Gait recognition using sub-vector quantisation technique

Neel K. Pandey*

Department of Electrical Engineering and Trades,
Faculty of Engineering and Trades,
Manukau Institute of Technology,
Private Bag 94006, Manukau 2241, Auckland, New Zealand
E-mail: neel.pandey@manukau.ac.nz
*Corresponding author

Waleed H. Abdulla and Zoran Salcic

Department of Electrical and Computer Engineering,
The University of Auckland,
Private Bag 92019, Auckland Mail Centre,
Auckland 1142, New Zealand
E-mail: w.abdulla@auckland.ac.nz
E-mail: z.salcic@auckland.ac.nz

Abstract: Recognising people from their gait is a challenging problem in biometric research. In this paper, we address the problem of gait identification based on a novel approach of sub-vector quantisation (SVQ) technique. A silhouette-based algorithm is utilised to capture the spatial-temporal information of the gait. A sequence of temporally ordered outer contour widths of binarised silhouettes of a walking person represents the feature vectors set. The feature vectors are segmented into sub vectors and vector quantised independently to represent the gait signatures using low dimensional vectors. Dynamic time warping (DTW) technique is used for gait feature sequence matching. The proposed method is validated on several well known benchmarked databases as well as on our own database. The experimental results confirm the validity and robustness of the proposed SVQ method for gait recognition.

Biographical notes: Neel K. Pandey is a Principal Lecturer at School of Electrical Engineering and Trades in Manukau Institute of Technology. He holds a BE (hons.) from University of Technology, Sydney (1995) and an ME (hons.) from University of Auckland (2000). He is currently finishing his PhD part-time from University of Auckland. His interest is in biometric authentication, image processing and pattern recognition. He is a member of IPENZ and is a Chartered Professional Engineer.

Waleed H. Abdulla holds a PhD from the University of Otago, New Zealand. He is an editorial board member of six journals and Vice President-Member Relations and Development of APSIPA. His research interests are in signal and speech processing, and human biometrics. He has published more than 80 refereed publications including a patent and a book. He has supervised more than 25 postgraduate students. He is a recipient of many awards and funded projects exceeding $900 K and was awarded JSPS, ETRI, Tsinghua and HKPU fellowships. He is a senior member of IEEE and member of APSIPA.

Zoran Salcic is a Professor of Computer Systems Engineering and Head of the Department of Electrical and Computer Engineering, University of Auckland, New Zealand. He received his BE (1972), ME (1974), and PhD (1976) degrees in Electrical Engineering from the University of Sarajevo. His main research interests include custom-computing machines, reconfigurable computing, FPGAs, processor and computer systems architectures, embedded systems and their implementation, design automation tools for embedded systems and models of computation and languages for concurrent and reactive systems. He is a Fellow of Royal Society (Academy of Science) New Zealand and recipient of Alexander von Humboldt Research Award.

1 Introduction

Person identification using biometric attributes such as voice, face, fingerprint, gait, etc. has always been attributed to human beings. Gait recognition is a non-invasive process which can be hard to be disguised or concealed. These prominent features have made gait recognition attractive alternative to the other biometric approaches. People identification was investigated in a series of early studies by Johansson (1973) and Cutting and Kozlowski (1977), which showed that gait signatures could be used as a reliable cue to identify people.

Despite the efforts of many researchers, there are still many challenges in order to implement a gait-based person identification system. The limitations extend to capturing of video sequence in controlled or uncontrolled environments, the acceptable level of accuracy rate, the scalability or the size of the gait database, and the usability of the identification system (Nixon at al., 2006).

The main purpose and contribution of this paper can be summarised as follows:

- We developed a simple and effective method for gait-based-human identification using silhouette-based analysis. The silhouette boundary (i.e., outer contour) is analysed instead of silhouettes images usually used in silhouette-based work. This reduces the computational cost of the subsequent processes. We apply sub-vector quantisation (SVQ) technique to extract the gait features.

- The proposed method implicitly captures the finer details of both structural and transitional characteristics of a gait through the use of sub-segmentation. Dynamic time warping (DTW) is used to measure the similarity between known and unknown sample of features. It is envisaged that time normalisation process achieved here will allow more accurate distance estimation and will yield a more accurate result.

- The performance evaluation, in terms of correct recognition and verification rate in is performed on three standard databases: the CMU MOBO database, SOTON database and CASIA-B image databases. The proposed algorithm is also evaluated on our own AUCKLAND image database.

- Many literature reported good recognition results on the standard databases but lacks informed comparisons among different algorithms (Bashir et al., 2008; Yu et al., 2006; Han and Bhanu, 2006; Phillips et al., 2002; Ekinci, 2006; Chai et al., 2006; Ye and Wen, 2006; Kale et al., 2003; Kusakunniran et al., 2009; Bouchrika and Nixon, 2008; Iwashita and Petrou, 2008). Here, we provide some comparison of recognition results to examine the performance of the proposed approach and other recent approaches on the standard databases.

This rest of the paper is organised as follows. The proposed approach and methodology are described in Section 2. Section 3 describes the experiments conducted and presents the results. In this section, we compare the recognition/verification accuracy rate of our algorithm against other works in this area using the benchmarked image databases. Finally, Section 4 concludes the paper.

2   The proposed approach

The proposed work is based on the implicit use of vector quantisation (VQ) technique. Specifically, we extract spatiotemporal information from the sub-vectors of the outer contour width (SVOCW) which is obtained from the silhouettes of the gait sequence. It is envisaged that our method for gait feature generation improves by encapsulating the finer details of the gait pattern and overcomes the data dimensionality issue without any loss of critical information. The aim is to exploit the redundancy in the gait data and to derive a compact representation of the gait signature. To obtain a compressed form of gait signature we have employed VQ technique using LBG algorithm (Linde et al., 1980). It is also envisaged that the unique codebook designed for each temporally aligned segmented data vector in a gait sequence will not only capture intricate details of gait pattern but will also produce uniquely quantised data representation. The quantised and non-quantised probe segment is matched against the quantised reference set. Inter-variation in gait results in varying lengths of walk-cycles. The challenge is to eliminate the timing differences between two gait sequences for matching. Hence, DTW (Sakoe and Chiba, 1978), which has a non-linear time-normalisation effect, is used as a pattern matching algorithm. The methodology has been described in Pandey et al. (2005, 2007).

Our goal is to capture the gait features, using the holistic-based approach, from the image data from a single viewpoint in a constraint environment using a non-calibrated camera. Recently, it is more common that several viewpoints image data are collected by placing multiple cameras at different angles. In this study, we also evaluate our algorithm over different viewpoints to extract gait features. An overview of the proposed method is illustrated in Figure 1.

Figure 1 Data processing steps of the proposed method
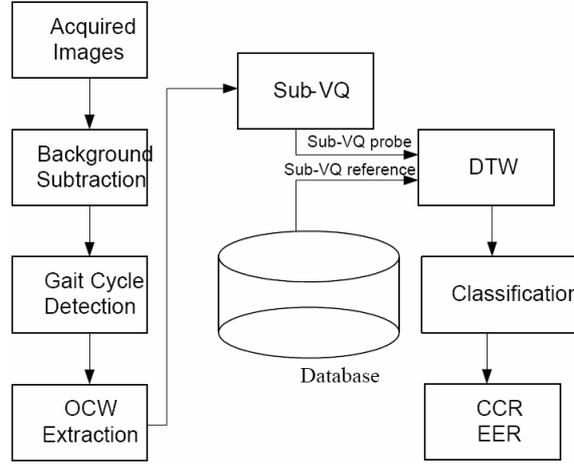
The robustness of the gait signatures needs to be tested on benchmarked databases so that valid comparisons and conclusions could be drawn in relation to other reported research results. There are several commonly used image databases that are available to benchmark results and to compare with other researchers. In our research, we have used CMU (Gross and Shi, 2001), CASIA-B (Yu et al., 2006) and SOTON (small) (Shutler et al., 2002) image databases. To further explore and gain insight of the complete system, we have created our own AUCKLAND image database of 25 people.

