Hierarchical and Multi-featured Fusion for Effective Gait Recognition under Variable Scenarios

Yanmei Chai[1], Jie Ren[2], Huimin Zhao[3], Yang Li[1], Jinchang Ren[4], Paul Murray[4]

[1]School of Information, Central University of Finance and Economics (CUFE), Beijing, China
[2]College of Electronics and Information, Xi'an Polytechnic University, Xi'an, China
[3]School of Electronics and Information, Guangdong Polytechnic Normal University, Guangzhou, China
[4]Centre for excellence in Signal and Image Processing, University of Strathclyde, Glasgow, U.K.

***Abstract*—Human identification by gait analysis has attracted a great deal of interest in the computer vision and forensics communities as an unobtrusive technique that is capable of recognizing humans at range. In recent years, significant progress has been made, and a number of approaches capable of this task have been proposed and developed. Among them, approaches based on single source features are the most popular. However the recognition rate of these methods is often unsatisfactory due to the lack of information contained in single feature sources. Consequently, in this paper, a hierarchal and multi-featured fusion approach is proposed for effective gait recognition. In practice, using more features for fusion does not necessarily mean a better recognition rate and features should in fact be carefully selected such that they are complementary to each other. Here, complementary features are extracted in three groups: *Dynamic Region Area*; *Extension and Space* features; and *2D Stick Figure Model* features. To balance the proportion of features used in fusion a hierarchical feature-level fusion method is proposed. Comprehensive results of applying the proposed techniques to three well-known datasets have demonstrated that our fusion based approach can improve the overall recognition rate when compared to a benchmark algorithm.**

***Index Terms*—Gait recognition, hierarchical and multi-featured fusion, extension and space features, 2D Stick Figure Model, dynamic region area**

I. INTRODUCTION

As a new technology of biometrics, gait recognition has its predominance among others though it is a challenging problem in computer vision. With the development of computer technology, many algorithms which aim to improve gait recognition performance have been proposed in the last 20 years.

The early algorithms often do not account for the mechanics of body and motion. Instead holistic features, such as body shape [1], silhouette or silhouette contour [2], paces [3], or body symmetry [4] etc., are extracted and used for gait recognition. These approaches have the advantage of low computational costs as they do not require the accurate calculation of parameters which can be used to model the motion patterns. However, such techniques generally achieve low recognition rates since the granularity of characteristic expression is too coarse.

In order to refine the granularity of characteristic expression, many model-based approaches have been proposed, such as the pendulum model of the leg [5], the ellipse model [6] and the five-link biped human model [7]. Not only do these model features reflect the dynamic characteristics of the gait, but they are also able to estimate the change trend of the gait since they are built according to the movement characteristics of the human body itself. Furthermore, model-based approaches have the additional advantage of being able to achieve view-invariant and scale-independent recognition. However, these techniques are sensitive to the quality of gait sequences captured and they suffer from high computational costs associated with determining the model parameters.

The aforementioned approaches all suffer, in our opinion, from a common shortcoming in that only part of the gait motion information that is actually available, is used, i.e. either holistic shape features or model parameters. The human vision perception system does not rely on single source gait features for person identification as there are many properties of the person that might serve as alternatives. Therefore, it is thought that an approach based on fusing multi-source information can overcome the limitations of these existing techniques and obtain better performance.

One approach which might be attractive is to combine the results from multiple biometric characteristics, such as gait, face and ear shape to achieve recognition. However, it is quite difficult to collect these characteristics in the same sequence. Thus, fusing different types of gait features becomes more desirable. Shape-based features and model-based features are used as fusion features in most existing approaches [8-11]. However, the granularity of shape-based features tends to be too coarse and that of model-based features too refined. As a result, the integration of these features does not yield the best results.

In this paper, a novel fusion based approach is proposed for effective gait recognition. The major contributions of the paper are summarized as follows:

- Three different types and granularities of features which complement each other are extracted and fused, to achieve effective gait recognition;
- A hierarchical feature-level fusion method is proposed to balance the proportion of features in fusion;
- Performance evaluations from various different aspects are adopted to verify the rationality and validity of the algorithm.

In the proposed technique, ***Dynamic Region Area*** features are incorporated into a fusion algorithm as a medium granularity feature, together with ***Extension*** and ***Space*** features and ***2D Stick Figure Model*** features which are used for gait recognition. Comprehensive results on the UCSD, the CMU and the CASIA databases have fully demonstrated the improved performance of our proposed approach with respect to several different variations such as the size of databases, walking speed and outdoor/indoor conditions.

The remainder of this paper is organized as follows. Section II provides details of how the three groups of features are extracted from the data. The proposed hierarchical and multi-featured fusion strategy is discussed in Section III and comprehensive results and evaluations are then presented in Section IV. This is followed by concluding remarks which are drawn in Section V.

II. EXTRACTING GAIT FEATURES

Three types of feature which complement each other are extracted by our algorithm to allow effective gait recognition. These are:

- ***Extension*** and ***Space*** features which describe the shape characteristic of a walking object;
- ***2D Stick Figure Model*** features which describe the trajectory-based joint kinematics characteristics;
- ***Dynamic Region Area*** features which describe the motion characteristics of the body parts.

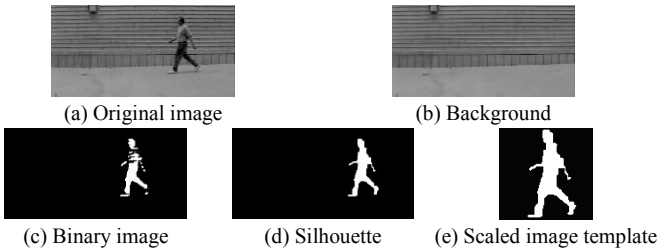
Details on pre-processing and extraction of these features are described below.

A. Preprocessing

Silhouette extraction is a necessary first step to extract the required human body by eliminating irrelevant background from each image. The detailed steps are described as follows.

1) Determine a background image using a running average over the entire image sequence (see in Fig.1b).
2) Detect moving objects in each frame by background subtraction using the determined background image (see in Fig.1c).
3) Apply morphological opening and closing (erosion and dilation) followed by component labeling to fill holes and also remove noise in the detected foreground object, i.e. the silhouette (see in Fig.1d).

4) Normalize the silhouette into a scaled template to eliminate redundancies and speed up the processing (see in Fig.1e).



