequence, two types of ACMs are constructed: the $ACM_{cycle}$ involves the gait images of one gait cycle and the $ACM_{sequence}$ involves the entire gait sequence. The probe view is $54^\circ$ and the gallery view varies from $0^\circ$ to $180^\circ$. As can be seen from Fig. 8, the $ACM_{sequence}$ achieves better recognition results than the $ACM_{cycle}$. This is because the performance of the ACM depends on the completeness of the involved gait information, and the involvement of the entire gait sequence builds a more complete gait feature and thus the ACM is able to convert the views in a more accurate way.

*4) The ELM parameters:* In this section, we test the performance of the ACM with different ELM parameters. We have tried a wide range of values ($\{1, 51, \cdots, 1451, 1501\}$) for the cost parameter $C$ and kernel parameter $P$. The probe view is $90^\circ$ and the gallery view is $54^\circ$. As can be seen from Fig. 9 that the proposed ACM is not sensitive to parameter selection, as long as the cost parameter $C$ and the kernel parameter $P$ are larger than

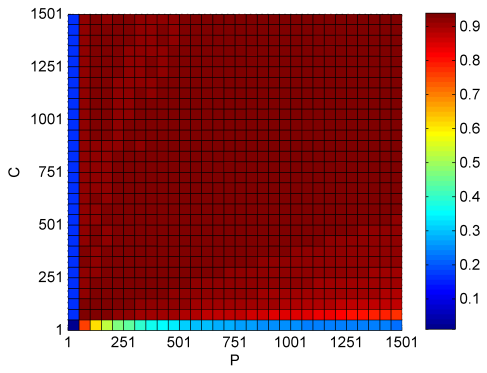



Fig. 9. The recognition rate of the ACM with different ELM parameters. The probe view is $90°$ and the gallery view is $54°$.

51. This experimental result ensures that no cumbersome human interference is needed for achieving proper ELM parameters that can achieve good recognition performance.

### *B. Viewing Angle Estimation*

9 viewing angles ($18°$, $36°$, $54°$, $72°$, $90°$, $108°$, $126°$, $144°$, $162°$) are involved for testing the performance of the viewing angle estimation method. 486 randomly selected gait sequences of 9 views ($54 \times 9 = 486$) are used as the training data and all gait sequences ($124 \times 9 \times 6 = 6696$) are participate as the testing data.

To test the impact of various ELM parameter values, namely, the cost parameter $C$ and the kernel parameter $P$, we have tried a wide range ($\{2^{-15}, 2^{-14}, \cdots, 2^{14}, 2^{15}\}$) of $C$ and $P$. It can be seen from Fig. 10 that the highest accuracy $93.37\%$ is achieved when $C = 2^{14}$ and $P = 2^6$. We also tested to training with more gait sequences, e.g. 900 sequences ($100 \times 9 = 900$), 1800 sequences ($200 \times 9 = 1800$), and 3600 sequences ($400 \times 9 = 3600$). The highest accuracy for 900 training gait sequences is $94.81\%$ which is achieved when $C = 2^{13}$ and $P = 2^6$. The highest accuracy for 1800 training gait sequences is $96.14\%$ which is achieved when $C = 2^{14}$ and $P = 2^2$. The highest accuracy for 3600 training gait sequences is $97.73\%$ which is achieved when $C = 2^{13}$ and $P = 2^4$. The achieved good results give credit to the particular selection of the representative view-angle feature and the characteristic of the ELM as discussed in Section IV.

An additional test that measures the number of misestimates in the experiment of Fig. 10(d) is shown in Fig. 10(e) where the ground truths are shown in blue line and the misestimates are shown as red-dotted lines. We count the number of misestimates which fall into the adjacent views $[-18°, 18°]$ and the number of misestimates which fall into distant views. It is observed that only 26 of 132 misestimates fall into distant views, most of which were found in $18°$ and $162°$. This is because obvious distortions can be found in the WPI constructed with views that are captured from $18°$ and $162°$, which may impair the performance of the TILT method. Regarding the obvious distortions, we are trying to say that the WPI generated from $18°$ or $162°$ has a quadrangle border with higher gradient. This can be easily deduced from Fig. 4(c) which is generated from the $36°$ averaged-gait-sequence image as shown in Fig. 4(a). Specifically, TILT aims at estimating the low-rank version of a given image, where an

obvious distorted quadrangle shape can significantly increase the difficulty of the estimation [26]. The rest of the misestimates fall into adjacent views, where most existing gait recognition methods are able to achieve acceptable recognition performance (higher than $90\%$). It is also noted that larger training data can construct a much more complete model and thus result in higher accuracy. Therefore, it is easy to conclude that the proposed viewing angle estimation method provides high enough accuracy for the subsequent multi-view recognition.