2.1 Silhouette extraction

In vision-based systems, detection is usually carried out by using adaptive background subtraction method. It is normally assumed that the background is fixed and that differences are solely caused by foreground objects. A simple static background segmentation approach is extremely sensitive to dynamic scene changes due to lighting and extraneous events. More complex methods build a model of the scene background, and for each pixel in the image detect deviations of pixel feature values from the model to classify the pixel as belonging either to background or to foreground.

In our research, we adopted an adaptive background subtraction method based on McKenna et al. (2000). We assume static camera and background. The method is capable of handling background changes that are relatively slow to the motion of people in the scene. We considered two aspects, the normalised *rgb* colour space and the edge for image extraction as we employ the method in a constrained environment. For each channel the models are generated and the current image is converted to the adequate form such as edge image and RGB image, which are used apart for the further classification. The combination of both classification results gives the final segmentation mask.

The adaptive process is only performed in regions of the image which has higher-level grouping labelled as background. Given a new pixel value, (R, G, B), the following updates as shown in equations (1) and (2) are performed:

$$\mu_r = \alpha r + (1 - \alpha)\mu_r \tag{1}$$

$$\sigma_r^2 = \alpha \left(r - \mu_r \right)^2 + (1-\alpha)\sigma_r^2 \qquad (2)$$

where μ_r is the mean pixel value computed up to frame t, α is the learning rate of the model, and r is the red pixel value in frame t. Similarly, the update is carried out for the green (g) and blue (b) pixel values. The current pixel $x(r, g, b)$ is compared to the model. If $| r - u | > 3\sigma$ or a Euclidean threshold then the pixels are considered as true and set as foreground. Similar test for g and b is conducted. A mask is produced which is considered as a region of interest for further processing.

Edge gradients are estimated using the Sobel masks in horizontal, x, and vertical, y, directions. Given a new pixel value $x(r, g, b)$, its spatial gradients (r_x, r_y), (g_x, g_y), (b_x, b_y) estimated using the Sobel operator. If $\sqrt{(r_x - \mu_{xr})^2 + (r_y - \mu_{yr})^2} > 3\sigma$, or if the similar test for g or b is true, then the pixel is set to the foreground. Otherwise it is set to the background. Figure 2 shows the extracted silhouette, edge gradient and the picture overlay extracted based on adaptive background subtraction method. Figure 2(a) shows the filled binary region of the silhouette based on the edge detection. Figure 2(b) shows the extraction of the actual image based on the binary silhouette which indicates the accuracy of the size of the extracted binary silhouette for the formation of the gait features.

Figure 2   Extracted silhouette using adaptive background subtraction method

(a)   (b)

2.2   Gait signature representation

The underlying principle used in our approach is that given a sequence of image features of a person, $X = \{x(1), x(2),\ldots,x(T)\}$, we develop a gait signature for that person and use it to identify the person from a database of N different subjects. Fronto-parallel view of a person, which is illustrated in Figure 3, provides the most information on a person's gait.

Figure 3   (a) Illustration of the fronto-parallel view and (b) the calculation of outer contour width

           (a)   (b)
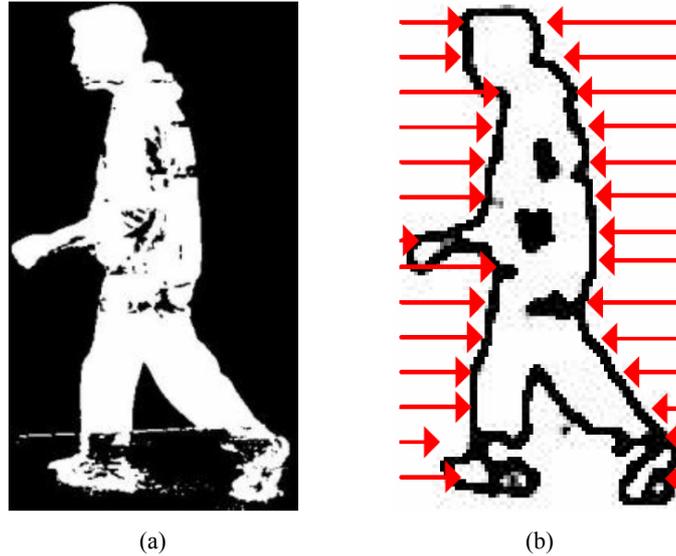
Features like the stride length, arm swing and physical structure are quite coherent in sequence of side-view images. Thus, extraction of feature set, while preserving the structural component and motion dynamics, is of great importance. The emphasis is that the side view provides crucial information on a person's gait. However, other views could provide valuable complementary information. Gait features are extracted from heel-to-heel strike of sequence of silhouettes.

2.2.1   Basic feature vectors set

The basic feature vector set is derived from the width of the outer contour of the temporally ordered silhouettes (Kale et al., 2003). The width along a given row is simply the difference in the location of rightmost and leftmost boundary pixel in that row as shown on Figure 3(b). The gait signature aims to capture the static (structural) component that captures the physical build of a person, e.g., body dimensions, length of limbs, etc., and the motion dynamics of the body during a gait cycle. The physical structure of the subject, as well as the swing of the limbs and other details of the body, is retained in the width vector. The discriminatory features of the width vectors are directly dependant on the quality of the silhouettes. To reduce the noise effect of silhouettes, a 3×3 averaging filter is applied on the width vectors for smoothing. The width vectors as a function of time emphasises the individuality of the person. The overlay of the smoothed width vectors is shown in Figure 4 for one individual.

Figure 4   Segmented blocks of width vectors
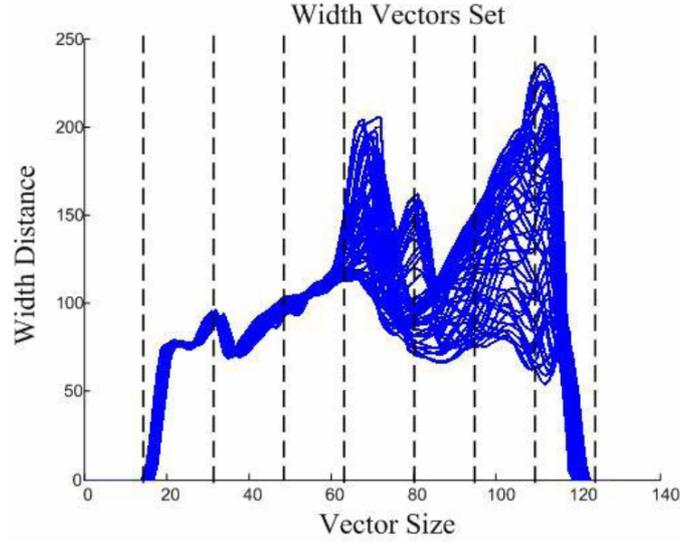
2.2.2 Sub-vector extraction method

The problem of finding and exploiting low-dimensional structures in high-dimensional data is arduous (Boulgouris et al., 2004). The curse of dimensionality is a challenge and we often need to conduct meaningful inference with limited number of samples in a very high-dimensional space. There are several motivations to reduce the number of dimensions such as: efficiency, classification performance and ease of interpretations/ modelling.