(a) Original image (b) Background

(c) Binary image (d) Silhouette (e) Scaled image template

Fig.1 the preprocessing of gait sequence

From the theoretical perspective, the human gait is a form of periodic motion. This is especially true when a person who is walking is observed from a lateral viewpoint. In this situation, we can estimate the gait periodicity over time by counting the number of foreground pixels in the bottom half of the silhouette in each frame [12]. The number will reach a maximum when the two legs are farthest apart (full stride stance) and drop to a minimum when the legs overlap (heels together stance). To demonstrate this, Fig. 2 shows an instance of a sequence's pixel numbers curve which has been calculated as described and smoothed by Gaussian filtering. Notice that two consecutive strides constitute a gait cycle.

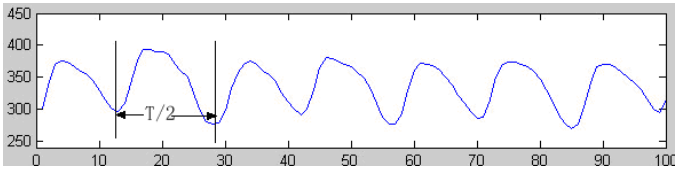


Fig.2 the smoothed curve of pixel number of a gait sequence

B. Extracting the shape features

The width and height of the body are important cues for identifying individuals in the human perception system [13]. However, these quantities always vary with the focus distance of the camera and, as a result, it is not advisable to use them as gait features directly. Instead, we use the ***Extension*** feature E which is a comparatively stable alternative to represent this property of the gait.

$$E = \frac{w}{h} \tag{1}$$

where w is the width of silhouette and h is its height.

While this feature is robust to changes in the focus distance of the camera, the ***Extension*** feature does eliminate some useful information by dividing. For example if a person is tall and fat and another is short and slim, the ***Extension*** of both people may be the same or at least very similar [14]. To overcome this potential shortcoming, we also compute a ***Space*** feature to represent the holistic shape of person. The ***Space,*** S , is obtained by counting the number of foreground pixels in the silhouette in each frame.

$$S = \sum_{i=1}^{n}\sum_{j=1}^{m} P(i,j), \qquad P(i,j) \in (0,1) \tag{2}$$

where, $P(i,j)$ is the pixel value located the i^{th} row and j^{th} column in the silhouette image.

It is important to note that the ***Extension*** and the ***Space*** represent respectively the physical characteristics of an object in images from two perspectives which are grouped together to form the Shape feature of the detected human body. Furthermore, to eliminate the influence of spatial scale, we normalize these features to a uniform magnitude which is consistent with the kinematic parameters of the gait (see II(C)).

C. Extracting kinematic features

The basic idea of a ***2D Stick Figure Model*** was first introduced in [14]. In our work, we use the body segment properties guided by anatomical knowledge to determine the position of the main parts of the body including the: head, neck, shoulder, pelvis, kneel and ankle. When the position of these body parts has been estimated, we can compute kinematic parameters of the gait. Please note that the upper limbs are ignored in our experiment due to the occlusion of these body parts in the side-view sequence. The steps for establishing the ***2D Stick Figure Model*** and extracting the kinematic features are described are as follows.

1) Extract the skeleton from each silhouette using the Medial Axis Transformation algorithm [15]. This is a reversible transformation which has the effect of compressing the image while retaining enough information to allow it to be accurately restored from the central axis and its values.

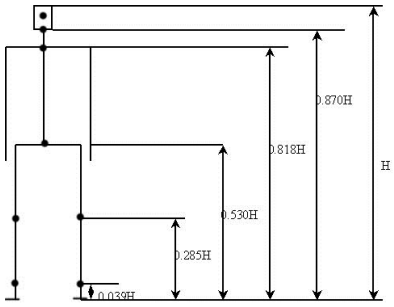
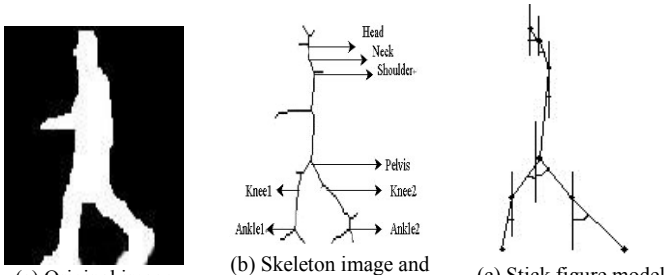


Fig.3 Body segment properties

2) Scan the image skeleton row by row from top to bottom according to the body segment properties (see in Fig. 3). The junction of the scan line and skeleton is defined as joint points. There are 8 coordinates (joint points) in a human body, which are (x_{head}, y_{head}) , (x_{neck}, y_{neck}) , $(x_{shoulder}, y_{shoulder})$, (x_{pelvis}, y_{pelvis}) , (x_{knee1}, y_{knee1}) , (x_{knee2}, y_{knee2}) , (x_{ankle1}, y_{ankle1}) and (x_{ankle2}, y_{ankle2}) . The skeleton image and joint points are shown in Fig. 4 (b). When the positions of all joint points have been computed from the image, their coordinates are connected in order to form a ***2D Stick Figure Model***.



(a) Original image (b) Skeleton image and joint points (c) Stick figure model and limb angles

Fig.4 Joint positions and limb angles in the gait

3) Calculate the angles between the main limbs and a vertical normal line (illustrated in Fig 4(c)) using the following formula:

$$\theta = \tan^{-1} \frac{x_1 - x_0}{y_1 - y_0} \tag{4}$$

where, (x_0, y_0) and (x_1, y_1) are the coordinates of two connected joints. Moreover, there are 7 angles associated with these positions, including θ_{head} , θ_{neck} , θ_{back} , θ_{thigh1} , θ_{thigh2} , θ_{shin1} and θ_{shin2} . The ***Stick Figure Model*** and the definition of limb angles are show in Fig. 4(c).

4) Obtain the kinematic parameters of the gait. According to the calculations described in steps 2) and 3), there are 23 kinematic parameters in the human stick figure model, where 2x8=16 parameters are for joint coordinates and the remaining 7 are for the limb angles. As the x values of joint coordinates are usually fixed, they can be ignored such that the overall number of parameters is reduced to 15. Furthermore, to eliminate the influence of spatial scale, we normalize all parameters to a uniform magnitude in the range $[\pi/2, 3\pi/2]$.

