

Estimation of view angles for gait using a robust regression method

De Zhang · Yunhong Wang · Zhaoxiang Zhang ·
Maodi Hu

Published online: 28 February 2012
© Springer Science+Business Media, LLC 2012

Abstract The performance of most gait recognition methods would drop down if the viewpoint of test data is different from the viewpoint of training data. In this paper, we present an idea of estimating the view angle of a test sample in advance so as to compare it with the corresponding training samples with the same or approximate viewpoint. In order to obtain reliable estimation results, the view-sensitive features should be extracted. We propose a novel and effective feature extraction method to characterize the silhouettes from different views. The discrimination power of this representation is also verified through experiments. Afterwards, the robust regression method is employed to estimate the viewpoint of gait. The view angles of test samples from BUAA-IRIP Gait Database are estimated with the regression models learned from CASIA Gait Database. Compared with the ground truth angles, such estimation is satisfactory with a small error level. Therefore, it can provide necessary help for gait application systems when the view angles of test data are uncertain. This point is verified experimentally through integrating the view angle estimation into a gait based gender classification system.

Keywords Gait · Robust regression · View estimation · Gender classification

D. Zhang · Y. Wang (✉) · Z. Zhang · M. Hu
State Key Laboratory of Virtual Reality Technology and Systems,
Beihang University, Beijing 100191, China
e-mail: yhwang@buaa.edu.cn

D. Zhang
e-mail: zhangde2000@163.com

Z. Zhang
e-mail: zxzhang@buaa.edu.cn

M. Hu
e-mail: londeehu@gmail.com

D. Zhang · Y. Wang · Z. Zhang · M. Hu
Laboratory of Intelligent Recognition and Image Processing, School of Computer Science
and Engineering, Beihang University, Beijing 100191, China

1 Introduction

Gait is defined as the manner of walking in Oxford Dictionary. People need to walk and their gait is usually apparent. As a relatively new biometric feature, gait offers several unique advantages such as non-contact, non-invasive and perceivable at a distance when compared with other traditional biometrics. Therefore, gait has been studied for automatic human identification and classification a lot in recent years [16]. The side view of gait receives considerable interest in most of earlier works [1, 3, 17, 21] and manifests superior performance. However, gait can be observed and captured from arbitrary viewing angles in visual surveillance, especially due to the rapid growth of security cameras. When the view angles between gallery data and probe data are different, the recognition rate drops to a rather low level [24]. Hence, it is necessary to automatically estimate the view angles presented by walking people in surveillance systems before recognizing.

How to reduce or avoid the effects of the view uncertainty on gait analysis and recognition has been investigated much by a great number of researchers. Since human gait is best reflected in the side view, which is justified by the fact that this view has proven its recognition capability in many gait recognition works, lots of the approaches found in literatures aim at synthesizing the side view of the human body from any other arbitrary views [2, 6, 10, 11, 18, 20]. In [11], Kale et al. estimate the walking angle in the 3D world according to a video sequence from a planar scene by using the perspective projection approach and optical flow based structure from motion (SfM) equations firstly. Then a camera calibration scheme is designed to transform an arbitrary view into the canonical view. In [18], Rogez et al. use several training views to create 2D point distribution model (PDM) and fit it to all the possible images captured by a single calibrated camera with perspective effect. In [10], Jean et al. propose a method which is based on homography transformations computed for each gait half-cycle. The body part trajectories from a walk at an arbitrary view are mapped to a simulated side view of the walk. In [2, 6], the markerless motion estimation is performed firstly to extract the limb's joint positions and trajectories as gait features. Then a viewpoint independent gait reconstruction algorithm is applied to normalize gait features extracted from an arbitrary view into the side-view plane. In [20], Shakhnarovich et al. compute an image based visual hull from a set of monocular views which is then used to render virtual views for tracking and recognition. Some other studies are interested in matching the view angle of test data without transformation to a canonical view [7, 13–15]. In [13–15], Makihara et al. apply a view transformation model (VTM) to frequency-domain features extracted from gait silhouette sequences. The features are transformed from one or more reference views in gallery data to match gait features of different walking directions in probe data. In [7], Han et al. utilize the fact that the walking view ranges along different directions might be overlapping to carry out the view-insensitive gait recognition.

However, these aforementioned approaches suffer from some limitations in surveillance scenarios. Camera calibration is needed in [11, 18] but it is not available in real surveillance environment. Moreover, the methods proposed in [7, 11] doesn't work well when view difference is large. The trajectories transformation method in [10] requires an assumption that the velocity of the observed walking is constant. The visual hull based method in [20] needs images taken synchronously from multiple view directions for all subjects. The VTM proposed in [13–15] does not consider the case in which the view direction of a probe set is different from any view direction in the gallery set. The viewpoint reconstruction method in [2, 6] needs a straight walk and constant distances between the bone joints both of which are not usually possible in surveillance applications.

In this paper, we propose a novel way to handle the problem of view uncertainty. It is that we implement view angle estimation prior to the recognition or classification stage. This

approach is capable of overcoming the restrictions mentioned above in real application. The proposed view estimation doesn't need the camera calibration or synchronizing image sequences, etc. Here the specific requirement is a large database consisting of different view angles, which is easier to be obtained. We intend to estimate the viewpoint of the test gait sequence using a regression model which is learned on the large training data set including multiple views. Because the smaller view differences between test data and training data could result in better recognition rates, we aim to find the most accurate view estimation of the test gait sequence through a robust regression method. Then the recognition process is carried out by using training data with the most approximate view angle. The publicly available gait database which has the most number of viewing angles is CASIA Gait Dataset B [24]. It provides 11 view directions at every 18 degrees from 0° to 180° . The view estimation using robust regression is experimented on this database firstly. For the sake of practicability, next we take another database as the testing set to estimate the view angles which are different from those view angles in the training set. Our BUAA-IRIP Gait Database including 7 views at every 30 degrees from 0° to 180° is used here. Moreover, a gait based gender classification system with prior view angle estimation is designed and realized to provide a verification. To improve the accuracy of estimation, we also propose a view-sensitive feature extraction method to represent gait silhouette.

The contribution of this paper is two-fold. Firstly, we put forward the idea of view angle estimation using robust regression to explore the problem of view uncertainty in gait applications. In surveillance scenarios, the walking directions of people are unexpected and the gait recognition system does not work well when the viewpoints of test data are not known. So prior view estimation is necessary and helpful. Secondly, we propose a gait representation method which is view-sensitive and low cost of computation. This representation reflects the view difference through a compact description of the shape and appearance of the silhouette.

The rest of this paper is organized as follows. Section 2 presents the framework of view angle estimation. Section 3 introduces how to extract view-sensitive features in detail and describes the validation of view sensitivity. Section 4 discusses the robust regression method which is adopted as the estimation tool. In Section 5, experimental evaluations of the proposed view angle estimation are presented. Section 6 concludes the paper and provides the future work.

2 View angle estimation framework

The proposed view angle estimation framework mainly consists of five parts, which are silhouette images extraction, silhouettes normalization, view-sensitive feature extraction, regression model generation and view angle computation as shown in the diagram of Fig. 1. For training, binary silhouette images are extracted from original gait videos using background subtraction technique. Then the silhouettes undergo a normalization which includes scaling the foreground regions to the same height while keeping the ratio of their height to width and moving them to the center of silhouette images. Next, a new gait representation is taken into account to extract the view-sensitive features. These features are extracted from silhouette images directly. Then a robust regression function is applied to fit the feature data and generate the final regression model. For testing, the gait video sequences are pre-processed in the same way as for the training data. Then view-sensitive features are extracted from normalized silhouettes. Afterwards, we can estimate the view angles of test samples by feeding these features into the trained regression model.