In this study, the width vector dimension is the number of rows representing the silhouette. In other words for a given frame size, the width vector would be of a fixed dimension. In a multi-view environment where several cameras are placed at an angle, the size of silhouette image will change according to the projection geometry. To reduce the effect of changes in silhouette size, silhouette image is size-normalised. The size-normalisation proportionally resizes each silhouette image so that all silhouettes have the same height. The width-to-height aspect ratio of each silhouette remains constant throughout the sequence. If the original silhouette size is denoted as $H_o x W_o$ and the normalised silhouette as HxW, then $f_h = H / H_o$ and $f_w = W / W_o$. To keep the same aspect ratio we maintain $f_h = f_w$. In the case of height normalisation we keep the height at a constant pixel length, hence the width of silhouette is resized in the horizontal direction with a scale factor of $f_w = f_h = H / H_o$. Therefore, the width of the normalised silhouette is $f_h W_o$. We have selected to normalise the height to a length of 128 pixels. This means a silhouette image is normalised to a ratio of $128 / H_o$. Empirically, we have found that a bounding box of size 128×88 pixels would incorporate all normalised images in a gait walk cycle. Hence, the normalised silhouette is aligned by shifting it horizontally to the centre of the image frame of size 128×88 pixels.

We take this result further to represent our feature vector in a lower dimension space. In order to better model the signature of a sequence of silhouette without losing vital information, the image width vectors of size *MxK* are split into *MxPxR* blocks, where *M*

is number of width vectors (image frames); K is the dimension of each vector. *P* and *R* are the dimension and number of the sub-vectors block respectively.

2.2.3 *VQ of width vectors set*

The motivation of feature extraction is to exploit the redundancy in the gait data for dimensionality reduction and to derive a compact representation of the gait signature. We have employed VQ technique using LBG algorithm to obtain a compressed form of gait signature. By applying the VQ technique on several gait cycles of sub-width vector, we derive a codebook for each sub-width vector set that represents the gait signature in the compressed format while encapsulating the static and dynamic features. It is envisaged that by generating a codebook for each of the *R* blocks of *P*-dimension dataset, a better model of the signature is obtained by capturing the discriminatory features of each block. Each segment of the test and reference sequences is quantised against the corresponding codebook segment. The LBG algorithm is described in detail in the paper by Linde et al. (1980). Figure 4 illustrates the segmentation of a set of the un-normalised width vectors extracted from two-walk-cycle span of silhouette images. In this example the span size is 61×128. The width vectors are segmented into eight sub-vectors, each of length 16 (i.e., $61 \times 16 \times 8$).

2.2.4 *Selection of the number of codevectors (codebook size)*

Heuristics approach is one of the methods which could be taken in selecting the size of the codebook. In order to select the codebook size with some objectiveness, mean squared error (MSE) for different codebook sizes (*N*) namely, 128, 64, 32, 8, 4, 2 and 1, were evaluated in this study. MSE provides a good measure in terms of closest representation of similarity between codevector $\hat{x}$ and the vector *x*. The codebook size: ($K = 32$) was selected at knee-point of the quantisation vs. codebook size graph plot as shown in Figure 5. Empirically, the selected value is found to be a good compromise between the MSE and the exhaustive quantisation process.

Figure 5   Codebook selection graph plot
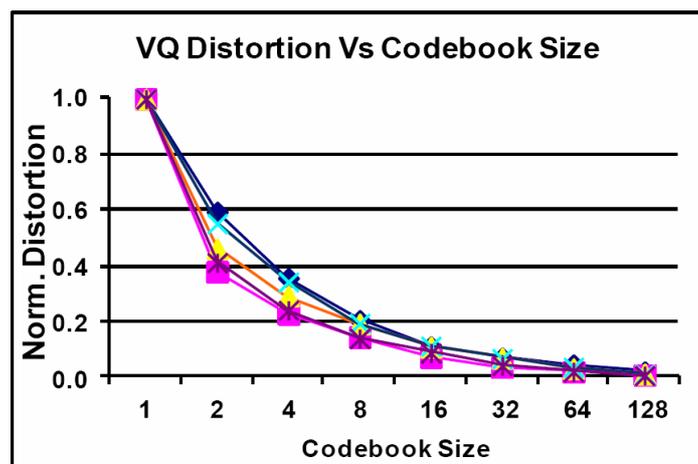
2.3   Gait cycle detection

The detection of gait cycle is also important in gait analysis processes. The gait sequence is partitioned into cycles that depict a complete walking period by locating the frame indices at which the temporal sum of the width vectors is maximised. A *walk cycle* comprises two consecutive double-support stances as shown in Figure 6.

Figure 6   Illustration of a gait walk-cycle

Double Support   Midstance   Double Support   Midstance   Double Support
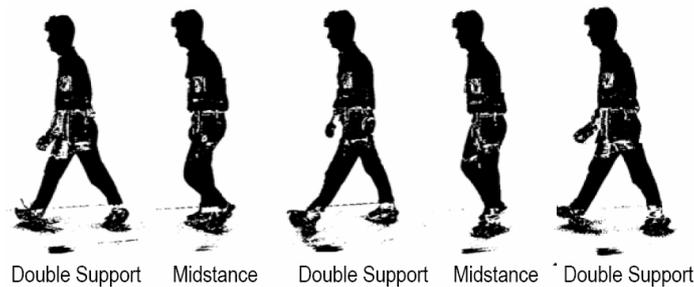
To derive the gait cycle length, temporal sum of the width vectors is performed. Since the width vectors are dependent on the quality of the silhouettes, sum of the width vectors with respect to time could be noisy. Hence, we use a simple two-element running average filter to clean any noisy signal. Figure 7 shows the variance-normalised temporal sum of the width vectors. We identify the cycle length TG from the local maxima, which is represented by two consecutive double-support stances. Image frame indices of peaks illustrated in Figure 8 represent the half-cycle double-support stance.