### *C. Cross-view Recognition*

Comparison with other methods is conducted to test the performance of the proposed ACM. Since the method proposed in [19] achieves much better recognition performance than other existing methods according to various authors [6, 15–17, 23], a concise comparison is conducted by comparing the proposed ACM only with methods that achieved the state-of-the-art performance under certain views, e.g., the method proposed in [19], the GEI-SVR method in [17], the view verification method in [6], and the method in [23]. To provide a fair comparison, we use 24 subjects to construct the ACM with the same database as the one used in [19]. Moreover, since no viewing angle estimation is performed in other methods, all tests are conducted with known viewing angles rather than estimated. Four probe views are involved to represent four common situations: $0^\circ$ for front view, $54^\circ$ and $126^\circ$ for oblique view, and $90^\circ$ for side view.

From Table I, it can be seen that our method outperforms other methods in almost every view combinations. For the majority of the cross-view tests, ACM is $10 \sim 20\%$ more accurate than other methods. Moreover, comparable performances are achieved in the remaining view combinations. As mentioned in Section.II, the symbol $(t)$ in the table indicates constructing the ACM with the left-right overturned versions of the gallery image. The reason of using the left-right overturned version of the gallery images is that front-views are not strictly symmetric. Moreover, for subjects captured with large viewing-angle variation in comparison with the gallery view, their gait appearances are left-right overturned version of the gallery images at certain extent. Therefore, the left-right overturned appearances provide more accurate information for ACM conversion and lead to better recognition results. Overall, the proposed method selects the ROI by simply cropping a rectangular patch around each pixel, a process which is easy to implement, and demonstrates high accuracy and good generalization performance.

### *D. Multi-view Gait Recognition*

To provide a fair comparison, 24 subjects are used to construct the extended ACM, which is the same as in [17, 19, 24]. Since similar comparison is already shown in [19], to give a concise comparison, only methods with comparable performance are listed. As shown in Table II, the proposed method and the methods in [17, 19] achieve similar performance for similar views' recognition. However, the proposed method significantly outperforms the methods in [17, 19, 24] under large view variations. This result assures the possibility of constructing a human gait recognition system with high recognition accuracy, where 2 or more cameras are used to monitor distant views (e.g. $0^\circ$ and $108^\circ$).