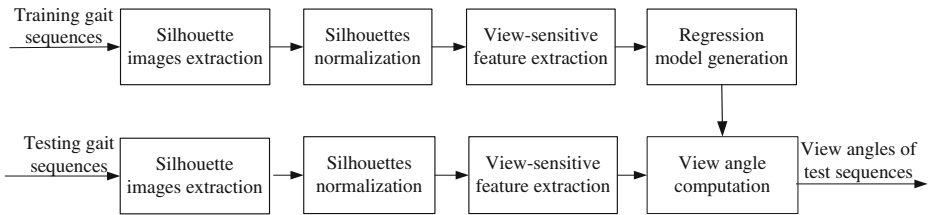


Fig. 1 Gait view angle estimation framework

3 View-sensitive feature extraction and verification

Feature extraction is an important part in the view estimation framework. Previous studies about gait analysis focus on feature extraction from the side view or try to transform features from other views into the side view. We need such features that are easy to be acquired from any viewing angle and are embedded with the difference of viewing angles. In this section, we will introduce the proposed feature extraction method and the verification of its view sensitivity.

3.1 Feature extraction

The features used in this paper are extracted from gait silhouettes. For video sequences, the preprocessing contains two stages. Firstly, the original walking videos are dealt with the technique of background subtraction. For BUAA-IRIP Gait Database, we use the mean value of multiple frames to update the background and threshold the difference images between frames and the background to generate silhouette images. For CASIA Gait Dataset B, we use the silhouette images in the download package directly. Secondly, the silhouette images need to be normalized. The foreground region (body region) in silhouette images is scaled to the same height of 140 pixels while keeping the aspect ratio of the region. Afterwards, we compute the center of the foreground region and move it to the center of its corresponding image which has a uniform size of 155-by-100. Thus, all the silhouette images are normalized to the same size and centered with foreground regions of the same height. Figure 2 illustrates some original frames and the preprocessed silhouette images of different viewpoints from BUAA-IRIP Gait Database. Here the view angle is defined as the angle formed by an optical axis and a walking direction.

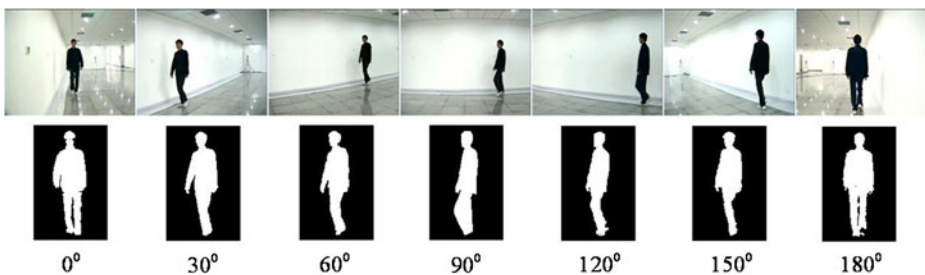


Fig. 2 The first row shows some frame samples. The second row shows the corresponding silhouettes. The third row lists the degrees of viewing angles

From these silhouettes shown in Fig. 2, it can be seen that they look apparently different because of view variation. We would like our gait representation to have the ability to represent the differences among different view angles in an effective but simple way. Considering the appearance level, we find that the whole body shape difference is the most salient because the stances of a walker captured at the same moment by different cameras which are located at different viewing directions differ from each other a lot. So we propose a gait representation method utilizing the shape and contour information. Note that the silhouette of 0° view looks quite similar with the silhouette of 180° view, as shown in Fig. 2. Hence, we don't make a distinction between these two views in this paper.

For a silhouette from a gait video sequence, we calculate the centroid and set the vertical line through the centroid as the central axis. On each row of the silhouette, there are two border points. One is on the left end and the other is on the right end. A triangle can be formed by these two end points and the centroid. The angle located on the vertex of centroid and the straight distances from the two end points to central axis are combined into a three-dimensional vector (θ_i, dl_i, dr_i) , where i denotes the i_{th} row. This vector is used as the feature expression of the i_{th} row. The top portion including head and neck is not sensitive to camera view changes since either head or neck in this portion is nearly round. Therefore, we only use the portion below neck to extract view-sensitive features. Figure 3 gives some examples of this representation on the silhouettes of 0° , 30° , 60° , 90° , 120° , and 150° view. In Fig. 3, there is a vertical central line and a triangle in each silhouette used to illustrate how to render the row vector (θ_i, dl_i, dr_i) . The triangle is shown in dotted line. Its lower vertex is the centroid and the angle bound to the centroid is denoted by θ_i . The opposite side of θ_i is the i_{th} row and it is split into two parts by the vertical central line. The length of the left part is dl_i and the length of the right part is dr_i . Moreover, it is observed from Fig. 3 that the triangles look apparently different when viewing angles change. According to this observation, we can say that this gait representation is view-sensitive. For each row of the silhouette, a triangle like the one shown in Fig. 3 is formed and three parameters of each triangle are used as view-sensitive features.

For a body height H , an initial estimate of the vertical position of the lower border of the neck (just above the shoulder) is $0.818H$ [4]. The H in a normalized silhouette image is 140 pixels according to the preprocessing. We select the portion starting from the bottom of feet and ending around the shoulder and compute the three-dimensional vector of each row in this portion as described above. The number of rows in this portion is set as 110 which is an approximate value of 140×0.818 . Since the feature vector corresponding to the i_{th} row is

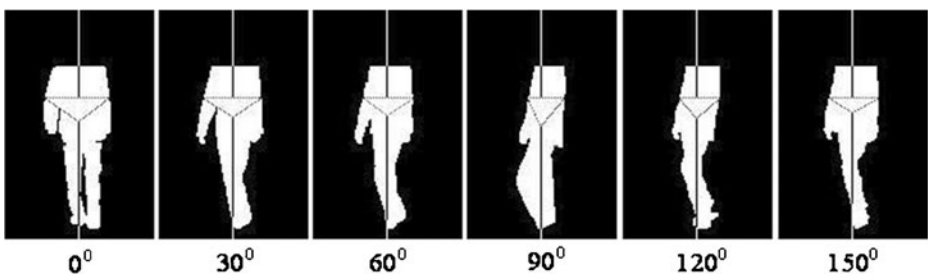


Fig. 3 Examples of view-sensitive feature extraction for different views. The head and neck portions are removed. Center axis and the triangles are marked

denoted by (θ_i, dl_i, dr_i) , the extracted feature of the j th silhouette image in a gait sequence can be expressed by concatenating these row vectors:

$$F_j = (\theta_{j,1} dl_{j,1} dr_{j,1} \cdots \theta_{j,i} dl_{j,i} dr_{j,i} \cdots \theta_{j,110} dl_{j,110} dr_{j,110}) \tag{1}$$

where F_j is a $110 \times 3 = 330$ dimensional vector. These features are robust to noises produced from background subtraction because we mainly use the pixels on the outer boundary of the silhouettes to extract features. The more important is that these features reflect the differences caused by varying view angles at the appearance level.

Given the feature vectors from a gait sequence, we need a concise representation over time. Our method is to compute the average feature vector of a gait cycle. Thus we don't need to consider whether the starting frames of the gait sequences with different views appear at the same time or not. Gait period is detected using the approach described in [25]. Let N denote the number of frames within a gait period. N is the same for different view angle as long as the walking is exposed to these view angles simultaneously. So, the gait average appearance feature vector \bar{F} of a gait cycle is calculated as follows:

$$\bar{F} = \frac{1}{N} \sum_{j=1}^N F_j = (\bar{\theta}_1 \bar{dl}_1 \bar{dr}_1 \cdots \bar{\theta}_i \bar{dl}_i \bar{dr}_i \cdots \bar{\theta}_{110} \bar{dl}_{110} \bar{dr}_{110}) \tag{2}$$

where \bar{F} is a $110 \times 3 = 330$ dimensional vector as F_j .