Figure 7   Normalised plot of temporal width vectors indicating the walk cycle (see online version for colours)
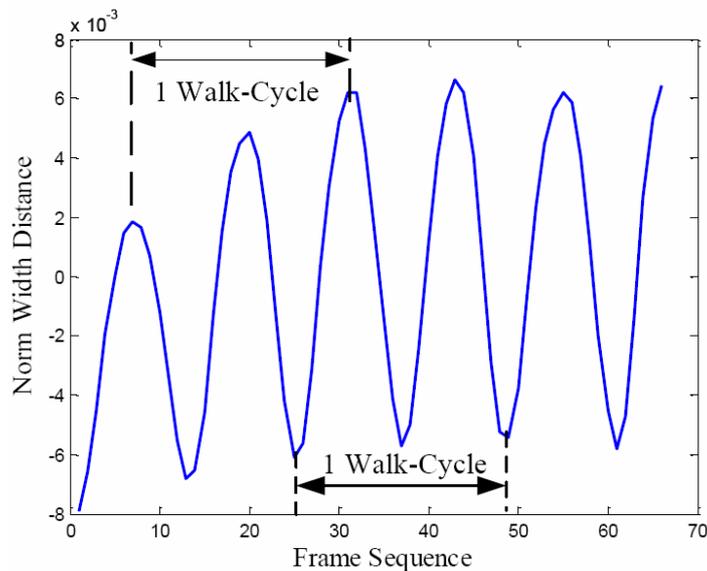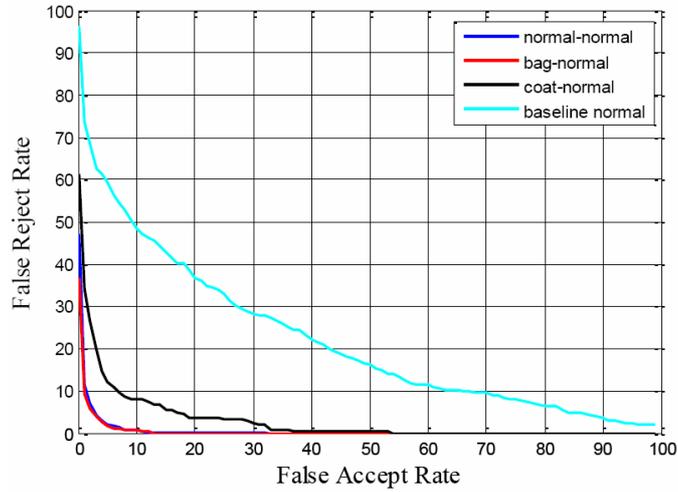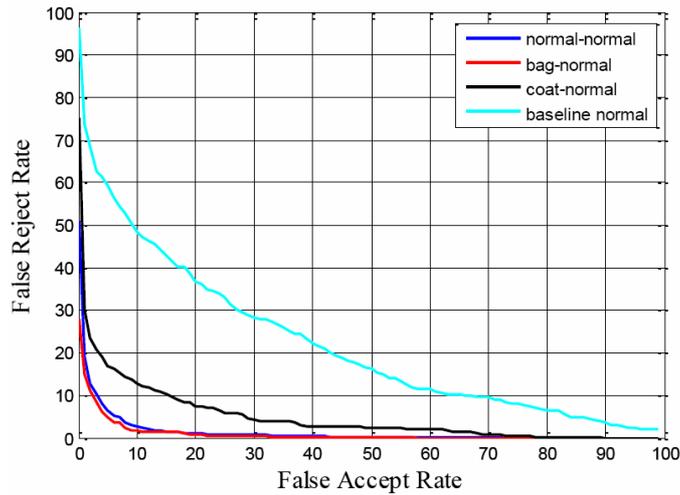
Figure 8 Verification curves of using CASIA-B database for (a) VQ probe vs. VQ reference and (b) non-VQ probe vs. VQ reference condition for 90 degree view angle

(a)

(b)

2.4 Template matching

The collection of the gait width vector sequences, from which the codebook of a person is derived, will be termed *gallery* sequence. A walk-cycle sequence, which is a subset of *gallery* sequence, is used to represent the known gait sequence and will be termed *reference* sequence. The unknown test gait sequence will be termed *probe* sequence. After the extraction of outer-contour width vectors from the sequence of silhouette, the probe and the reference sub-segmented sequences are quantised against the corresponding codebook as described in Section 3.1. DTW is performed on sub-segment

sequences for matching. The final matching score in this experiment is the sum of all the sub-segments score for the given test sub-vectors set.

Inter-variation in gait is a well known phenomenon and results in varying lengths of walk-cycles that requires elimination of the timing differences between two gait sequences. DTW (Sakoe and Chiba, 1978) is used to model the non-linear time-normalisation effect and pattern matching. Given two cycles of different duration, the warping function maps the time axis of the probe cycle to the time axis of the reference cycle.

2.5 Classification

A simple nearest neighbour (NN) classifier is used to perform classification in this study. To cater for any variation is the gait style and to increase the chances of correctly recognising a person, several walk-cycle sequences of a person can be employed to form a probe and reference sets. A probe set, which consists of several walk-cycles of probe sequences can be represented as: $G_p = \{X_p^1, X_p^2,…,X_p^{np}\}$, np is the number of probe walk-cycles. Likewise, a reference set which consists of several walk-cycles of reference sequences can be represented as: $G_r = \{X_r^1, X_r^2,…,X_r^{nr}\}$, where nr is the number of reference walk-cycles and X represents one walk-cycle gait sequence data.

DTW is applied between a probe sequence, p, and a kth reference sequence, r, in the database. The sum of scores for all combinations of sequences between the probe and reference set is calculated and can be represented as (3):

$$Dist\left(G_p, G_r^k\right) = DTW\left(X_p^i, X_r^{jk}\right), \text{ where } i = 1,…,np \text{ and } j = 1,…,nr \tag{3}$$

The matching score for the given probe set with those of a given reference set is given as the minimum distance between the probe set and the reference set can be represented as (4).

$$Matching\ score = Min\left(Dist\left(G_p, G_r^k\right)\right) = Min\left(DTW\left(X_p^i, X_r^{jk}\right)\right) \tag{4}$$

For classification, the probe set will be classified into class k if the kth reference set produces the minimum score, shown in (5).

$$k = \arg\min_{(n=1…K)} Dist\left(G_p, G_r^k\right) \tag{5}$$

where $n = 1,…,K$ for K reference set in database.

To elaborate the method using one of the experimental conditions used in this research, we used a probe set G_p which consisted of two probe sequences; each sequence is of a walk-cycle length. Likewise, the reference set G_r consisted of four reference sequences. We denote the probe set as $G_p = \{X_p^1, X_p^2\}$ and reference set as $G_r = \{X_r^1, X_r^2, X_r^3, X_r^4\}$. Table 1 illustrates the score set $S = \{S_{11}, S_{12},…,S_{24}\}$ obtained from matching each of the probe sequence set to each of the reference sequence set using DTW.

Table 1   Score set for all combination of probe and reference set

Probe set G_p	Reference set G_r			
	X_r^1	X_r^2	X_r^3	X_r^4
X_p^1	S_{11}	S_{12}	S_{13}	S_{14}
X_p^2	S_{21}	S_{22}	S_{23}	S_{24}

Using the minimum method, the minimum score $S_{i,j}$ represents the matching score of the probe set with that to the reference set and can be represented as (6).

$$Matching\ score\ S_{pr} = Min\left(Dist\left(G_p, G_r\right)\right) = Min\left(DTW\left(X_p^i, X_r^j\right)\right)$$
$$for\ i = 1, 2\ and\ j = 1,\dots,4 \tag{6}$$

3   Experimental results

The performance of our proposed SVQ-based gait algorithm has been tested on three well known gait databases, CMU Mobo gait database (Gross and Shi, 2001), CASIA-B gait database (Yu et al., 2006), SOTON small gait database (Shutler et al., 2002), and also on our own AUCKLAND database. The codebook was generated from the gallery sequences, and the reference sequences were represented by the trials of gait sequences within the gallery sequence. The probe and reference sequences are of one walk-cycle length.

Given a probe sequence data, two conditions of experiments were conducted:

1   vector quantised probe vs. vector quantised reference (VQ-VQ)

2   non-vector quantised probe vs. vector quantised reference (non-VQ-VQ) dataset.

In Phillips et al. (2002), a human ID gait challenge framework and a referential gait recognition baseline algorithm are proposed. For comparison of our results, we implemented the baseline algorithm and applied it to the same sets of gait sequences data as used on our proposed algorithm. We also compared our results of the proposed method with other well cited gait recognition approaches using the same gait databases used in this study.