TABLE I

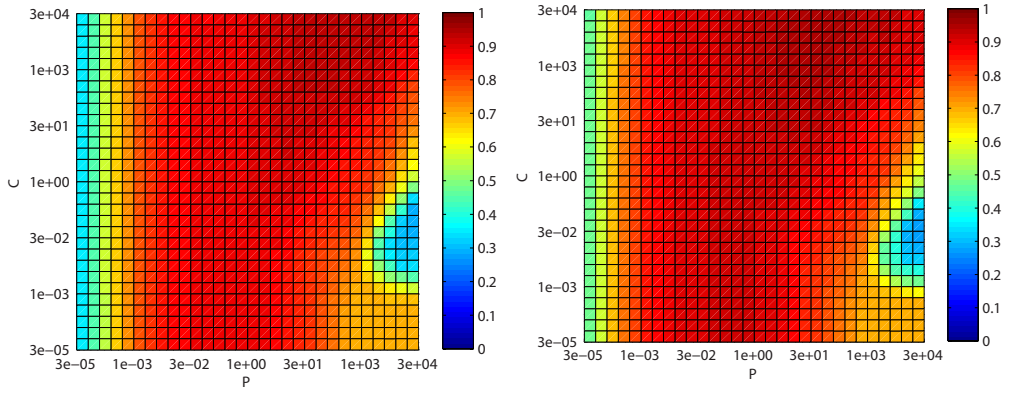
COMPARISON BETWEEN THE PERFORMANCE OF THE ACM AND OTHER METHODS

| Probe view $\theta_j$ | $0°$ | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Gallery view $\theta_i$ | $18°$ | $36°$ | $54°$ | $72°$ | $90°$ | $108°$ | $126°$ | $144°$ | $162°$ | $180°$ |
| Method in [23] | - | - | - | - | - | - | - | - | - | - |
| View-rectification [6] | - | - | - | - | - | - | - | - | - | - |
| GEI-SVR [17] | 84 | 45 | 23 | 23 | 25 | 21 | 23 | 29 | 67 | 94 |
| Method in [19] | 85 | 47 | 26 | 25 | **28** | 25 | 27 | 37 | 68 | **95** |
| Proposed ACM | **88** | **70** | **46** | **31** | 24 | **29**$(t)$ | **36**$(t)$ | **56**$(t)$ | **73**$(t)$ | $94(t)$ |
| Probe view $\theta_j$ | $54°$ | | | | | | | | | |
| Gallery view $\theta_i$ | $0°$ | $18°$ | $36°$ | $72°$ | $90°$ | $108°$ | $126°$ | $144°$ | $162°$ | $180°$ |
| Method in [23] | - | - | - | - | - | - | - | - | - | - |
| View-rectification [6] | - | - | 57 | 65 | 62 | 63 | **63** | - | - | - |
| GEI-SVR [17] | 22 | 64 | 95 | 93 | 59 | 51 | 42 | 27 | 20 | 21 |
| Method in [19] | 24 | **65** | **97** | 95 | 63 | 53 | 48 | 34 | 23 | 22 |
| Proposed ACM | **46** | 63 | 95 | **97** | **83** | **71** | 62 | **50** | **40** | **43**$(t)$ |
| Probe view $\theta_j$ | $90°$ | | | | | | | | | |
| Gallery view $\theta_i$ | $0°$ | $18°$ | $36°$ | $54°$ | $72°$ | $108°$ | $126°$ | $144°$ | $162°$ | $180°$ |
| Method in [23] | - | - | - | 68 | 94 | **96** | 70 | - | - | - |
| View-rectification [6] | - | - | **53** | 74 | 73 | 69 | 67 | - | - | - |
| GEI-SVR [17] | 16 | 22 | 35 | 63 | 95 | 95 | 65 | 38 | 20 | 13 |
| Method in [19] | 18 | 24 | 41 | 66 | 96 | 95 | 68 | 41 | 21 | 13 |
| Proposed ACM | **24** | **28** | 50 | **83** | **97** | **96** | **80** | **54** | **32** | **24** |
| Probe view $\theta_j$ | $126°$ | | | | | | | | | |
| Gallery view $\theta_i$ | $0°$ | $18°$ | $36°$ | $54°$ | $72°$ | $90°$ | $108°$ | $144°$ | $162°$ | $180°$ |
| Method in [23] | - | - | - | - | - | - | - | - | - | - |
| View-rectification [6] | - | - | 45 | 57 | 60 | 70 | 68 | - | - | - |
| GEI-SVR [17] | 22 | 26 | 26 | 42 | 57 | 78 | **98** | **98** | 74 | 19 |
| Method in [19] | 25 | **29** | 35 | 49 | 60 | 78 | **98** | **98** | 75 | 22 |
| Proposed ACM | **43**$(t)$ | 27 | **48** | **63** | **71** | **86** | 96 | 95 | **78** | **45** |