3.2 Feature verification

The view sensitivity of the proposed gait representation needs to be verified before we use it in the regression experiments. Each view is treated as a single class and the class separable criterion can be used to measure the separability between every pair of classes. In this way, we are able to know how sensitive the extracted features are to view changes. The Bhattacharyya distance is a kind of measurement of class separability and is known to provide the upper and lower bounds of the Bayes error [5]. Especially, it has a monotonic relationship with the upper bound of the error probability. We adopt Bhattacharyya distance to measure the separability of different views. Although Bhattacharyya distance is used here, the proposed gait representation is not limited to this particular measurement and other methods are also applicable. For two classes of normal distribution, the Bhattacharyya distance is defined as follows:

$$J_B = \frac{1}{8} (\mu_2 - \mu_1)^T \left[\frac{\Sigma_1 + \Sigma_2}{2} \right] (\mu_2 - \mu_1) + \frac{1}{2} \ln \frac{|(\Sigma_1 + \Sigma_2)/2|}{|\Sigma_1|^{1/2} |\Sigma_2|^{1/2}} \tag{3}$$

where μ_i and Σ_i are the mean vector and covariance matrix of class i , respectively.

We use BUAA-IRIP Gait Database to evaluate the view discriminability of the extracted features. An assumption is needed before calculating the Bhattacharyya distance that the feature data of each view are normally distributed. Such assumption is theoretically reasonable for most sample data even without the support of the Central Limit theorem. The view angles available in this database are $0^\circ, 30^\circ, 60^\circ, 90^\circ, 120^\circ, 150^\circ$, and 180° . We don't use the view of 180° as aforementioned in this paper. For each gait video sequence in this database, we extract a gait cycle and compute the average feature vector \bar{F} using the method described earlier. There are 430 gait sequences for every individual view in this database. The feature vector \bar{F} is 330 dimensional. So the sample data size of a single view is 430-by-330. The

Eq. 3 is used to calculate the Bhattacharyya distances between any two of these views and the results are listed in Table 1. It is clearly seen that these distances are large enough to manifest the differences existing in different views. This proves that the proposed feature extraction method is view-sensitive.

We also do experiments of view classification based on the proposed feature representation to further verify the view sensitivity. A high correct classification rate accounts for a high view sensitivity of the features. As shown in Table 1, there are six different view angles. The number of all the combinations of two different views out of six view angles is:

$$\binom{6}{2} = \frac{6!}{2!(6-2)!} = 15 \quad (4)$$

In every combination, the sample data is composed of two classes which are actually two kinds of viewing angles in this paper. As mentioned above, the sample size for one view angle is 430-by-330 where each row is an observation, and each column is a feature element. We perform the classification experiment for every combination of two view angles. Linear discriminant analysis (LDA) is used to find an optimal combination of feature elements to separate two classes. In BUAA-IRIP Gait Database, there are 86 subjects in total and every one has 5 gait video sequences for each single view. We use the leave-one-out cross-validation method. Therefore, the classifier is trained 86 times and during each time the 5 gait sequences from one person are held out as a validation set. The mean value of these 86 classification results is used as the final estimated performance. The correct classification rates cross different views are listed in Table 2. It is shown that the classification results based on the proposed gait representation are rather satisfactory. Any pair of different view angles can be distinguished with a high accuracy. Therefore, in addition to the Bhattacharyya distances shown in Table 1, these experimental results listed in Table 2 make the view sensitivity of our proposed feature extraction method more believable and reliable to be utilized for view angle estimation.

4 Robust regression method

In this section, we describe how view angles are estimated using a regression method, given a set of training data with the above extracted features.

4.1 Multiple linear regression

After finding the compact representation of the original gait video sequences, we define the view angle estimation as a multiple linear regression problem [22] in a low-dimensional

Table 1 Bhattacharyya distances between different view angles

View angle	0°	30°	60°	90°	120°	150°
0°	0	17.51	28.46	35.86	23.15	14.75
30°	17.51	0	15.71	21.85	13.02	15.01
60°	28.46	15.71	0	13.29	14.41	29.44
90°	35.86	21.85	13.29	0	14.78	30.25
120°	23.15	13.02	14.41	14.78	0	12.57
150°	14.75	15.01	29.44	30.25	12.57	0

Table 2 Correct classification rate between different view angles

View angle	0°	30°	60°	90°	120°	150°
0°	N/A	95.35%	99.07%	99.77%	98.95%	94.77%
30°	95.35%	N/A	98.14%	99.81%	98.95%	94.79%
60°	99.07%	98.14%	N/A	99.19%	94.65%	98.49%
90°	99.77%	99.81%	99.19%	N/A	99.19%	99.31%
120°	98.95%	98.95%	94.65%	99.19%	N/A	97.79%
150°	94.77%	94.79%	98.49%	99.31%	97.79%	N/A

space. Linear regression is the oldest and most widely used predictive model. It attempts to model the relationship between two variables by fitting a linear equation to observed data. One variable is considered to be a predictor variable, and the other is considered to be a response variable. When there are several predictor variables or it can be said that the predictor is multidimensional, this case is referred to as multiple linear regression.

For the purpose of this paper, Eq. 5 specifically describes the regression model:

$$view\ angle = f(v) \Leftrightarrow \widehat{L} = \widehat{f}(v) \tag{5}$$

where \widehat{L} denotes the estimated view angle label, $f(\cdot)$ is the unknown regression function, $\widehat{f}(\cdot)$ is the estimated regression function and v is the feature space. Given n independent observations $(x_1, l_1), \dots, (x_n, l_n)$ of the feature vector x and the view angle label l , the linear regression model becomes an n -by- p system of equations:

$$\underbrace{\begin{pmatrix} l_1 \\ \vdots \\ l_n \end{pmatrix}}_L = \underbrace{\begin{pmatrix} m_1(x_1) & \cdots & m_p(x_1) \\ \vdots & \ddots & \vdots \\ m_1(x_n) & \cdots & m_p(x_n) \end{pmatrix}}_X \underbrace{\begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}}_B + \underbrace{\begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}}_e \tag{6}$$

where L is the view angle label vector, X is the predictor matrix including designed model terms, B denotes the unknown coefficient vector which we need to estimate during the learning stage, and e is the error vector consisting of unobservable random variables. To fit the model to the data, B is estimated by ordinary least squares:

$$\widehat{B} = (X^T X)^{-1} X^T L \tag{7}$$

where T is the matrix transpose operator. Then, the fitted value of \widehat{L} is computed by:

$$\widehat{L} = X \widehat{B} = X (X^T X)^{-1} X^T L \tag{8}$$

The residual values in $L - \widehat{L}$ are useful for detecting failures in model assumptions.

4.2 Robust regression

All estimation methods rely on assumptions for their validity. The widely used least-squares solution mentioned above can lead to unreliable results if the assumptions are not true. So it is not robust to violations of its assumptions. An estimator is said to be robust if it provides useful information even when some of the assumptions used to justify the estimation method

are not available. In this paper, we adopt robust regression method to solve the problem of view angle estimation.

Robust regression uses robust fitting methods that are less sensitive than ordinary least squares to large changes in small parts of the data. The estimators of robust regression are designed to be not overly affected by violations of assumptions by the underlying data-generating process [9]. Due to their superior performance over least squares estimation in many situations, we use robust estimation methods to train the regression models in this work.

M-estimation proposed in [8] can be used for robust regression. The M in M-estimation stands for maximum likelihood type. This method is robust to outliers in the responsible variable. However, it turns out not to be resistant to outliers in the explanatory variables. Another solution is S-estimation proposed in [19]. This method computes a hyperplane that minimizes a robust estimate of the scale (from which the method gets the S in its name) of the residuals. This method is highly resistant to explanatory variables, and is robust to outliers in the response. However, it turns out not to be efficient.

The later proposed MM-estimation [23] attempts to retain not only the robustness and resistance of S-estimation but also the efficiency of M-estimation. Hence, it is our choice for the robust estimation method in this paper. This method is realized by finding a highly robust and resistant S-estimate which minimizes an M-estimate of the scale of the residuals (the first M in the method's name), and the estimated scale is then kept constant while a close-by M-estimate of the parameters is located (the second M).