In order to study the efficiency of the gait recognition, we analyse the results in terms of

a   correct recognition rate

b   verification rate.

Correct recognition rate is measured by finding the ratio of the number of correctly identified persons against the total number of persons in the database. Verification rate is measured using the area-under-the-curve (AUC) of the receiver operating characteristic (ROC), which is a plot of the probability of detection (i.e., correct recognition or true positive), p(true positive), vs. the probability of a false alarm (i.e., false acceptance or false positive), p(false positive), for the experiments. Equal error rate (EER), which is the threshold value where the probability of false rejection is equal to the probability of false

acceptance, is also obtained from ROC plot to provide an objective measurement for comparison between different systems. In general, the lower the EER, the more accurate the system is considered to be.

Based on the outlined performance criteria, comprehensive experiments have been conducted to evaluate the algorithm using the standard databases. Since the focus of each database differed in terms of the co-variables such as the camera view angles, clothing and speed conditions, we evaluated the performance of our algorithm based on these covariates accordingly. The type of evaluation performed is as follows and will be detailed in later sections:

- *CASIA-B database:* we evaluate our algorithm in terms of view-dependent recognition, view-invariance performance and shape invariance performance.

- *CMU and SOTON database:* we evaluate our algorithm in terms of covariate analysis of shape invariance and gait speed invariance performance of a fixed view angle. A similar experiment was conducted for our locally produced AUCKLAND database.

3.1   CASIA-B gait database

The CASIA gait database is an indoor gait database consisting of 124 subjects captured from 11 different views simultaneously starting from 0 to 180 degrees with an increment of 18 degrees. For each view angle, each subject has ten gait sequences which comprise of six normal, two with a coat and two with a bag. In this experiment, gait data of 105 people with view direction angle of 54, 72, 90, 108 and 126 degrees is used. This gives a view range of 36 degrees on either side of fronto-parallel (90 degrees) view. The experiments carried out using this database can be categorised in two ways:

a   the performance of algorithm based on acquisition of invariant gait features for different view angles and silhouette conditions

b   the study of the number of probes, references and the gallery size for feature extraction and recognition rate.

For the case in experiment (a), with the five selected view angles of 54, 72, 90, 108 and 126 degrees, we conducted the following experiments:

- *View-dependent performance:* For each of the view angle image data, we recognised a probe sequence against each of the reference sequences, i.e., the same view angle for the probe and reference sequence. The recognition rate will give us the measure of the performance of the proposed algorithm for the same viewing condition. The results are presented in Table 3.

- *View-invariance performance:* We conducted this evaluation by attempting to recognise a probe sequence of one view against reference samples of other view angles. For instance, a probe sequence of 90 degree view is tested against all other view angles reference sequences. The results are presented in Table 4.

- *Shape-invariance performance:* We tested the two shape-differed probe sequences against normal reference sequences: one where the subject wears a coat and in the other the subject carries the bag. The results are presented in Table 5.

For the case in experiment (b), to study the effectiveness of algorithm due to limited sample size, we conducted the following experiments:

- *Leave-one-out method (condition 1):* In this case, a codebook was generated by leaving one probe sequence out and using five normal sequences to form the gallery of sequences. Each reference sequence was vector quantised using the codebook generated. This method effectively gave six probe sequences and six sets of reference sequences.

- *Two probe sequences:* In the first case (condition 2), we based our experiments using leave-one-out method but using only two probes as opposed to all six sequences as probe sequences. For instance, for the first probe sequence trial 1 data sequence was used as the probe sequence and trials 2 to 6 data sequences formed the gallery to generate the codebook and respectively formed the five quantised reference sequences, and for the second probe sequence trial 2 data sequence was used as the probe sequence and remaining five data sequences as the reference sequences. In the *second case (condition 3)*, we took a simpler approach of selecting two probe sequences and the remainder of the four sequences formed the gallery to generate the codebook and respectively formed the four quantised reference sequences data. In this experiment, trials 1 and 2 data sequences formed the probe and trials 3 to 6 data sequences formed the reference sequences. The results of all experiments are summarised in Tables 3, 4 and 5.

As seen in Table 3, the performance of our algorithm performs best when all probes are utilised using leave-one-method (condition 1). Although the proposed representation works best for condition 1, the difference in the performance, taking into account of the simpler approach of condition 3 is negligible and similar. The view dependent performance of the proposed algorithm seems to give better results for the case of non-VQ probe when compared to VQ-VQ probe and for the view angle of 90 degrees as reported in Table 3 and Table 4. This difference in performance between non-VQ-VQ and VQ-VQ could be attributed to the high distortion cost in the case of VQ probe tested against VQ reference if the compared vectors are assigned to different codewords.

The high identification rate for 90 degree view reinforces the point that the fronto-parallel view angle captures the most gait information. Table 4 shows the variation of recognition rate with respect to different view angles. For the non-diagonal cases, the recognition rate within the 18 degree view range is an average of 40.5% and drops largely to an average of 11.8% outside this view range. This indicates that due to the excessive viewpoint difference between the reference and test sequence, the distortion on extracted feature vectors is quite considerable and has a detrimental effect on the recognition performance.

Table 5 shows the covariate effects of clothing and carrying conditions affecting the silhouette shape independent from viewpoint. Based on the proposed approach for deriving the gait feature, an average recognition of 70.3% is achieved for bag carrying and 33.9% for coat wearing conditions. The results indicate that the algorithm is sensitive to structural changes and the unique dynamics features of the gait is better captured in bag carrying conditions when compared to coat wearing condition. However, it is important to note that gait information is compromised in both conditions.

The recognition rate results in Table 2 show that the proposed method performs better than the baseline method in all experiments for 90 degree view angle condition. The result of person wearing coat indicates that dominant features are occluded during feature extraction process, which has adverse effect when compared to person carrying bag. The proposed approach for non-VQ-VQ condition performed better than majority of the compared methods. Bashir et al. (2008) report higher recognition rate for supervised learning method, however, it comes with a computationally intensive feature extraction process due to the data dimension and the use of greedy search algorithm. The proposed method uses the sub-vector technique which not only overcomes the data dimension issue but is less computationally intensive.

Table 2   Comparison of fronto-parallel view results

Method	NM-NM	BG-NM	CT-NM
Bashir et al. (2008) (un-sup)	99.4	79.9	31.3
Bashir et al. (2008) (sup)	98.6	85.5	88.8
Yu et al. (2006)	97.6	52.0	32.7
Han and Bhanu (2006)	99.4	60.2	22.0
Phillips et al. (2002)	49.3	20.7	11.9
Proposed method (VQ-VQ)	*98.1*	*56.2*	*45.7*
Proposed method (non-VQ-VQ)	*100*	*77.1*	*43.8*

Table 3   Summary results of view dependent recognition rate (%)

Condition	View angle					
	90 deg		*108 deg*		*126 deg*	
	Non-VQ probe	*VQ probe*	*Non-VQ probe*	*VQ probe*	*Non-VQ probe*	*VQ probe*
Condition 1	100	98.1	97.1	88.6	91.7	89.5
Condition 2	99.1	92.4	95.2	84.8	92.4	89.5
Condition 3	97.1	87.6	94.3	85.7	88.6	84.8
	72 deg		*54 deg*		*Average results*	
	Non-VQ probe	*VQ probe*	*Non-VQ probe*	*VQ probe*	*Non-VQ Probe*	*VQ probe*
Condition 1	98.1	96.2	100	100	97.4	94.5
Condition 2	95.2	92.4	97.1	96.2	95.8	91.1
Condition 3	94.3	88.6	96.2	93.3	94.1	88.0