TABLE II

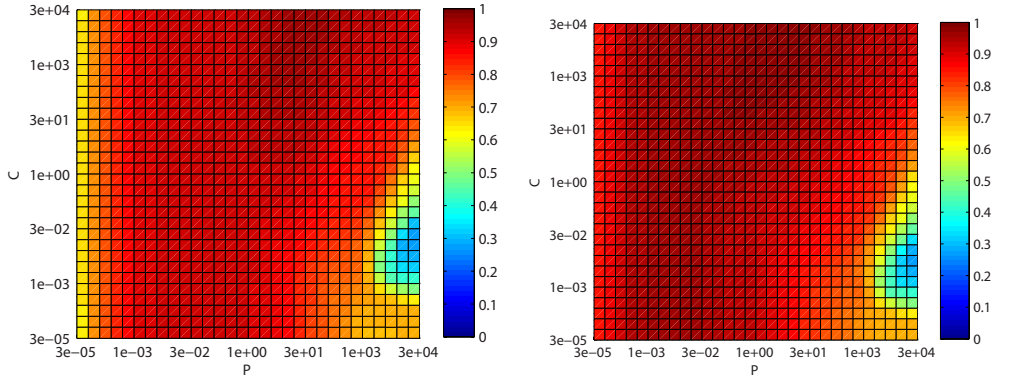
COMPARISON OF THE PERFORMANCE OF THE EXTENDED ACM WITH OTHER METHODS

| Gallery view $\theta_i$ | $54°$ | | | $126°$ | | |
|---|---|---|---|---|---|---|
| Probe view $\theta_j$ | $36°$, $72°$ | $18°$, $90°$ | $0°$,$108°$ | $108°$, $144°$ | $90°$, $162°$ | $72°$,$180°$ |
| FT-SVD [24] | 72 | 55 | 33 | 86 | 63 | 32 |
| GEI-SVR [17] | **99** | 80 | 54 | 98 | 88 | 54 |
| Method in [19] | **99** | 83 | 57 | **99** | 90 | 60 |
| extended ACM | 98 | **96** | **88** | 96 | **93** | **82** |



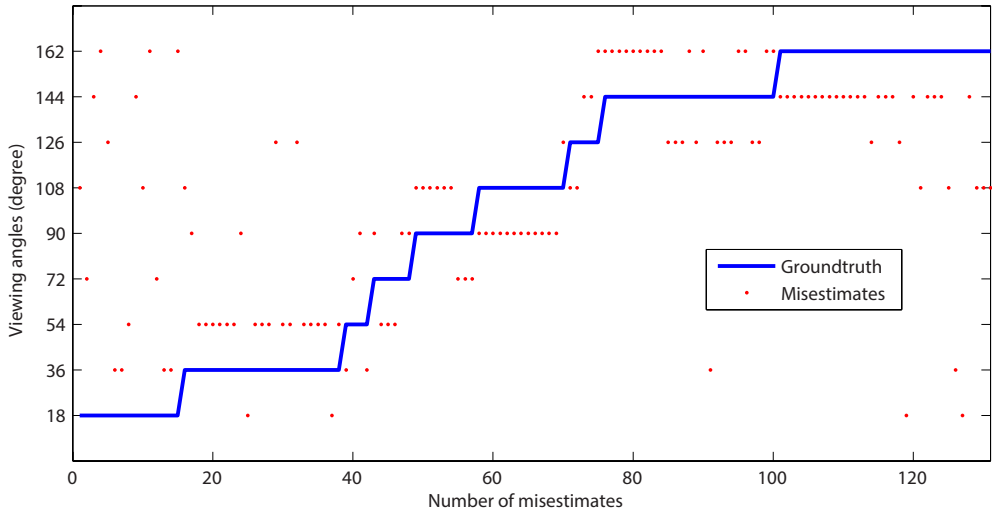

(a) 486 training sequences and 6696 testing sequences

(b) 900 training sequences and 6696 testing sequences




(c) 1800 training sequences and 6696 testing sequences

(d) 3600 training sequences and 6696 testing sequences




(e) 132 misestimates of the test in (d)

Fig. 10. The generalization performance of the proposed viewing angle estimation method, where the estimation accuracy is indicated by its colour.

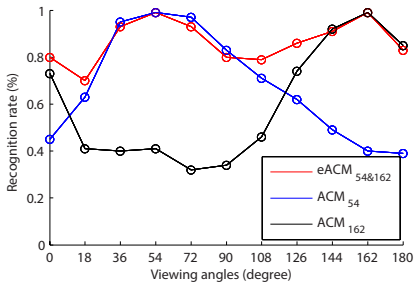



Fig. 11. The complementary performance of the ACM and the extended ACM under various views.

### *E. Comparing the ACM and the extended ACM*

In this section, we set up an additional experiment that concerns a hypothetical scenario where two fixed cameras are mounted in a scene. This experiment looks at the performance comparison of the ACM and the extended ACM under 11 different views (from $0^\circ$ to $180^\circ$). Specifically, the gait appearance of $54^\circ$ and $162^\circ$ are involved as the gallery data, and the gait appearance of 11 views ($0^\circ$, $18^\circ$, $36^\circ$, $54^\circ$, $72^\circ$, $90^\circ$, $108^\circ$, $126^\circ$, $144^\circ$ and $162^\circ$) are involved as the probe data. It can be seen from Fig. 11 that two ACMs achieve high recognition accuracy (higher than $95\%$) around slightly varying views ($-18^\circ \sim 18^\circ$) while the extended ACM achieves better performance than the ACM under distant views. Therefore, we can conclude from the results that the combination of the ACM and the extended ACM show complementary recognition performance and thus it is able to cover the viewing range from $0^\circ$ to $180^\circ$ with relatively high accuracy.