D. Extracting dynamic region feature

It has been shown that the shape features focus on the holistic object, and their granularity is coarse. It has also been shown that the kinematic features focus on the joint positions, and their granularity is comparatively more refined. We now describe ***Dynamic Region*** features which take into account the movement of body parts thus allowing more comprehensive gait information to be obtained.

In practice, one person's body parts will move differently from another's when they are walking. For example, some peoples' heads may have only a slight movement while others movements may be more exaggerated and more frequent. Some peoples' torsos will remain almost still when walking while others will oscillate severely. Furthermore, the oscillation of different peoples' legs will likely differ too. In order to extract the region feature, we therefore divide the two dimensional silhouette of the walker into three regions, namely the: head region, trunk region and leg region as shown in Fig. 5.

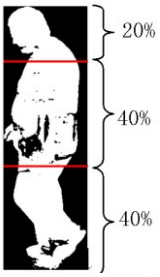


Fig.5 Region segmentation of parts of body

Following this partitioning of the image, the *area* feature is extracted from each region respectively, and the sequence of features from one region over the entire sequence is used to characterize the salient motion of the corresponding region. Let k be the number of regions, R_1, R_2 and R_3 represent the head region, trunk region and legs region respectively, and let $f(i,j)$ denote the binary value at position (i,j). The area feature can then be calculated as [16]:

$$A_k = \sum_{(i,j)\in R_k} f(i,j), \qquad \text{where} \tag{5}$$

$$f(i,j) = \begin{cases} 0 & (i,j) \in R^k_{background} \\ 1 & (i,j) \in R^k_{object} \end{cases} \tag{6}$$

III. HIERARCHICAL AND MULTI-FEATURED FUSION

Since the shape features, model-based features and dynamic region features describe the characteristics of the gait from different granularity, it is anticipated that fusion of these features may yield tangible benefits. In this paper, two different fusion strategies are employed: feature-level fusion and decision-level fusion. In feature-level fusion, multiple features are grouped together to form a multivariate single feature known as a combined gait signature. Pattern training and classification is then carried out using the combined signature for recognition purposes. In decision-level fusion, multiple sources of features are regarded as independent gait patterns and these are used separately for training and classification. Fusion is then carried out using the classified results from each individual feature to obtain the final results by employing either Sum rules or Product rules [17]. These rules are frequently used in decision-level fusion strategies where the final results are calculated by summing (or multiplying) the scores acquired from multi-source information, respectively. It is worth noting that the fused features (or decision scores) should be normalized into the same interval, respectively, for consistency.

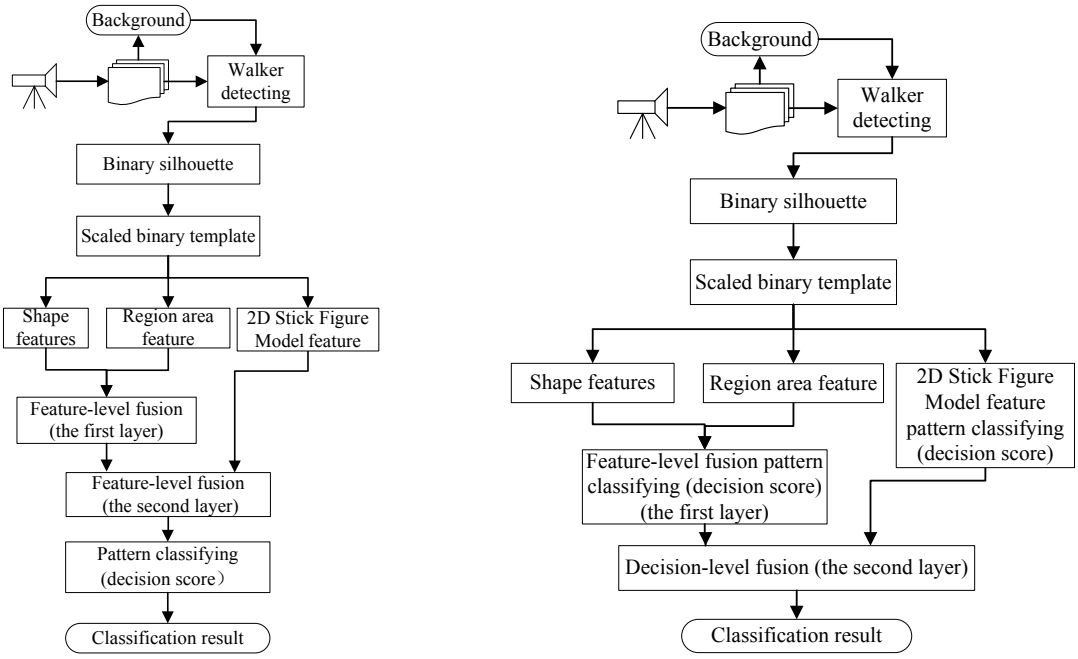


Fig.6 Diagram of two different fusion strategies: Fusion strategy #1 (left) and Fusion strategy #2 (right).

Since the dimension of the shape and region features is much fewer than that of the 2D Stick Figure Model features, a hierarchal fusion approach is adopted to balance the proportion of features in fusion. To achieve this, the shape feature and region features are firstly fused using the feature-level fusion. Then the combined features are fused with the 2D Stick Figure Model features using both feature-level fusion and decision-level fusion, respectively. Flow charts which clearly show these two fusion strategies and the differences between them are given in Fig. 6.

In this paper, the classification process is carried out using the Nearest Neighbor classifier (NN). Considering the gait's periodicity, a spatio-temporal similarity measurement based on the gait cycles is adopted. The detailed steps of our approach are described as follows.

1) Let $\mathbf{x}_g = \{\mathbf{x}_{g,1}, \mathbf{x}_{g,2}, \cdots$ be a sequence in the gallery, and $\mathbf{x}_p = \{\mathbf{x}_{p,1}, \mathbf{x}_{p,2}, \cdots$ be an arbitrary one in the probe, where m and n are the numbers of the frames in the two sequences, respectively. $\mathbf{x}_{i,j}$ is the gait feature vector from the j^{th} frame of the i^{th} sequence.