Table 4   Summary results of view invariance recognition rate (%)

	126 deg	*108 deg*	*90 deg*	*72 deg*	*54 deg*
126 deg	92.4	36.2	20	9.5	3.8
108 deg	43.8	95.2	53.3	18.1	7.6
90 deg	20	49.5	99.1	42.9	11.4
72 deg	11.4	20	43.8	95.2	33.3
54 deg	5.7	4.8	9.5	20.9	97.1

Table 5   Summary results of shape invariance recognition rate (%)

	View angle					
	90 deg		*108 deg*		*126 deg*	
	Non-VQ probe	*VQ probe*	*Non-VQ probe*	*VQ probe*	*Non-VQ probe*	*VQ probe*
Coat vs. Nm	43.8	45.7	29.5	18.1	27.6	21.9
Bag vs. Nm	77.1	56.2	62.8	31.4	73.3	36.2
	72 deg		*54 deg*		*Average results*	
	Non-VQ probe	*VQ probe*	*Non-VQ probe*	*VQ probe*	*Non-VQ Probe*	*VQ probe*
Coat vs. Nm	32.4	36.2	36.2	38.1	33.9	32
Bag vs. Nm	71.4	56.2	66.7	57.14	70.3	47.4

The proposed approach was also evaluated in terms of verification performance. An EER plot is used to measure the verification performance in this study, as illustrated in Figure 8 for 90 degrees view. It can be seen that the proposed algorithm outperforms the baseline method for both conditions. An EER between 3.5–5.7% is obtained for the normal-to-normal condition testing and 3.5–11.9% is obtained for cross-condition testing compared to 29–35.5% for baseline method. The low EER values indicate the verification accuracy of the proposed algorithm when compared to the baseline algorithm. Therefore, a 3.5% EER indicates that there is a possibility of 3.5 people out 100 users who pose as intruders would be given a false acceptance and likewise 3.5 genuine users will be denied access to the system. Furthermore, an AUC of 0.992 and 0.984 is obtained in VQ-VQ and non-VQ-VQ testing for the normal-to-normal condition compared to 0.775 for the baseline algorithm. In the case of bag-to-normal condition, an AUC of 0.994 for VQ-VQ and 0.989 for non-VQ-VQ testing is obtained compared to 0.653 for the baseline algorithm. Similarly, for coat-to-normal testing, AUC of 0.968 for VQ-VQ and 0.947 for non-VQ-VQ testing is obtained compared to 0.747 for the baseline algorithm. The overall AUC results show a high classification performance using the proposed approach.

3.2   CMU gait database

To further study the effect of covariate factors in terms of shape and speed on view angle independent gait, experimental analyses were carried out on the CMU, SOTON and our own AUCKLAND database using only the fronto-parallel gait image data. CMU database has 25 subjects (23 males, 2 females) walking on a treadmill. There are about 8 walk-cycles in each sequence and each sequence is recorded at 30 frames per second. It also contains six different views of simultaneous motion sequences of 25 subjects. The image size was 480×640 pixels. Each frame produced 1×640 width vector which was down-sampled by 5-pt to obtain 1×128 raw width vector.

In this study, only six side-view walking sequences of each subject are considered. The probe and reference sequences were extracted and prepared as explained in Section 3. The following experiments for both VQ-VQ and non-VQ-VQ conditions were conducted using this database:

1   train on slow walk (SW) and test on SW

2   train on fast walk (FW) and test on FW

3   train on SW and test on FW

4   train on FW and test on SW.

From Table 6, it is clear that the recognition performance of the proposed method for the non-VQ-VQ condition is much better than that of the baseline algorithm and the other methods reported for fast-fast and slow-slow testing.

Table 6   Gait recognition results of CMU database

Method	FW-FW	SW-SW	FW-SW	SW-FW
Ekinci (2006)	92	84	32	52
Chai et al. (2006)	97	95	76	68
Ye and Wen (2006)	96	96	64	48
Kale et al. (2003)	68	72	58	32
Kusakunniran et al. (2009)	100	100	92	92
Phillips et al. (2002)	88	92	82	75
Proposed method (VQ-VQ)	*100*	*100*	*76*	*40*
Proposed method (non-VQ-VQ)	*100*	*100*	*100*	*80*

Figure 9   Verification curves for (a) proposed algorithm for non-VQ vs. VQ condition and (b) baseline algorithm using CMU database
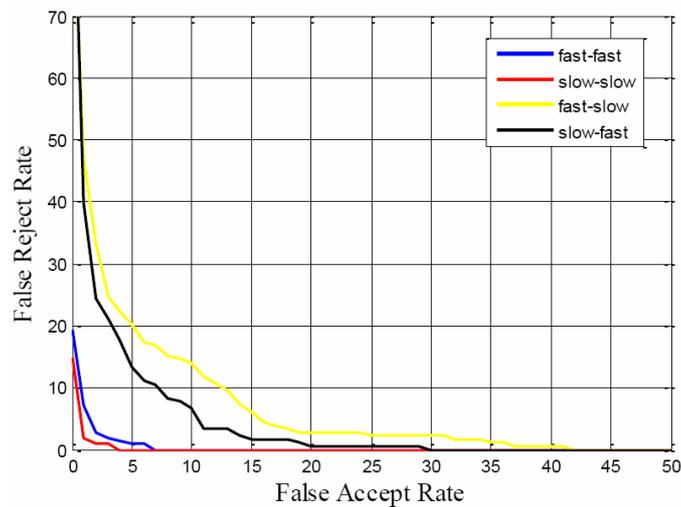
(a)

Figure 9 Verification curves for (a) proposed algorithm for non-VQ vs. VQ condition and (b) baseline algorithm using CMU database (continued)
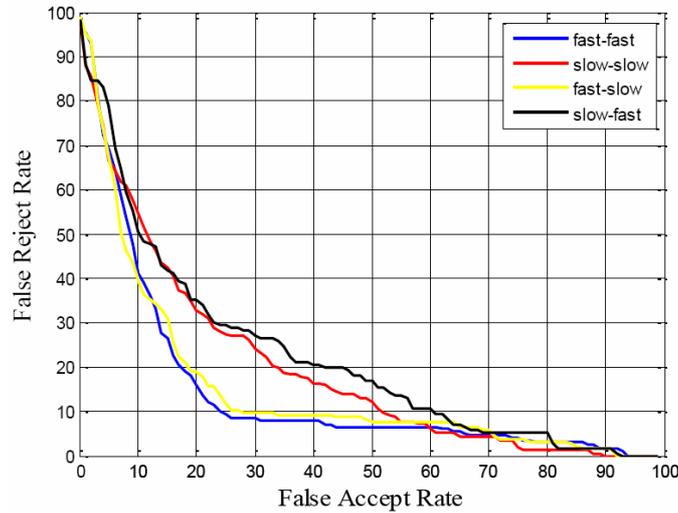
(b)

The proposed method performs better (100%) when compared to Kusakunniran et al. (2009) (92%) and all the other methods for the fast-slow case. However, it is slightly down (80%) for slow-fast case in comparison to Kusakunniran et al. (2009) (92%). The slight difference in performance could be attributed to the difference in the selection of the silhouettes for gait feature representation, changes in body dynamics and stride lengths due to variation in walking speed. Figures 9(a) and 9(b) shows the proposed system evaluation in terms of verification performance using the EER plot. An EER between 1.44–2.4% is obtained for the same condition testing and 8.2–11.4% is obtained for cross-condition testing compared to 18.5–27.2% for the baseline method. The AUC between 0.992–0.998 is obtained for the same condition testing and 0.957–0.98 is obtained for cross-condition testing compared to 0.777–0.853 for the baseline method.