## VI. Discussion and Conclusions

This paper proposed a gait recognition method for subjects walking in arbitrary straight directions. More specifically, the viewing angle of the monitor equipment in terms of the probe subject is estimated using the proposed viewing angle estimation method. Then, the gallery gait appearance is converted to the estimated view using the proposed ACM, and finally conduct the similarity measurement. We have explored the proposed method on the CASIA-B multi-view gait database and achieved better recognition results than other methods under most views. This is because the proposed pixel's correlation extraction method is consistent in the ACM construction process, the gallery training process, and the testing process, and because the ELM based ACM is able to achieve good generalization performance. The limitation of the ACM is that good conversion results are only achieved for similar views. This is because the similar views share more correlated information for appearance conversion. To overcome this limitation, the extended ACM achieves better recognition result for distant views by taking advantage of the fact that using two views can introduce more correlated information for appearance conversion. Experimental results have shown the effectiveness of the extended ACM.

## ACKNOWLEDGEMENT

The authors would like to thank the anonymous reviewers for their valuable comments and suggestions to improve the manuscript. This paper was supported by the China Scholarship Council (CSC) for 1 year study at Imperial College London, file No. 201306120111.

## REFERENCES

[1] I. Bouchrika and M. S. Nixon. Model-based feature extraction for gait analysis and recognition, 2007.

[2] N. V. Boulgouris and Z. W. X. Chi. Gait recognition using radon transform and linear discriminant analysis. *IEEE Transactions on Image Processing*, 16(3):731–740, 2007.

[3] Sanghavi S. Parrilo P. A. Chandrasekaran, V. Rank-sparsity incoherence for matrix decomposition. *SIAM Journal on Optimization*, 21(2):572–596, 2011.

[4] C. Cortes and V. Vapnik. Support-vector networks. *Machine Learning*, 20(3):273–297, 1995.

[5] A. Elgammal and C.-S. Lee. Gait tracking and recognition using person-dependent dynamic shape model. In *IEEE International Conference Workshops on Automatic Face Gesture Recognition.*, pages 553–559, 2006.

[6] Bouchrika I. Carter J. N. Goffredo, M. Self-calibrating view-invariant gait biometrics. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 40(4):997–1008, 2010.

[7] J. Han and B. Bhanu. Individual recognition using gait energy image. *IEEE Transactions on Pattern Analysis And Machine Intelligence*, 28(2):316–322, 2006.

[8] Chen L. Huang, G. B. and C. K. Siew. Universal approximation using incremental constructive feedforward networks with random hidden nodes. *IEEE Transactions on Neural Networks*, 17(4):879–892, 2006.

[9] Ding X. J. Huang, G. B. and H. M. Zhou. Optimization method based extreme learning machine for classification. *Neurocomputing*, 74(1-3):155–163, 2010.

[10] G. B. Huang and L. Chen. Convex incremental extreme learning machine. *Neurocomputing*, 70(16-18):3056–3062, 2007.

[11] G. B. Huang and L. Chen. Enhanced random search based incremental extreme learning machine. *Neurocomputing*, 71(16-18):3460–3468, 2008.

[12] X. Huang and N. V. Boulgouris. Gait recognition with shifted energy image and structural feature extraction. *IEEE Transactions on Image Processing*, 21(4):2256–68, 2012.

[13] Zhou H. Ding X. Huang, G. B. Extreme learning machine for regression and multiclass classification. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 42(2):513–29, 2012.

[14] Zhu Q. Y. Huang, G. B. and C. K. Siew. Extreme learning machine: Theory and applications. *Neurocomputing*, 70(1-3):489–501, 2006.

[15] Wu Q. Li H. Kusakunniran, W. Multiple views gait recognition using view transformation model based on optimized gait energy image. In *2009 12th IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, pages 1058–1064. IEEE, 2009.

[16] Wu Q. Zhang J. Kusakunniran, W. Multi-view gait recognition based on motion regression using multilayer perceptron. In *2010 20th International Conference on Pattern Recognition (ICPR)*, pages 2186–2189. IEEE, 2010.

[17] Wu Q. Zhang J. Kusakunniran, W. Support vector regression for multi-view gait recognition based on local motion feature selection. In *2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 974–981. IEEE, 2010.