2) Calculate the period of an arbitrary gait sequence $\mathbf{X} = \{\mathbf{X}_1, \mathbf{X}_2, \cdots \mathbf{X}$, and N is the length of the sequence. Then we can partition the whole sequence into $\lfloor N / N_p \rfloor$ subsequence, where N_p ($N_p < N$) is the gait period of the sequence. The k^{th} subsequence is denoted by $\mathbf{X}(k) = \{\mathbf{X}_{k+1}, \mathbf{X}_{k+2}, \cdots, \mathbf{X}_{k+N_p}\}$.

3) Determine the distance between a subsequence that begins from the j^{th} frame in the gallery and the k^{th} one in the probe by using

$$dis_{(\mathbf{X}_p(k), \mathbf{X}_g)}(l) = \sum_{j=1}^{N} \left\| \mathbf{X}_{p,k+j} - \mathbf{X}_{g,l+j} \right\| \tag{7}$$

4) Obtaining the similarity of two whole sequences is defined as

$$Sim(\mathbf{X}_p, \mathbf{X}_g) = 1 - \frac{1}{K} \sum_{k=1}^{K} \min_{l} (dis_{(\mathbf{X}_p(k), \mathbf{X}_g)}(l)) \tag{8}$$

where, $K = \lfloor n / N_p \rfloor$ and l ($l = 0, 1, \cdots m \quad N_p$) is the starting frame in the gallery sequence, from which we compare the two subsequences. The greater the similarity value, the more similar the two sequences.

IV. Experimental Results and Discussions

In this section, we analyze the performance of our proposed algorithm when applied to sequences from 3 different databases. Our experiments aim to find out how well our method performs with respect to several different variations such as the size of databases, the speed of walking, etc. We have also considered data captured in both indoor and outdoor environments for our analysis.

A. Datasets and Experimental Settings

1) **The University of California, San Diego (UCSD)** database - all videos in this database were captured in an outdoor environment where the distance between the camera and subjects is comparatively far. There are 6 subjects in the database, each subject has 7 sequences, and each sequence covers 2-3 gait cycles. The dimensions of the original images are 320×160 (pixels) which were normalized into a 64×64 image template for efficiency in our experiments.

2) **The Carnegie Mellon University (CMU)** database - all videos were taken on an indoor treadmill where the distance between the camera and subjects is comparatively small. This database has 25 subjects walking at a fast pace (about 55 gait cycles per minutes) and slow pace (about 48 gait cycles per minutes), respectively, and there are around 7-8 gait cycles in each sequence. The original images in this dataset are 640×480 (pixels) and these too were normalized into a 64×64 image template for the reasons stated above.

3) **Chinese Academy of Sciences (CASIA) Dataset B** - videos in this database were acquired outdoors and the distance between the camera and subjects is comparatively far. There are 124 subjects in this dataset, and the gait data is captured from 11 different viewpoints. Three variations, namely view angle, clothing and changes in carrying condition, are considered separately. For the purposes of our experiment, 107 subjects with view angle "090" were selected from the database and, as with the other datasets, the selected images were normalized into a 64×64 image template.

Using the three databases described above, a total of 10 different experiments were designed and these are described in Table 1. In the first three experiments, a leave-one-out cross-validation rule is used for validation purposes. For the UCSD database which

contains 42 normal-walk sequences (6 subjects and 7 sequences for each subject), we leave one example out as the probe, and the rest constitutes the gallery. This process is repeated 42 times, and the recognition rate is obtained as the number of correctly classified test examples out of all 42 tests (experiment #1). A similar process is adopted for the CASIA database where a total of 624 leave-one-out cross-validations are executed (experiment #4). For the CMU database which contains 8 gait cycles in each Fast Walk sequence and 7 cycles in each Slow Walk sequence the process is a little different. Here, each gait cycle is considered as a subsequence. The leave-one-out process is then applied by leaving out one subsequence as the probe and allowing the rest of them (6 for the Slow Walk or 7 for the Fast Walk respectively) to form the gallery (experiments #2-#3).

TABLE 1.
DESIGNED EIGHT GROUPS OF EXPERIMENTS IN OUR SYSTEM

Database	No.	The gallery sequence	The probe sequence
UCSD	#1	All others (41) as gallery sequences	Leave one as probe sequence (normal walk)
CMU	#2	All other gait cycles	Each gait cycle of fast walking sequence
	#3	All other gait cycles	Each gait cycle of slowing walking sequence
CASIA	#4	All others as gallery sequences	Leave one as probe sequence (normal walk)
UCSD	#5	6 sequences	The rest 36 sequences
CMU	#6	25 fast-walk subsequences (Fast)	The rest 175 fast-walk subsequences (Fast)
	#7	25 slow-walk subsequences (Slow)	The rest 150 slow-walk subsequences (Slow)
	#8	25 fast-walk entire sequences (Fast)	25 slow-walk entire sequence (Slow)
	#9	25 slow-walk entire sequences (Slow)	25 fast-walk entire sequence (Fast)
CASIA	#10	107 sequences	The rest 535 sequences

Experiment #5 is designed for the UCSD database where 6 sequences are used as gallery and the remaining 36 are used as probe sequences. Experiments #6-#9 are for the CMU database. In experiments #6-#7, 25 subsequences which constitute a single gait cycle are used for training and the remaining subsequences (175 for Fast Walk and 150 for Slow Walk) are used as testing samples. Then, in experiments #8-#9, we use the entire original sequences with all the gait cycles in order to evaluate the influence of speed to our algorithm. This is achieved by training on fast-walk sequences and testing on slow-walk sequences in Experiment #8, and vice versa in Experiment #9. Experiment #10 is designed for CASIA database, where 107 sequences make up the gallery and the remaining 535 are used as probe sequences.

B. Evaluation Criteria and Results

Based on the 10 groups of experiments which we have designed and described above, three criteria are used for performance evaluation. The first is the classification rate or recognition rate which is quantified using the Correct Classification Rates (CCR) and the Cumulative Match Score (CMS). The second uses Receiver Operating Characteristic (ROC) analysis to evaluate performance in terms of correct recognition rate and false-alarm rate. The third is a comparison of the proposed technique with a benchmarking algorithm [19] . The results from these designed experiments and their evaluation are presented and compared below.

1) Performance of CCR and CMS

The CMS curve proposed in the FERET protocol [18] indicates the probability of the correct match being included in the top n matches. The size of the gallery is the number of different people contained in the gallery and there is one sequence per person in the gallery. When computing CMS, the horizontal axis of the graph corresponds to rank and the vertical axis is the accumulated probability $p(i)$ of the identification. Obviously, a result of $p(1)$ is equivalent to the correct classification rates (CCR) of the NN classifier. For simplicity, CCR is applied to the experiments #1-#4 whilst plots of CMS are used to evaluate experiments #5-#10.