After the predictor matrix X in Eq. 6 is generated by using a predefined formulation, we implement the MM-estimation to calculate the coefficient values in B . As suggested by regression related work in [12], we have investigated three formulations for the regression function: a linear, a pure quadratic, and a pure cubic formulation. Given a feature vector $x = (x_1, \dots, x_n)$, we have these three formulations shown in Eq. 9, Eq. 10, and Eq. 11, respectively.

$$\varphi = x_1 + \dots + x_k \quad (9)$$

$$\varphi = x_1 + \dots + x_k + x_1^2 + \dots + x_k^2 \quad (10)$$

$$\varphi = x_1 + \dots + x_k + x_1^2 + \dots + x_k^2 + x_1^3 + \dots + x_k^3 \quad (11)$$

5 Experiments and analysis

5.1 Databases

For the task of view angle estimation, obviously we need databases including multiple view angles. CASIA Gait Dataset B [24] is a publicly available database for gait analysis with 11 different view angles ranging from 0° to 180° at an interval of 18° . Figure 4 shows the cameras setup for data collection and how the view angle is defined in this database. The angle θ formed by the optical axis and the walking direction is specified as the view angle of the corresponding camera. In CASIA Gait Dataset B, the camera 1 faced by the walking person is assigned with the angle label of 0° . With an interval of 18° between two adjacent

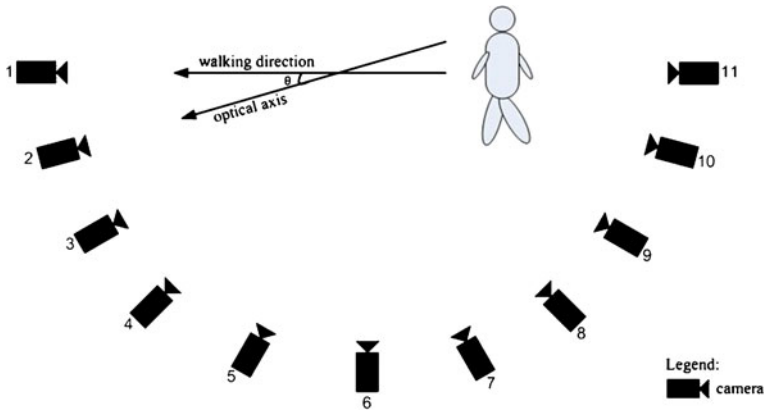


Fig. 4 Cameras setup and view angle definition

cameras, the camera 2 has the label of 18° . Likewise, the camera 3 is assigned with the label of 36° , and till camera 11, the angle label is right 180° . The walking direction in this database is from right to left, as illustrated in Fig. 4.

We also have our own database which is named as BUAA-IRIP Gait Database. Although it is built for gender classification initially, there are 7 different views in this database such that it is also appropriate for the purpose of view angle estimation. The cameras used for gait data collection are placed in the same way as shown in Fig. 4, forming a semicircle. But the interval between two adjacent cameras is 30° and so the number of different view angles is 7. The view angle label of 0° is also assigned to the camera which the person walks toward. The next neighbor one is assigned with 30° , and till the last one right with the label 180° .

In Fig. 2, some examples of original images in the BUAA-IRIP Gait Database can be found to show the capturing effect of different view angles. As mentioned in Section 3, we don't make a distinction between 0° and 180° . So we don't use the gait sequences with the view angle label of 180° in this paper. In CASIA Gait Dataset B, there are 124 subjects in total. For each viewpoint, there are 6 normal walking video sequences per subject. In BUAA-IRIP Gait Database, there are 86 subjects each of which has 5 normal walking video sequences for each view angle.

The concerned difference between these two databases is the scale of interval angle. One is 18° and the other is 30° . As mentioned in the beginning, the view direction to be estimated might appear in any angle. A smaller interval angle and a larger number of different views in the training database can make the estimation more precise. So, in our experiments, we use CASIA Gait Dataset B as training data to learn the parameters of a regression model.

5.2 Experimental evaluation

We have two sets of experiments carried out in this part. Firstly, we want to find out how well the view angle estimation could perform, especially by means of the robust regression method mentioned in Section 4. CASIA Gait Dataset B is used to explore the performance in the first set of experiments and it is separated into a training part and a testing part. Secondly, it is necessary to consider the case like a real environment in which the view angle of test data is different from any angle available in training data. So, in the second set of experiments, we use BUAA-IRIP Gait Database as the test data and CASIA Gait Dataset B is still employed for training.

In both sets of experiments, we use the method described in Section 3 to extract features. For each gait sequence, we extract a gait cycle and calculate the feature vector \bar{F} in Eq. 2. Every \bar{F} in the training data and its corresponding view angle are the input to learn the regression model. \bar{F} is composed of 110 row vectors as shown in Eq. 2 and the data redundancy does exist in \bar{F} because of the similarity of neighboring row vectors. Such similarity might appear on some contour regions where the shape changes slowly, in terms of the proposed gait representation. So, when concatenating the row vectors to generate \bar{F} , we consider selecting rows by an interval to reduce the data redundancy. We test different intervals to find out the most appropriate one. The interval starts from 1 row, which means all rows are chosen and the feature vector \bar{F} is $110 \times 3 = 330$ dimensional. Then we use 2 rows, 3 rows, until 40 rows as the interval one by one. In this way we obtain feature vectors with different dimensions and they are all tested through experiments to compare their performance. For example, when the interval is 3 rows, the selected result will include 37 rows which are the 1st row, the 4th row, the 7th row, and so on, until the 109th row. The ending row number is the maximum value which is less than 110 when stepping up with the interval. The corresponding feature vector \bar{F} is $37 \times 3 = 101$ dimensional and a regression result will be generated under this condition. The maximal interval is 40 rows here because the dimension of feature vector is already decreased to $3 \times 3 = 9$ for this interval and the information of gait representation cannot be lost any more.

5.2.1 Experiment Set 1

We separate the CASIA Gait Dataset B into two subsets, 84 persons in training set and the remaining 40 persons in testing set. Every subject has 6 gait video sequences for every view angle. The degrees of the 11 viewing angles in this database are 0° , 18° , 36° , 54° , 72° , 90° , 108° , 126° , 144° , 162° , and 180° . As mentioned before, we don't use the view of 180° . So the total number of gait sequences in training set is $84 \times 6 \times 10 = 5040$. The dimension of feature vector \bar{F} varies from 330 to 9 according to different interval rows. We implement the robust regression method on the training data to estimate the parameter vector B in Eq. 6. Then B is used to calculate the estimation of view angles of test data. We use the MM estimator mentioned in Section 4 to carry out the robust regression. As described in [23], MM-estimation is defined in three steps. In the first step, the least trimmed squares (LTS) estimate is computed as the initial estimate because of its speed and high breakdown value. In the second step, the scale parameter is computed using Tukey's bisquare function. In the third step, the iteratively reweighted least squares (IRLS) algorithm is used to compute the final MM estimate. In addition, we consider using three formulations shown in Eq. 9–11 to evaluate the performance of view angle estimation.

The experimental results are illustrated and compared in Fig. 5. The absolute error between estimated view angle and the ground truth in the testing set is introduced to give an analysis of the results. Figure 5(a) shows the Mean Absolute Error (MAE) of the absolute error against the different row interval. The row interval implies the dimension of the feature vector. Figure 5(b) shows the variance of the absolute error against the different row interval. It is observed that all the three formulations achieve the best results when the number of interval rows is around 4. Comparatively speaking, the pure cubic formulation outperforms the linear and pure quadratic formulation. From both the evaluations shown in Fig. 5, we can see that the pure cubic formulation performs the best at the interval of 4 rows where the MAE is around 7.5 degrees and the variance of absolute error is less than 30. Such a result is rather acceptable in terms of the minimal view difference of 18 degrees in the database.