[18] Wu Q. Zhang J. Kusakunniran, W. A new view-invariant feature for cross-view gait recognition. *IEEE Transactions on Information Forensics And Security*, 8(10):1642–1653, 2013.

[19] Wu Q. Zhang J. Kusakunniran, W. Recognizing gaits across views through correlated motion co-clustering. *IEEE Transactions on Image Processing*, 23(2):696–709, 2014.

[20] Lu J. W. Liu, N. N. and Y. P. Tan. Joint subspace learning for view-invariant gait recognition. *IEEE Signal Processing Letters*, 18(7):431–434, 2011.

[21] Z. Liu and S. Sarkar. Simplest representation yet for gait recognition: Averaged silhouette. In *2004 Proceedings of the 17th International Conference on Pattern Recognition (ICPR)*, volume 4, pages 211–214. IEEE, 2004.

[22] J. Lu and Y.-P. Tan. Uncorrelated discriminant simplex analysis for view-invariant gait signal computing. *Pattern Recognition Letters*, 31(5):382393, 2010.

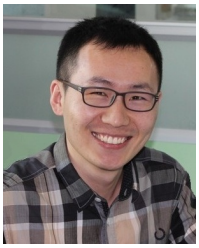
[23] Wang G. Lu, J. and P. Moulin. Human identity and gender recognition from gait sequences with arbitrary walking directions. *IEEE Transactions on Information Forensics And Security*, 9(1):51–61, 2014.

[24] Sagawa R. Mukaigawa Y. Makihara, Y. Gait recognition using a view transformation model in the frequency domain, 2006.

[25] Tan D. Yu, S. and T. Tan. Modelling the effect of view angle variation on appearance-based gait recognition. In *Computer Vision ACCV 2006*, pages 807–816. Springer Berlin Heidelberg, 2006.

[26] Ganesh A. Liang X. Zhang, Z. D. Tilt: Transform invariant low-rank textures. *International Journal of Computer Vision*, 99(1):1–24, 2012.

[27] Wang D. Zhou Z. Zhang, X. Simultaneous rectification and alignment via robust recovery of low-rank tensors. *Advances in Neural Information Processing Systems*, page 18, 2013.



**Xiaohui Zhao** was born in inner Mongolia, China. He is currently pursuing the Ph.D. degree in Information and Communication Engineering in the Research Institute of Electronic Engineering Technology, Harbin Institute of Technology. He was also studied in Imperial College, London, U.K. in 2013 and 2014. He has intensive research experience in image processing and computer vision and, more specifically, object detection and recognition. His research interests include signal processing, computer vision, and machine learning.

**Yicheng Jiang** was born in Heilongjiang, China, in November 1964. He received the Ph.D. degree in information and communication engineering from the Harbin Institute of Technology (HIT), Harbin, China, in 1996. He is currently a Professor with the Research Institute of Electronic Engineering Technology, HIT. His research interest includes radar signal processing.

**Tania Stathaki** was born in Athens, Greece. She received the Masters degree in electronics and computer engineering from the Department of Electrical and Computer Engineering, National Technical University of Athens, Athens, Greece, and the Ph.D. degree in signal processing from Imperial College, London, U.K. She was a Lecturer with the Department of Information Systems and Computing, Brunel University, London, U.K., and an Assistant Professor with the Department of Technology Education and Digital Systems, University of Piraeus, Piraeus, Greece. She is currently a Reader with the Department of Electrical and Electronic Engineering, Imperial College. She has intensive research experience in image processing and computer vision and, more specifically, image fusion, image registration, change detection, object detection and recognition, and object tracking. She has mainly been involved in defense and security applications and has collaborated with the U.K. companies Dstl, General Dynamics, Selex Galileo, and BAE Systems. She is actively involved in various defense programs, such as the Data Information Fusion Defence Technology Centre, the Systems Engineering for Autonomous Systems Defence Technology Centre, and the University Defence Research Centre. She has authored or co-authored many papers on signal and image processing and computer vision.

**Huisheng Zhang** received the MS degree from Xiamen University in 2003 and Ph.D degree from Dalian University of Technology in 2009.He is currently an associate professor of Dalian Maritime University. His research interests include neural networks, signal processing, and learning theory. He has published several research papers in IEEE Transactions on Neural Networks, Neural Computation, Neurocomputing, Neural processing Letters, respectively.