3.3 SOTON gait database

The SOTON-small gait database consists of 11 subjects filmed walking around the inside track, with a green chroma-key backdrop. Each subject was filmed wearing a variety of footwear, clothes and carrying various bags. They were also filmed walking at different speeds. Each subject's data was captured during one long walking session. We used six side-view gait sequences for experimentation. The first three sequences of each individual are used as the probe sequences and the last three sequences are used as reference sequences. Experiments were carried out for two view-types:

a   person walking from left to right

b   person walking from right to left.

The following experiments for both VQ-VQ and non-VQ-VQ conditions were conducted for each of the view-type using this database:

1   train on normal (NM) and test on NM

2   train on fast (FT) and test on FT

3   train on NM and test on FT

4   train on FT and test on NM

5   train on bag (BG) and test on BG

6   train on NM and test on BG

7   train on FT and test of BG.

The average correct recognition results of both view types are summarised in Table 7.

Table 7   Gait recognition results of SOTON database

| Method	FT vs. FT	NM vs. NM	FT vs. NM	NM vs. FT	BG vs. BG	BG vs. FT	BG vs. NM
Bouchrika and Nixon (2008)	-	95.7	-	-	-	-	-
Iwashita and Petrou (2008)	-	94	-	-	-	-	-
Phillips et al. (2002)	81.8	86.8	79.8	77.3	94.5	40.9	43.9
Proposed method (VQ-VQ)	*95.5*	*100*	*81.8*	*81.8*	*100*	*36.4*	*54.5*
Proposed method (non-VQ-VQ)	*100*	*100*	*90.0*	*90.9*	*100*	*63.6*	*68.1*

The results of our proposed method compare favourably with others. Obviously, the performance decreases for cross condition testing when the person is carrying a bag. This indicates the change of dynamic and static information in the feature vectors caused by the bag carrying condition. However, the results for testing against different speed perform well.

Figure 10 Verification curves for (a) proposed algorithm for non-VQ vs. VQ condition and (b) baseline algorithm using SOTON database
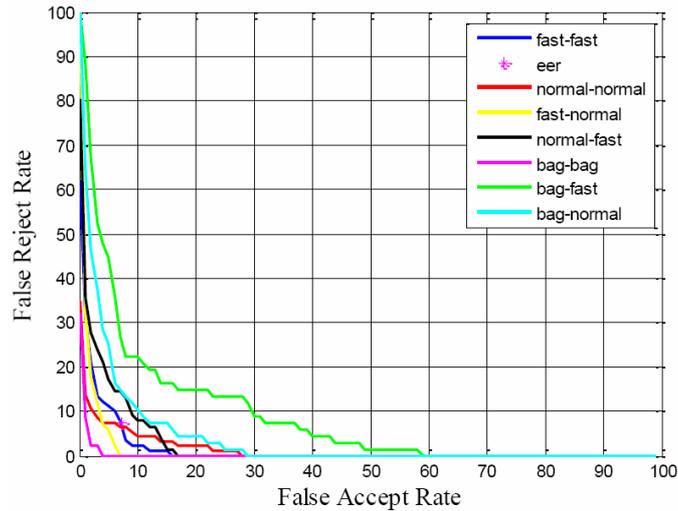
(a)

Figure 10 Verification curves for (a) proposed algorithm for non-VQ vs. VQ condition and
　　　　　　(b) baseline algorithm using SOTON database (continued)
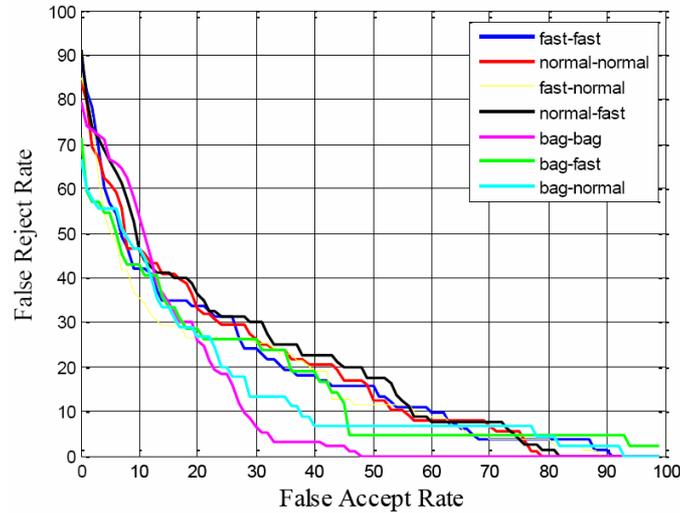
(b)

Figures 10(a) and 10(b) show the EER curves for the proposed algorithm for non-VQ-VQ condition and baseline method. The EER of the proposed method varies between 2.2–16.2% compared to 20.4–27.7% for baseline method. The AUC between 0.962–0.995 for same-condition testing and 0.906–0.982 for cross-condition testing is obtained, which again indicates a high classification rate.

3.4 AUCKLAND gait database

To gain a better understanding of the whole process of the gait recognition method, we created our own image database: AUCKLAND database. This database has 25 subjects (20 males, five females). Each subject is recorded performing two different types of walk: SW and FW. There are about two and half walk cycles in each sequence recorded at 25 frames per second. The database has 24 walking sequences for each individual consisting of six normal walking sequences and six fast walking sequences in right-to-left direction and vice-versa in fronto-parallel view. For each condition, the first three sequences of each individual are used as the probe sequences, and the last three sequences are used as reference sequences.

　　Experiments were carried out for two view-types:

a　person walking from left to right

b　person walking from right to left and similar to those using CMU database.

The correct recognition results are summarised in Table 8. Once again, the proposed method performs better than the baseline method. Figures 11(a) and (b) shows the EER curves for the proposed and baseline method. The EER for same condition ranges from 2.75–3.35% and 6.3–6.9% for cross condition in non-VQ-VQ case. It can be seen that these results are significantly better than the baseline method which ranges from

19.2–25.8%. The AUC between 0.992–0.994 is obtained for the same condition testing and 0.974–0.983 is obtained for cross-condition testing compared to 0.819–0.852 for the baseline method.
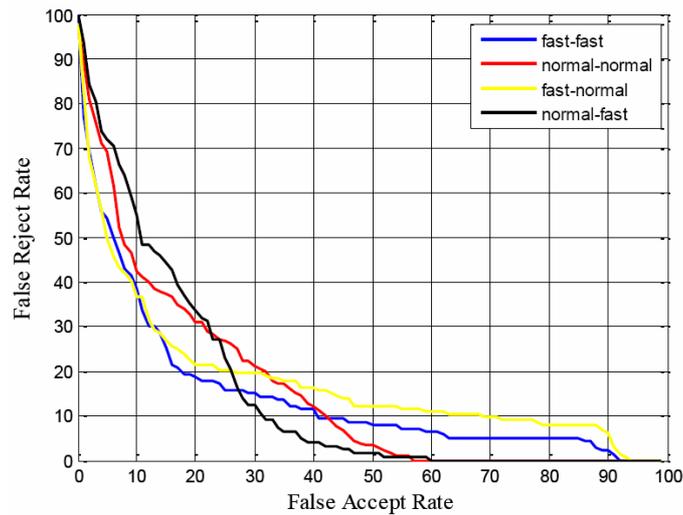
Table 8   Gait recognition results of SOTON database