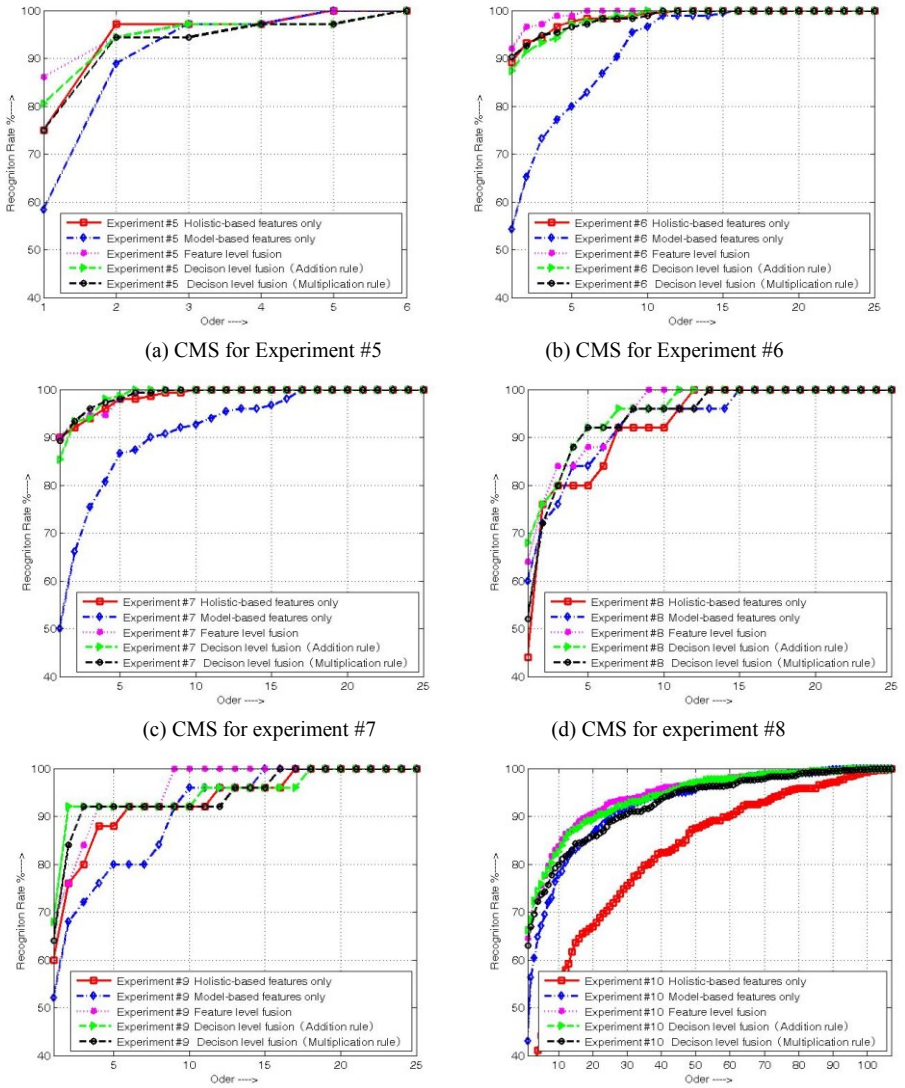
As explained previously, in experiments #1-#4, leave-one-out cross-validation is utilized. For these experiments, a number of different feature sets were extracted from each dataset, including: shape features, region features, holistic features (shape plus region features), model-based features, feature-level fusion and decision-level feature (including the Sum and Product rules), respectively. The CCR values, as defined previously in this section, were computed for each experiment with results summarized in Table 2.

From Table 2 it is clear that, on average, the best identification performance using leave-one-out criterion is obtained by using the feature-level fusion algorithm. This is closely followed in terms of performance by the approach using decision-level fusion with the product rule and the sum rule. On further observation, we can also notice that the size of the dataset affects the CCR of the fusion

algorithms. That is, a greater improvement can be seen in the results obtained using fusion algorithms over the other approaches when applied to the larger datasets.

TABLE 2.
CCR OF DIFFERENT ALGORITHMS UNDER LEAVE-ONE-OUT SCHEME WITH THE NN CLASSIFIER

Database		UCSD	CMU		CASIA
No.		#1	#2	#3	#4
Walking speed		Normal	Fast	Slow	Normal
Shape features		80.95	91.0	89.71	40.03
Region features		95.24	98.50	**100.0**	51.62
Holistic features (Shape + region)		**97.62**	98.50	**100.0**	56.85
Model-based features		83.33	72.50	73.71	66.51
Feature-level fusion		**97.62**	**99.0**	**100.0**	78.35
Decision -level fusion	Sum rule	95.24	97.0	98.29	78.51
	Product rule	95.24	**99.0**	99.43	**78.97**



(a) CMS for Experiment #5 (b) CMS for Experiment #6

(c) CMS for experiment #7 (d) CMS for experiment #8

(e) CMS for experiment #9 (f) CMS for experiment #10

Fig. 7 CMS performances from Experiments #5-#10

As the holistic features have produced much better results than those obtained by using Shape features or Region features in isolation, only the holistic features will be used for comparison in the remaining experiments. For experiments #5-#10, the CMS curves which have been calculated are shown in Figure 7. As it is not straightforward to directly compare performance

using these plots, the average CMS values of rank 1-5 were computed and these are given in Table 3. As in Experiments #1-#4 it is clear that the fusion algorithms significantly outperform the algorithms which exploit only the holistic features or model-based features in isolation. The advantages of using the fusion approaches are very clear, especially when applied to a very large database.

If we look further into these results, several interesting findings can be made and these are summarized as follows. Firstly, the fused results from different strategies are always the best in the table when compared to those without fusion. Secondly, feature level fusion produces better results in simple cases when the walking patterns are consistent in the sequences (Experiments #1-#3, #5, #6). However, when the walking speeds in the training and testing sequences are different (Experiments #8, #9), decision level fusion with the sum rule generates the best results. Thirdly, the results obtained by using only holistic features are better than those obtained using only model-based features to process the UCSD and CMU datasets (Experiments #1-#3 and #5-#9). Finally, for the CASIA dataset, the results are completely different and in fact using only the model-based features yields better results than if only the holistic features are used. This is probably due to large variations among shape features extracted from the data, especially when the number of subjects is small.