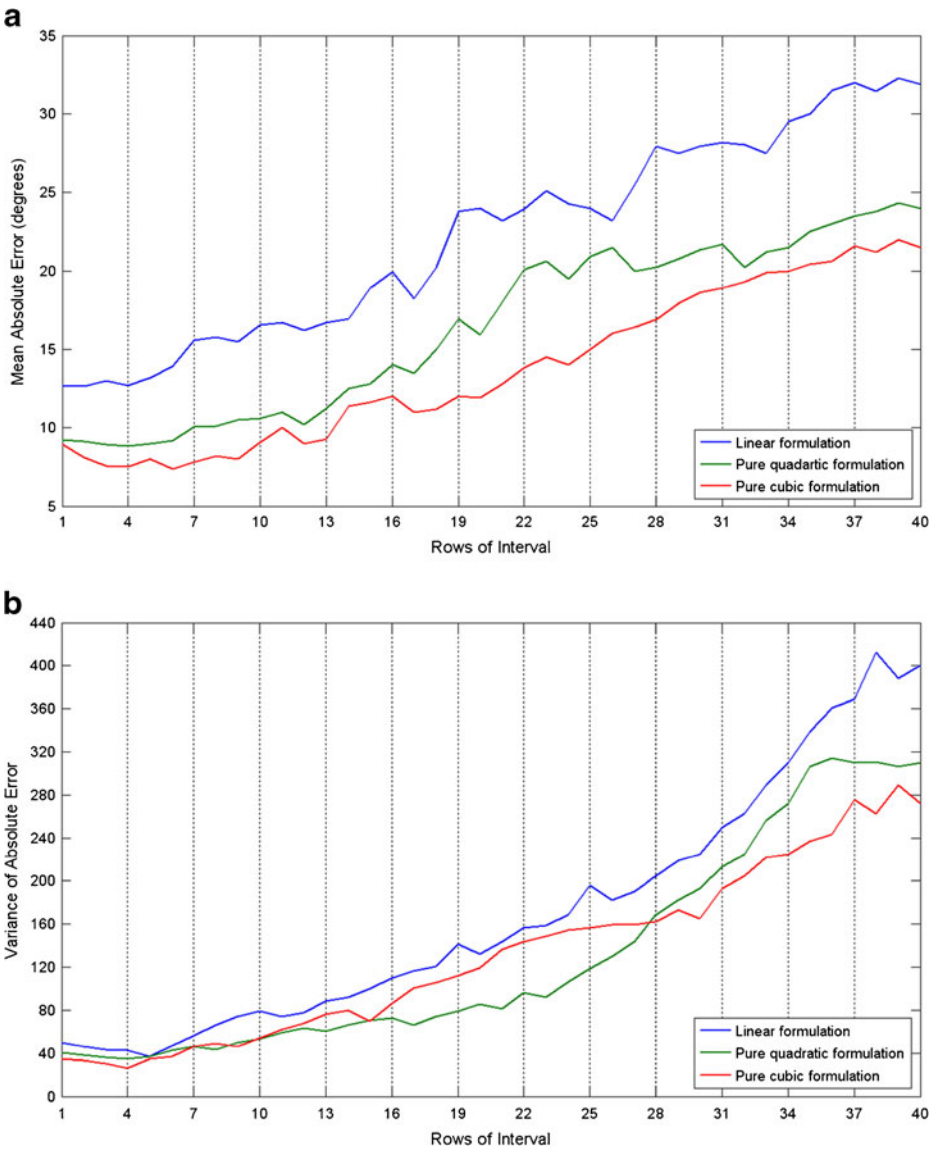


Fig. 5 Absolute error analysis of robust regression with different formulations on CASIA Gait Dataset B. **a** MAE, **(b)** Variance of absolute error

Therefore, we illustrate the accuracy scores against error level of various degrees in Fig. 6 specifically for the interval of 4 rows. Apparently, the pure cubic formulation gives the superior results than the other two. So in the later experiments, we choose the pure cubic formulation expressed by Eq. 14 as the regression model function.

In order to demonstrate the effectiveness of the robust regression method, we duplicate the aforementioned experiments but using the multiple linear regression to make a comparison. The results are shown in Figs. 7 and 8. Note that only pure cubic formulation is taken into account. Figure 7(a) shows the MAE of the absolute error versus the interval rows.

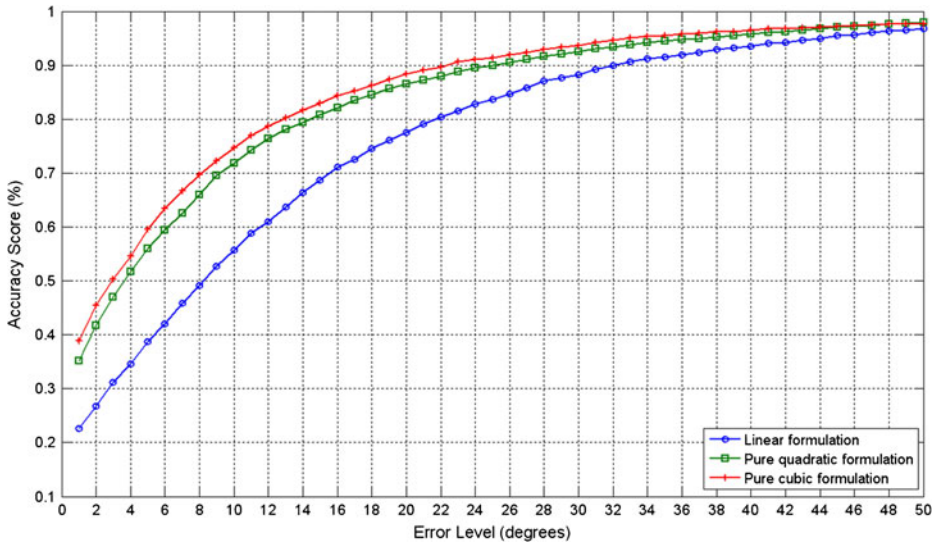


Fig. 6 Accuracy scores of robust regression with different formulations on CASIA Gait Dataset B

Figure 7(b) shows the variance of the absolute error against the interval rows. Figure 8 presents the accuracy rates versus the estimation error levels when the interval is 4 rows. As shown in Fig. 7, robust regression has a lower MAE and variance than linear regression. In Fig. 8, when the error level is less than 28 degrees, robust regression has a higher accuracy score than linear regression. We are interested in the accuracy rate at the error level of 18 degrees which is equal to the view angle interval in CASIA Gait Dataset B. For robust regression, the corresponding value is 86.3% as shown in Fig. 8. It is fairly good to show the feasibility of estimating view angles with robust regression method.

Additionally, it can be observed from Figs. 5 and 7 that the overall estimation results look better when the value of interval rows is smaller. As mentioned before, the final feature vector is composed of the row vectors selected with an interval. From Figs. 5 and 7, we see that the results become unstable and unacceptable along with the expanding of interval rows. A small interval could reduce some data redundancy, but a big interval could lead to losing too much feature information. When the interval is less than 10 rows, the results are relatively satisfactory according to Figs. 5 and 7. Furthermore, it is observed that the interval of 4 rows gives the best result for the analysis of absolute error. The accuracy score is plotted against different error levels in Figs. 6 and 8 corresponding to this interval.