Method	FT-FT	NM-NM	FT-NM	NM-FT
Phillips et al. (2002)	64.9	78.7	70	53.7
Proposed method (VQ-VQ)	*98*	*100*	*84*	*60*
Proposed method (non-VQ-VQ)	*98*	*100*	*94*	*84*

Figure 11 Verification curves for (a) proposed algorithm for non-VQ vs. VQ condition and (b) baseline algorithm using MANUKAU database
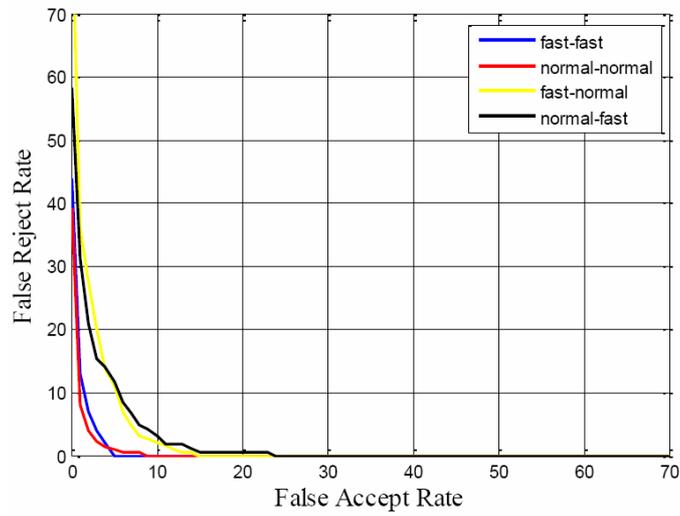
(a)

(b)

4 Conclusions

This paper has proposed a new gait recognition approach using SVQ technique for analysing sequence of width vectors obtained from the silhouette of the walking human. In order achieve the goals, we extracted outer contour width of silhouette and segmented it in to blocks (know as sub-vectors), while maintaining the data dimension and preserving all vital static and dynamic feature information. LBG VQ technique was applied to sub-vectors to obtain the compressed form of gait signature. The probe sequence and reference sequence are quantised using the trained codebooks. DTW was applied to find the match. The proposed approach was applied to three well known databases and on our own image database.

Extensive experiments were carried on five different view angle gait datasets for a large CASIA-B image database. As validated experimentally, the drop in recognition performance is negligible using a simple approach for gait feature extraction when compared to the leave-one-out method. Furthermore, experiments were conducted to measure the performance of the proposed algorithm for different covariate factors such as speed, clothing and bag carrying. The proposed approach achieved a consistent overall recognition and verification performance when tested on the three standard databases and on our own database. This confirms the stability of the algorithm and is very promising. On average across the three standard databases and our own database, the proposed algorithm achieved 99.7% identification rate for same condition testing, 89.8% for different walking speed and 69.6% for different clothing and carrying condition testing.

The work carried out thus far encourages further work to investigate the combination of different view angles score level fusion of results to improve the identification and verification rate. Moreover, further work also involves the implementation of the algorithm in a real-time system. The challenge envisaged is the efficiency of the system in identification of an unknown probe. As such, the system needs to be scalable relative to the size of the database. Different models of the real-time system, such as hardware implementation, and hybrid hardware-software co-design will be investigated.

References

Bashir, K., Xiang, T. et al. (2008) 'Feature selection on gait energy image for human identification', *ICASSP 2008. IEEE International Conference on Acoustics, Speech and Signal Processing*, pp.985–988.

Bouchrika, I. and Nixon, M.S. (2008) 'Gait recognition by dynamic cues', *19th International Conference on Pattern Recognition*, pp.1–4.

Boulgouris, N.V., Plataniotis, K.N. et al. (2004) 'Gait recognition using dynamic time warping', *IEEE 6th Workshop on Multimedia Signal Processing*, pp.263–266.

Chai, Y., Wang, Q. et al. (2006) 'A novel human gait recognition method by segmenting and extracting the region variance feature', *18th International Conference on Pattern Recognition*.

Cutting, J.E. and Kozlowski, L.T. (1977) 'Recognising friends by their walk: gait perception without familiarity cues', *Bulletin of the Psychonomic Society*, Vol. 9, pp.353–356.

Ekinci, M. (2006) 'Gait recognition using multiple projections', *Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition*.

Gross, R. and Shi, J. (2001) *The CMU Motion of Body (MoBo) Database Technical Report*, CMU-RI-TR-01-18, Robotics Institute, Carnegie Mellon University, June.

Han, J. and Bhanu, B. (2006) 'Individual gait recognition using gait energy image', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 28, No. 2, pp.316–322.

Iwashita, Y. and Petrou, M. (2008) 'Person identification from spatio-temporal volumes', *23rd International Conference Image and Vision Computing*, New Zealand, pp.1–6.

Johansson, G. (1973) 'Visual perception of biological motion and a model for its analysis', *Perception & Psychophysics*, Vol. 14, No. 2, pp.202–211.

Kale, A., Cuntoor, N. et al. (2003) 'Gait analysis of human identification', *4th International Conference on Audio-and Video-Based Biometrie Person Authentication*, pp.706–714.

Kusakunniran, W., Wu, Q. et al. (2009) 'Automatic gait recognition using weighted binary pattern on video', *Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance*, pp.49–54.

Linde, Y., Buzo, A. et al. (1980) 'An algorithm for vector quantization', *IEEE Transactions on Communications COM*, January, Vol. 28, pp.84–95.

Mckenna, J., Jabri, S. et al. (2000) 'Tracking interacting people', *Proceedings of the Fourth IEE Conference on Automatic Face and Gesture Recognition*.

Nixon, M.S., Tan, T. et al. (2006) *Human Identification Based on Gait*, Springer, New York, USA, ISBN 0-387-24424-7.

Pandey, N., Abdulla, W. et al. (2005) 'Human identification using vector quantisation technique', *International Symposium on Signal Processing and its Applications Sydney*, Australia, Vol. 2, pp.671–674.

Pandey, N., Abdulla, W. et al. (2007) 'Gait-based person identification using multiview sub-vector quantisation technique', *International Symposium on Signal Processing and its Applications*, Sharjah, United Arab Emirates.

Phillips, P.J., Sarkar, S. et al. (2002) 'The gait identification challenge problem: data sets and baseline algorithm', *Proceedings of the 16th International Conference on Pattern Recognition*.

Sakoe, H. and Chiba, S. (1978) 'Dynamic programming algorithm optimization for spoken word recognition', *IEEE Transactions on Acoustics, Speech, and Signal Processing ASSP*, Vol. 26, No. 1, pp.43–49.

Shutler, J.D., Grant, M.G. et al. (2002) 'On a large sequence-based human gait database', *Proceedings of 4th International Conference on Recent Advances in Soft Computing*, Nottingham, UK.

Ye, B. and Wen, Y. (2006) 'A new gait recognition method based on body contour', *9th International Conference on Control, Automation, Robotics and Vision*, pp.1–6.

Yu, S., Tan, D. et al. (2006) 'A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition', *The 18th International Conference on Pattern Recognition*.