It should be noted that there are some inconsistences between the results in Table 2 and those in Table 3. That is, CCR (Table 2) suggests feature level fusion is the best option whilst CMS (Table 3) recommends that decision level fusion with the sum rule should be used. One possible reason for this is over-fitting in training since only one testing sample was used when applying the leave-one-out cross-validation rule in Experiments #1 - #4. As such, the single test sample used in these experiments can always find consistent patterns. However, this situation does not reflect real world situations where such high consistency of features is void, especially when the walking speeds for training and testing sequences are different or the subjects are very large in number. In other words, it appears that CCR with leave-one-out strategy is a less suitable for performance evaluation for large datasets.

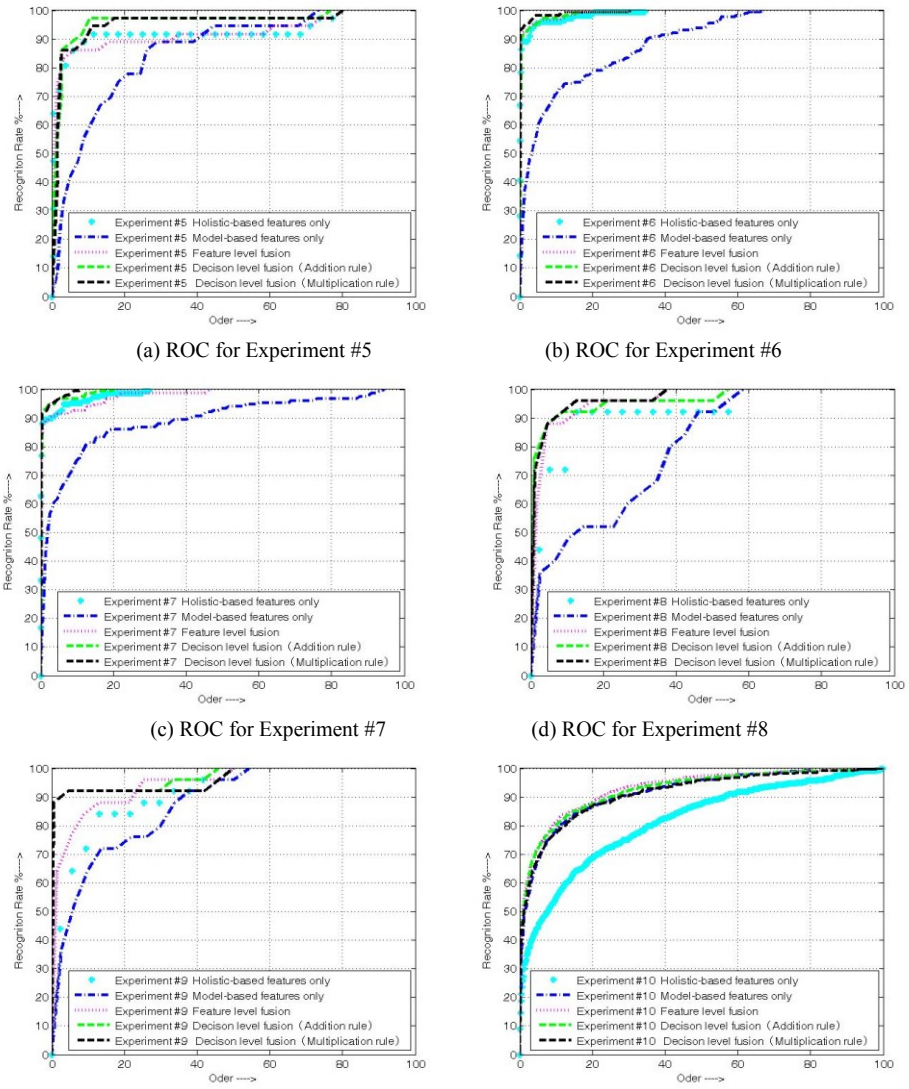
TABLE 3.
PERFORMANCE FROM EXPERIMENTS #4-#8 USING NORMALIZED CMS CRITERION

Database	No	Holistic features	Model-based features	Feature-level fusion	Decision-level fusion		Speed in training & testing sequences
					Sum rule	Product rule	
UCSD	#5	0.933	0.883	**0.949**	0.933	0.916	Normal/normal
CMU	#6	0.942	0.699	**0.967**	0.927	0.939	Fast/fast
	#7	0.94	0.717	0.941	0.939	**0.949**	Slow/slow
	#8	0.72	0.752	0.792	**0.808**	0.768	Fast/slow
	#9	0.784	0.696	0.824	**0.872**	0.848	Slow/fast
CASIA	#10	0.359	0.583	0.707	**0.714**	0.690	Normal/normal
Average		**0.779**	**0.721**	**0.863**	**0.865**	**0.851**	

2) Performance of ROC analysis

Receiver Operating Characteristic (ROC) analysis can be used to assess the trade-off between the probability of correct identification and the probability of false alarm. Correct identification occurs when the algorithm correctly reports the existence of a probe which is in the gallery. Conversely, a false alarm occurs when the algorithm reports that a probe is in the gallery when in fact it is not. In the false-alarm test, there are two primary categories of probes. The first are probes which are not in the gallery that generate false alarms. The false-alarm rate is the percentage of probes not in the gallery that are falsely reported as being in the gallery and is denoted by P_f. The second category of probes is the set that is in the gallery. This set characterized by the percentage of these probes that are correctly identified and is denoted by P_i. The pair of values P_f and P_i describe the operation of a system in an open universe in which not every probe is in the gallery. For an algorithm, performance is not characterized by a single pair of statistics (P_f, P_i) but rather by all pairs (P_f, P_i), and this set of values is an ROC.

For Experiments #5-#10, the corresponding ROC curves are plotted in Fig. 8, respectively. As can be seen, decision level fusion with the product rule and sum rule are found to generate the best results in these plots. These techniques are followed by those which use holistic features in experiments #5-#7 and feature level fusion in experiments #8-#10. Again, the shape features have a higher degree of differentiation when the subjects are fewer in number and there is a small change in speed between the training and testing sequences. On the other hand, the feature level fusion approach can obtain better results when applied to large databases and can perform well when processing data from highly dynamic environments and changing scenarios.