5.2.2 Experiment Set 2

In this set of experiments, we utilize two data sets in which there are no identical viewing angles to test the performance of robust regression method used in Experiment Set 1. All the 124 subjects in CASIA Gait Dataset B are used as the training set. The viewing angles in training set are 0° , 18° , 36° , 54° , 72° , 90° , 108° , 126° , 144° , and 162° . All the 86 subjects in BUAA-IRIP Gait Database are used as the testing set. There are seven different views in this database and four of them are different from the view angles available in CASIA Gait Dataset B. These four view angles, which are 30° , 60° , 120° , and 150° respectively, are used in testing set. Like Experiment Set 1, we implement MM-estimation method to compute the

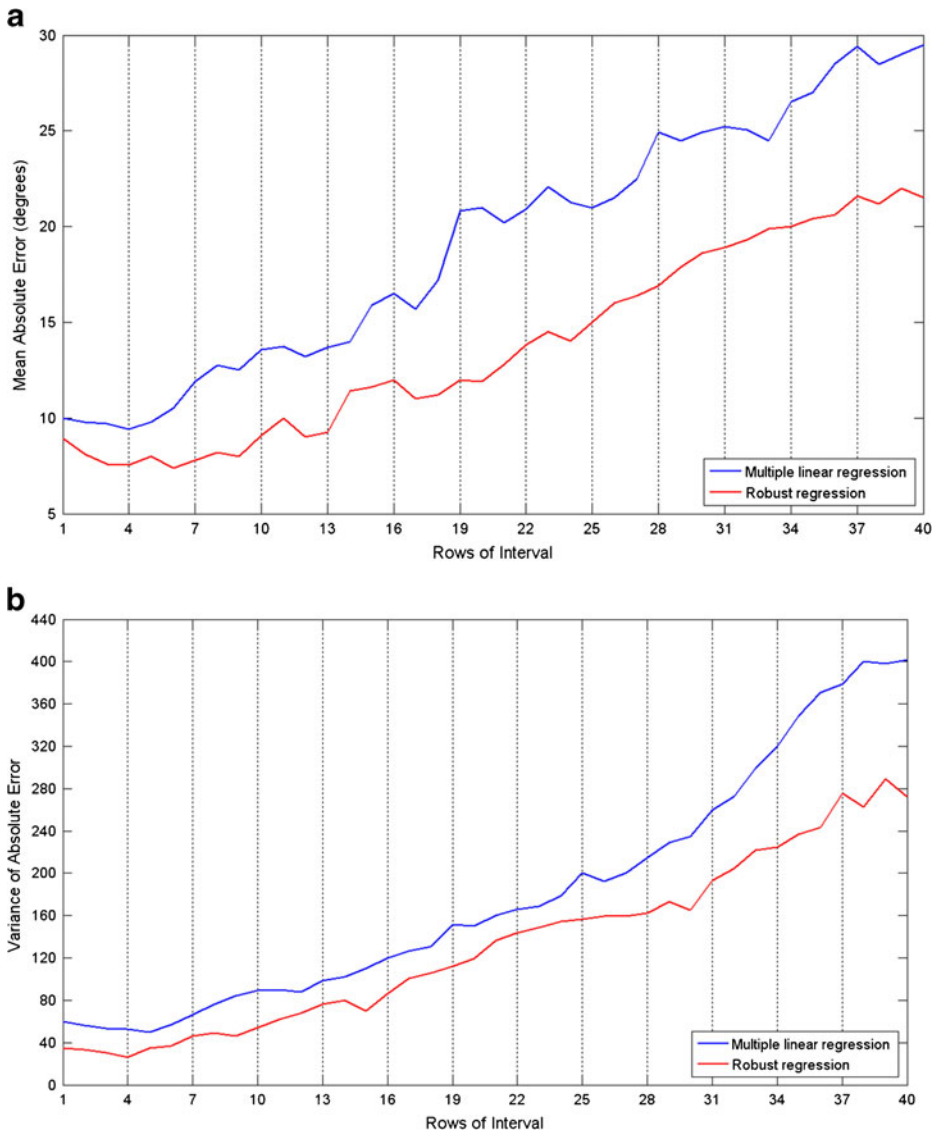


Fig. 7 Comparison of absolute error analysis between robust regression and multiple linear regression using pure cubic formulation. **a** MAE, **(b)** Variance of absolute error

parameter vector B and employ pure cubic formulation as the regression mode function. The MAE and variance of absolute error are shown in Fig. 9. For the interval of 4 rows, the accuracy score versus error level is shown in Fig. 10. For the convenience of comparison, we also plot the results obtained with robust regression and pure cubic formulation in Experiment Set 1 in Figs. 9 and 10.

The results obtained from this cross-database validation are better than the results of Experiment Set 1 as shown in Figs. 9 and 10. Under the same row interval, it can be seen from Figs. 9 and 10 that the MAE and variance of absolute error for the cross-database

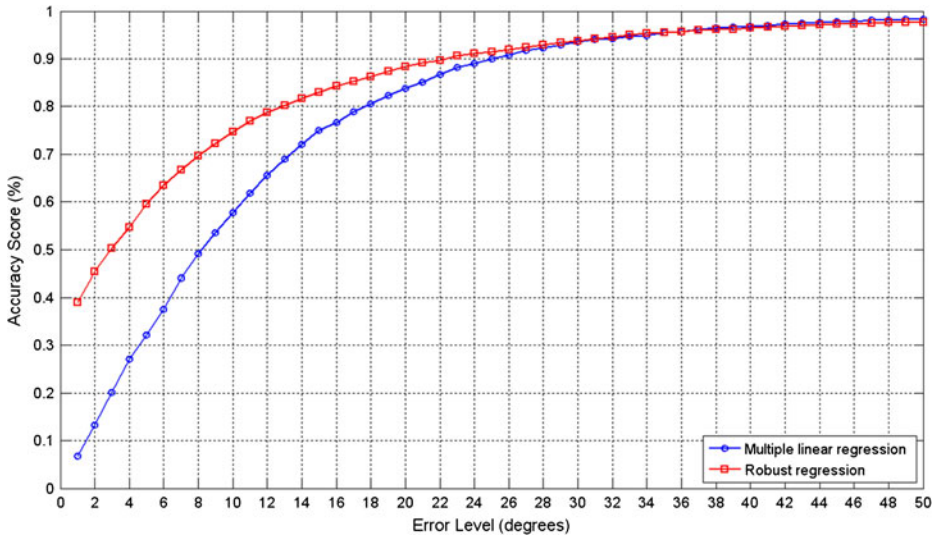


Fig. 8 Comparison of accuracy scores between robust regression and multiple linear regression using pure cubic formulation

validation are smaller and the accuracy scores for the cross-database validation are higher. Moreover, in Fig. 10 we can observe that the accuracy rate reaches 83.3% at the error level of 10 degrees for cross-database validation. That is to say, given a training data set with a view interval of 18° and a test sample with an unknown view angle, we could estimate the view angle of the test sample within a deviation range of 10° in an accuracy rate of 83.3%. Since Experiment Set 2 is designed more closely to the real application than Experiment Set 1, such results are very encouraging and significant for solving the problem of view angle variation. We also know that every method has its limitations and the proposed view estimation is not always sufficient. In cases such as walking on a curved path or tilted road, the performance of view estimation could be badly impacted. Under such situations, body tilt correction or silhouette image registration might be additionally required.

5.3 An application of view estimation

In this part, we integrate the view estimation into an application system to validate if it reduces the effect of view angle uncertainty. Using both CASIA Gait Dataset B and BUAA-IRIP Gait Database, we design a gender classification system. Like Experiment Set 2, we use CASIA database as the training set because it has the most number of view angles. The data sequences with the angles of 30° , 60° , 120° , and 150° in BUAA-IRIP database are taken as the walking sequences simulated from the real environment. Assume that the view angles of these sequences are unknown because of the unexpected walking directions towards the controlling cameras. Considering an application of gender recognition, we firstly estimate the view angles of these walking sequences and secondly implement gender classification by matching the data sequences with approximate view angles in the training set.