(a) ROC for Experiment #5 (b) ROC for Experiment #6

(c) ROC for Experiment #7 (d) ROC for Experiment #8

(e) ROC for Experiment #9 (f) ROC for Experiment #10

Fig. 8 Performance of ROC analysis

3) Performance comparison with the baseline algorithm

We now compare our approach with the baseline algorithm proposed in [19]. The baseline algorithm draws upon the success of template-based recognition algorithms in computer vision and has become the "benchmark" in evaluating gait recognition techniques. The method proposed in [19] estimates silhouettes by background subtraction and performs recognition using temporal correlation of silhouettes. In fact, the baseline algorithm has fairly good recognition performance when all of the original pixels are used. For example, in experiment #D of the HumanID Gait Challenge problem, several well-known algorithms were used to analyze a database containing 71 subjects. The identification rate of continuous HMM based recognition [20] is 36%, body shape [21] is 21%, body part based recognition [22] is 25%, and the baseline algorithm is 29%. While the performance of the baseline algorithm is good, it should be noted that the computational complexity and time consumption of the technique are considerably high.

The baseline algorithm is used as a standard comparison algorithm in many papers, and some recent examples published in 2011 can be found in [23]. Using the same experimental processes described in Section IV A, and using the same databases, we carried out the above experiments to evaluate the performance of the baseline algorithm for the purposes of comparison. The results from experiments #1-#10 are shown in the first two columns of Table 4. Furthermore, in order to compare our approach with the baseline algorithm, we computed the CMS value of rank 1-5 and obtained the average CCR of them for experiment #5-#10.

TABLE 4.
COMPARING THE PERFORMANCE OF OUR APPROACHES WITH THE BASELINE ALGORITHM

No.	Base-line	Feature-level fusion		Decision-level fusion			
				Sum rule		Product rule	
	Avg. CCR	Avg. CCR	Gains/ loss	Avg. CCR	Gains / loss	Avg. CCR	Gains / loss
#1	90.48	97.62	↑7.14	95.24	↑4.76	95.24	↑4.76
#2	96.50	99.0	↑2.50	97.0	↑0.5	99.0	↑2.5
#3	98.29	100.0	↑1.71	98.29	0.0	99.43	↑1.14
#4	80.99	78.35	↓ 2.64	78.51	↓ 2.48	78.97	↓ 2.02
#5	92.78	94.44	↑1.66	93.33	↑0.55	91.66	↓ 1.12
#6	84.69	94.97	↑10.28	92.68	↑7.99	93.94	↑9.25
#7	84.27	93.47	↑9.20	93.87	↑9.6	94.80	↑10.53
#8	79.2	72.0	↓ 7.20	80.80	↑1.6	76.8	↓ 2.4
#9	84.0	75.2	↓ 8.80	87.20	↑3.2	84.8	↑0.8
#10	61.57	70.73	↑9.16	71.44	↑9.87	69.05	↑7.48
Avg	85.28	87.58	↑2.3	88.84	↑3.56	88.37	↑3.09

Note that the results in Table 4 are actually from 10 different experimental settings (see details in Table 1). The reason we put them together is for better comparison, especially to show how consistently our proposed approach perform under different experimental settings, i.e. across a diverse range of data sets.

In Experiments #1 to #4, leave-one-out strategy was used, i.e. the probe set contains only one sample sequence. We can see the results from Experiment #4 on the CASIA dataset are actually the worst among these four groups of experiments. The reason behind this is that the CASIA dataset is the largest, which contains over 640 sequences. When the size of the database becomes large, there is a higher chance than a sequence can be mismatched when using the leave-one-out strategy due to the similarity that exists between gait sequences in dataset itself. Although feature-level fusion and decision-level fusion works well for the other three groups of experiments and produces an improved CCR, the results for experiment #4 are unsatisfactory. This has clearly shown that leave-one-out is not a good strategy for performance assessment, especially for large datasets.

In Experiments #5-#10, the probe set contains multiple sequences and there is no situation where size of the probe sequence is smaller than the gallery set. In all the 6 groups of experiments, decision-level fusion using sum-rule consistently yields improved average CCR, where the greatest improvement can be found in Experiments #7 and #10. In addition, the results from feature level fusion and decision-level fusion using product rule generates inconsistent results, though on average they can still improve the CCR among the 10 group of experiments performed.

In summary, the experiments shown in Table 4 can convince us of the following findings: Firstly, leave-one-out is not a good strategy for CCR based performance assessment, especially for large datasets. Secondly, fusion based approach can generally improve the overall performance. Thirdly, decision level fusion seems to outperform feature level fusion, especially when the sum rule is used. In fact, this approach consistently produces improved results in all experiments from #5-#10, which has demonstrated the added value of the proposed work.

V. CONCLUSIONS

Most existing gait recognition algorithms are based on single source features and, in general, their recognition rates are unsatisfactory due to the lack of information used. In this paper, we propose a novel hierarchal and multi-featured fusion approach to improve the performance of gait recognition. For our algorithm, three types of feature are extracted: ***Extension*** and ***Space*** features, ***2D Stick Figure Model*** features and ***Dynamic Region Area*** features. These features are found to complement to each other well in the fusion schemes thus allowing a better recognition rate than when using these features in isolation. Using the UCSD, CMU and CASIA databases which contain a wide range of samples exhibiting different walking speeds, movement patterns, and different indoor/outdoor scenarios, our proposed hierarchal and multi-featured fusion strategy has been fully validated. Furthermore, three different criteria have been used for evaluation, including CCR/CMS, ROC and a detailed comparison with the baseline algorithm is provided - this is the benchmark gait analysis technique making this a highly appropriate comparison.

The main findings of our experiments are summarized as follows. Firstly, the fused results from different strategies are always the best in comparison with those without fusion. Specifically, feature level fusion produces better results in simple cases when the subjects are few in number and the walking patterns are consistent. Otherwise, decision level fusion will generate the best results, especially when the walking speeds in training and testing sequences are different. Secondly, leave-one-out strategy is unsuitable for performance evaluation in large datasets as it tends to yield the over-fitting problem. This causes inconsistences in the results when

compared with cases in which fewer training samples are used. Thirdly, the shape features allow a higher degree of differentiation when the subjects are fewer in number and there is little difference in the speed of movement between training and testing sequences. On the contrary, feature level fusion can obtain better results in large databases or in situations where the environment is dynamic and changes significantly. Fourthly, model-based features are found to yield to the worst results in these experiments. This is most likely due to the large variations within the gait data used. Finally, when compared with the baseline algorithm, the decision-level fusion approach which uses the sum rule generates improved results in all the experiments with an average gain of CCR over 3.56%. The other fusion strategies also produce better results in these experiments too, but their overall average improvement slightly less with and average gain of 2.3 for feature level fusion and 3.09 for decision-level fusion using the product rule.