The view angle estimation described in Experiment Set 2 is carried out in the first step. The estimation results are transmitted into the second step. Given a test sequence, assume that its view angle is estimated to be $\hat{\theta}$ in the first step. Then we find the data sequences with the closest angle to $\hat{\theta}$ from the training set and use them as the corresponding training data

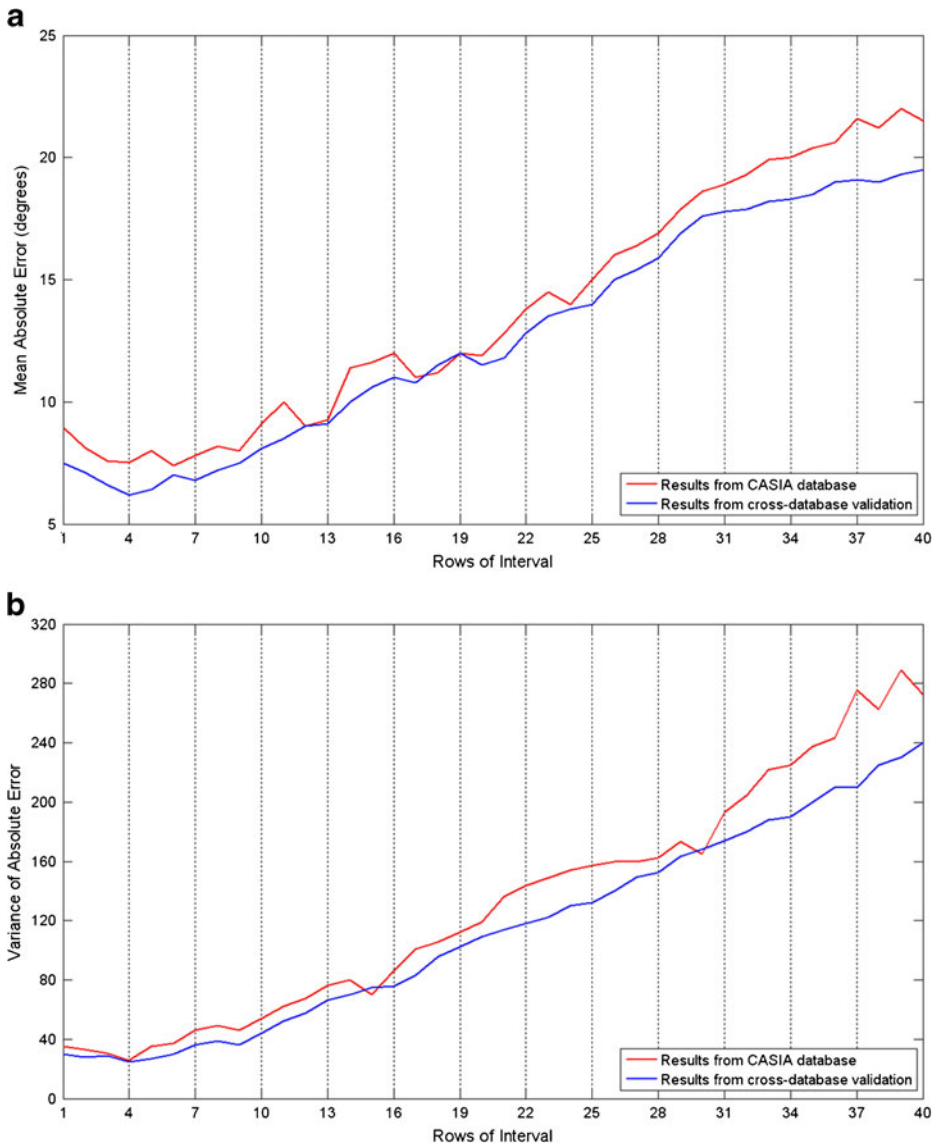


Fig. 9 Comparison of absolute error analysis for robust regression with pure cubic formulation. **a** MAE, **(b)** Variance of absolute error

for the given test sequence. For each view angle in CASIA database, there are 124 subjects including 31 females and 93 males. We compute two average Euclidean distances to determine the gender of the test sample. One is the average of distances between the test sample and all female samples in training data. The other average distance is between the test sample and all male samples in training data. The smaller one of these two average distances indicates the gender type that the test sample more likely belongs to. Note that the gait feature used in the second step is from GEI representation. We don't use the view-sensitive feature extraction method described in Section 3 because it is mainly proposed to

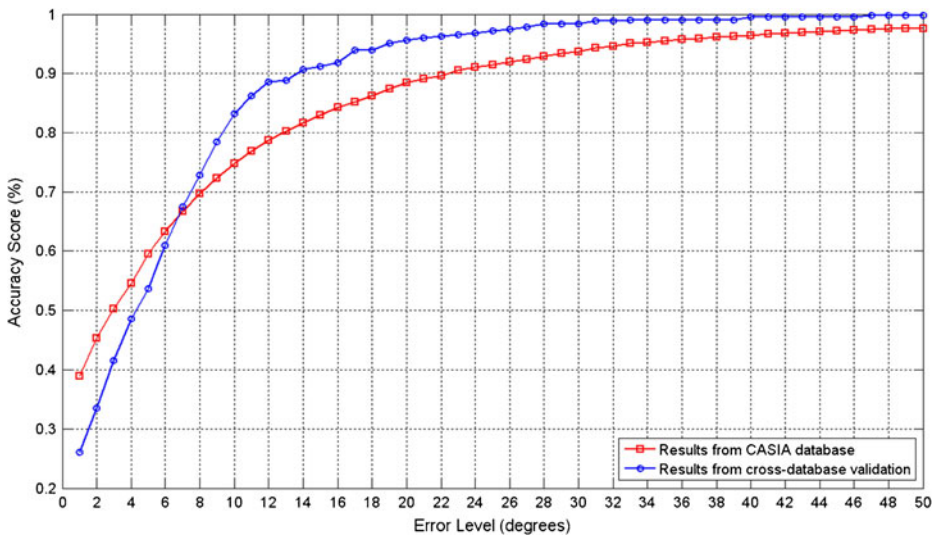


Fig. 10 Comparison of accuracy scores for robust regression with pure cubic formulation

differentiate the view angles and is not appropriate for gender classification. Table 3 lists the classification results for the four view angles mentioned above. One part of the results is the correct classification rate based on prior view estimation. The other is the correct classification rate based on ground truth training data. For the four view angles of 30° , 60° , 120° , and 150° in the testing data, the closest views in training data are 36° , 54° , 126° , and 144° , respectively. We take the data sequences from these views as the corresponding ground truth training data when carrying out gender classification. The classification results obtained with view estimation are not much worse than the results obtained by using the ground truth training data. This shows that the proposed view angle estimation is applicable and reliable for practical applications.

We also carry out some experiments using other solutions for view changes to compare with the results of the proposed method. The first experiment is to train a view transformation model as proposed in [13–15]. The second experiment is to render view-invariant features by using a view-point rectification method proposed in [2, 6]. The experimental results are listed and compared in Table 4. Overall, the proposed view estimation results in a better performance than other methods. For the view of 60° and 120° , the VTM method achieves competitive results which are even slightly better than the proposed estimation. But the view-invariant features don't perform well for all these four views. According to the experimental results shown in Tables 3 and 4, the proposed view estimation provides an effective way to handle the view uncertainty.

6 Discussion

According to the experiments, robust regression can provide satisfactory results when the row interval is relatively small during feature combination. The tradeoff between data redundancy and data loss is what we need to face. Based on the experimental results, we find that the interval of 4 rows is an appropriate choice. For the proposed gait representation, it could be a complement for extracting features more effectively. The robust regression

Table 3 Results of gender classification

View angle of test data	Correct classification rate	
	With view estimation	With ground truth training data
30°	83.5%	86.7%
60°	84.2%	87.5%
120°	85.7%	87.5%
150°	84.2%	86.7%

implemented by MM-estimation performs better than multiple linear regression. Under the same condition of error level, the estimation of gait view by robust regression is more accurate than the estimation results obtained by linear regression. One possible reason may help explain the superiority of robust regression. It is because robust regression methods are designed to be not overly affected by outliers. Due to the instability occurring in our feature generating process, there exist some outliers in the data. If the proportion of a body part for a subject doesn't fit the normal value proposed in [5] in some extent, the features extracted from this subject would not follow the pattern of the other observations and therefore become the outliers. For the formulation used to design the matrix X in Eq. 6, pure cubic formulation outperforms the linear and pure quadratic formulation as demonstrated by the analysis of results. Regression is an important and popular approach for estimation. Based on the exploration described above, the robust regression which is modeled with pure cubic formulation and performed with MM-estimation method is a good choice for estimating gait view. We also consider a real application scenario of gender classification to validate the usefulness of the proposed view estimation. The classification results provide an experimental support for the idea of estimating view angle to deal with the problem of view uncertainty.

7 Conclusion

We propose a novel way to deal with the problem of uncertain view angle in gait recognition systems. The observed subject could appear in any direction in visual surveillance and the recognition performance is impacted by the unknown view variation badly. In this paper, we use the robust regression to estimate the view angle of a test subject. The extensive experimental results demonstrate that the accuracy of estimation is rather reliable to provide necessary prior viewpoint information in a gait recognition system. The regression model

Table 4 Comparison of results from different methods

View angle of test data	Method		
	This paper	Proposed in [13–15]	Proposed in [2, 6]
30°	83.5%	80.6%	78.6%
60°	84.2%	84.5%	80.2%
120°	85.7%	85.0%	81.0%
150°	84.2%	81.1%	79.1%

using pure cubic formulation is better than using linear or pure quadratic formulation to fit the training data generated by our proposed feature extraction method. This feature representation is proved to be view-sensitive experimentally using both Bhattacharyya distance and classification between any pair of different view angles in our BUAA-IRIP Gait Database. Limited by the available multi-view databases, the variance of estimated view angle and the degree of error level extended to achieve a high accuracy are relatively large. Hence, in the future work, we will develop a new database including more different view angles and smaller degree difference between viewing angles. The proposed view angle estimation will be applied and integrated into other gait recognition systems in later experiments.

Acknowledgments This work was funded by the National Basic Research Program of China under Grant 2010CB327902, by the National Natural Science Foundation of China under Grants 61005016, and 61061130560, and by the Fundamental Research Funds for the Central Universities.

References

1. BenAbdelkader C, Cutler R, Nanda H, Davis L (2001) Eigengait: motion-based recognition of people using image self-similarity. In Proceedings of International Conference on Audio- and Video-based Biometric Person Authentication, pp 284–294
2. Bouchrika I, Goffredo M, Carter JN, Nixon MS (2009) Covariate analysis for view-point independent gait recognition. In International Conference on Biometrics, pp 990–999
3. Cunado D, Nash JM, Nixon MS, Carter JN (1999) Gait extraction and description by evidence-gathering. In Proceedings of International Conference on Audio- and Video-based Biometric Person Authentication, pp 43–48
4. Dempster WT, Gaughran GRL (1967) Properties of body segments based on size and weight. *Am J Anat* 120(1):33–54
5. Fukunaga K (1990) Introduction to statistical pattern recognition, 2nd edn. Academic Press, New York
6. Goffredo M, Seely RD, Carter JN, Nixon MS (2008) Markerless view independent gait analysis with self-camera calibration. In Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition, pp 1–6
7. Han J, Bhanu B, RoyChowdhury AK (2005) A study on view insensitive gait recognition. In Proceedings of IEEE International Conference on Image Processing, 1:297–300
8. Huber PJ (1964) Robust estimation of a location parameter. *Ann Math Stat* 35(1):73–101
9. Huber P, Ronchetti EM (2009) Robust statistics, 2nd edn. Wiley Interscience, Hoboken
10. Jean F, Bergevin R, Albu AB (2008) Trajectories normalization for viewpoint invariant gait recognition. In Proceedings of International Conference on Pattern Recognition, pp 1–4
11. Kale A, RoyChowdhury AK, Chellappa R (2003) Towards a view invariant gait recognition algorithm. In Proceedings of IEEE International Conference on Advanced Video and Signal Based Surveillance, pp 143–150
12. Lanitis A, Taylor C, Cootes T (2002) Toward automatic simulation of aging effects on face images. *IEEE Trans Pattern Anal Mach Intell* 24(4):442–455
13. Makihara Y, Sagawa R, Mukaigawa Y, Echigo T, Yagi Y (2006) Which reference view is effective for gait identification using a view transformation model. In Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshop, pp 45–55
14. Makihara Y, Sagawa R, Mukaigawa Y, Echigo T, Yagi Y (2006) Adaptation to walking direction changes for gait identification. In Proceedings of International Conference on Pattern Recognition, pp 96–99
15. Makihara Y, Sagawa R, Mukaigawa Y, Echigo T, Yagi Y (2006) Gait recognition using a view transformation model in the frequency domain. In European Conference on Computer Vision, pp 151–163
16. Nixon MS, Carter JD (2006) Automatic recognition by gait. *IEEE Proceedings* 94(11):2013–2024
17. Niyogi SA, Adelson EH (1994) Analyzing and recognizing walking figures in XYT. In Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp 469–474
18. Rogez G, Guerrero JJ, Martinez J, Orrite C (2004) Viewpoint independent human motion analysis in man-made environments. In Proceedings of British Machine Vision Conference, 2:659–668
19. Rousseeuw PJ, Yohai VJ (1984) Robust regression by means of S-estimators. *Robust and Nonlinear Time Series Analysis* 26:256–274

20. Shakhnarovich G, Lee L, Darrell T (2001) Integrated face and gait recognition from multiple views. In Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1:439–446
21. Tanawongsuwan R, Bobick A (2001) Gait recognition from time-normalized joint-angle trajectories in the walking plane. In Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2:726–731
22. Weisberg S (2004) Applied linear regression, 3rd edn. Wiley Interscience, Hoboken
23. Yohai VJ (1987) High breakdown point and high efficiency robust estimates for regression. *Ann Stat* 15 (2):642–656
24. Yu S, Tan D, Tan T (2006) A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition. In Proceedings of International Conference on Pattern Recognition, 4:441–444
25. Zhang D, Wang Y, Bhanu B (2010) Ethnicity classification based on gait using multi-view fusion. In Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshop, pp 108–115



De Zhang received the B.E degree in electric engineering from the North China University of Technology, Beijing, China, in 2001, and the second B.E. degree in computer science from Tsinghua University, Beijing, in 2003. He is currently working toward the Ph.D degree in the Laboratory of Intelligent Recognition and Image Processing, Beijing Key Laboratory of Digital Media, School of Computer Science and Engineering, Beihang University, Beijing, China. He worked at Visualization and Intelligent Systems Laboratory, University of California, Riverside, CA, USA as a Visiting Scholar from 2009 to 2010. His research interests include computer vision, pattern recognition, and image/video processing.



Yunhong Wang received the B.S. degree in electronic engineering from Northwestern Polytechnical University, Xi'an, China, in 1989, the M.S. degree and the Ph.D. degree in electronic engineering from the Nanjing University of Science and Technology, Nanjing, China, in 1995 and in 1998, respectively. She worked at the National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of

Sciences, Beijing, China, from 1998 to 2004. Since 2004, she has been a Professor with the School of Computer Science and Engineering, Beihang University, Beijing, China, where she is also the Director of Laboratory of Intelligent Recognition and Image Processing, Beijing Key Laboratory of Digital Media. Her research interests include biometrics, pattern recognition, computer vision, data fusion, and image processing.



Zhaoxiang Zhang received the B.S. degree in electronic science and technology from the University of Science and Technology of China, Hefei, China, in 2004 and the Ph.D. degree in pattern recognition and intelligent systems from the Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 2009. In October 2009, he joined the Laboratory of Intelligent Recognition and Image Processing, Beijing Key Laboratory of Digital Media, School of Computer Science and Engineering, Beihang University, Beijing, China as a Faculty Member. His research interests include computer vision, pattern recognition, image processing, and machine learning.



Maodi Hu received the B.E. degree in computer science from Southwest University, Chongqing, China, in 2008. He is currently working toward the Ph.D. degree in the Laboratory of Intelligent Recognition and Image Processing, Beijing Key Laboratory of Digital Media, School of Computer Science and Engineering, Beihang University, Beijing, China. His research interests include pattern recognition, computer vision, machine learning, and motion analysis.