Future work will focus on automatic selection of features for fusion, where feature collision needs be dealt with to improve the complementariness of features by using new techniques such as dynamic variance features [24]. It is also worth extending the work to weighted features to cope with more complex scenarios such as changes in viewing angles and carrying conditions for forensics applications [25][26][27]. In addition, some new concepts and approaches such as more accurate motion-based modelling [28-30], more effective classifier design using support vector machines (SVMs) and artificial neural networks (ANNs) [31], component-based segmentation [32], adaptive clustering [33] and confidence-based analysis [34] will also be investigated.

VI. Acknowledgements

This paper is partially supported by the Research Innovation Fund, Education Department of Shaanxi Province and National Natural Science Foundation of China (61272381). The authors also wish to thank the editors and anonymous reviewers for their valuable and constructive comments in further improving the quality of this paper.

References

[1] Das Choudhury S, Tjahjadi T. Gait recognition based on shape and motion analysis of silhouette contours. *Computer Vision and Image Understanding*, 117(12): 1770-1785, 2013.

[2] F. Dadashi, B.N. Araabi, H. Soltanian-Zadeh. Gait recognition using Wavelet packet silhouette representation and transductive support vector machines. in *Proc. 2nd Int. Congress on Image and Signal Processing*, 2009, pp. 1-5

[3] Ichino M, Kasahara H, Yoshii H, et al. A study on gait recognition using LPC cepstrum for mobile terminal, in Proc. *IEEE/ACIS 12th Int. Conf. on Computer and Information Science*. 2013: 11-16.

[4] J. B. Hayfron-Acquah, M. S. Nixon and J. N. Carter. Automatic gait recognition by symmetry analysis, *Pattern Recognition Letters*, 24(13): 2175-2183, 2003

[5] D. Cunado, M.S. Nixon, and J. N. Carter, Automatic extraction and description of human gait models for recognition purposes, *Computer Vision and Image Understanding*, 90(1): 1–41, Apr. 2003.

[6] L. Lee and W. E. L. Grimson, Gait analysis for recognition and classification, In: *Proc. 5th IEEE Int. Conf. Automatic Face and Gesture Recognition*, pp. 155–162 2002.

[7] R. Zhang, C. Vogler and D. Metaxas. Human gait recognition at sagittal plane, *Image and Vision Computing*, 25(3): 321-330, Mar. 2007.

[8] H. Zheng and K. Zhang. Decision Fusion Gait Recognition Based on Bayesian Rule and Support Vector Machine. *Applied Mechanics and Materials*, 411: 1287-1290, 2013

[9] G. Yu, C. Li, Y. Hu. Gait recognition under carrying condition: a static dynamic fusion method. *In Proc. of the Int. Society for Optical Engineering*. V 8436, 2012

[10] Y. Fu, L. Cao, G. Guo, and T. Huang. Multiple feature fusion by subspace learning. In Proc. *ACM Conf. on Content-based Image and Video Retrieval*, pp. 127-134, 2008

[11] W. Liu, and D. Tao. Multiview Hessian regularization for image annotation. *IEEE Trans. on Image Processing*, 22(7): 2676-2687, 2013

[12] Z. Liu and S. Sarkar. Simplest representation yet for gait recognition: averaged silhouette. *In Proc. of the 17th Int. Conf. on Pattern Recognition (ICPR)*. 2004, 4: 211 – 214

[13] J. Jiang, F. Chen and G. Zhang. Fusion of multi-region features for gait recognition. *Computer Engineering and Applications*, 2011, 47(7): 159-161.

[14] Y. Chai, Q. Wang, J. Jia and R. Zhao. A novel gait recognition method via fusing shape and kinematics features, *Lecture Notes in Computer Scienc*e, V4291, pp. 80-89, Lake Tahoe, NV, USA, 2006

[15] J. H. Yoo and M. S. Nixon. On laboratory gait analysis via vomputer vision. *In Proc. Int. Sympo. on Biologically-Inspired Machine Vision, Theory and Application. Aberystwyth,* UK, 2003, 109-113

[16] Y. Chai, Q. Wang, J. Jia and R Zhao. A novel human gait recognition method by segmenting and extracting the region variance feature, In Proc. *18th Int. Conf. on Pattern Recognition,* 425-428, Hong Kong, 2006

[17] N. Cuntoor, A. Kale and R. Chellappa. Combining multiple evidences for gait recognition. In Proc. *the IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, 2003, 3: III - 33-6

[18] P. J. Phillips, H. Moon, S. A. Rizvi and P. J. Rause, The FERET evaluation methodology for face recognition algorithms, *IEEE Trans Pattern Analysis and Machine Intelligence*, 22(10): 1090-1104, 2000.

[19] S. Sarkar, P. J. Phillips, Z. Liu, I. R. Vega, P. Grother and K.W. Bowyer. The humanID gait challenge problem: data sets, performance, and analysis, *IEEE Transactions on Pattern Analysis and Machine Intelligence. 27(2):* 162-177, 2005

[20] A. Sunderesan, A. K. Roy Chowdhury and R. Chellappa. A hidden Markov model based framework for recognition of humans from gait sequences. *In Proc. of IEEE Int. Conf. on Image Processing,* 2003.

[21] D. Tolliver and R. Collins. Gait shape estimation for identification. In Proc. *Int. Conf. on Audio- and Video-Based Biometric Person Authentication,* 2003.

[22] L. Lee, G. Dalley and K. Tieu. Learning pedestrian models for silhouette refinement. In Proc. *Int. Conf. on Computer Vision,* 2003.

[23] I. Venkat and P. D. Wilde. Robust Gait Recognition by Learning and Exploiting Sub-gait Characteristics. *International Journal of Computer Vision,* 91: 7–23, 2011

[24] Y. Chai, J. Ren, R. Zhao and J. Jia, Automatic gait recognition using dynamic variance features, In Proc. 7th *Int. Conf. on Automatic face and Gesture Recognition*, Southampton, U.K., pp. 475-480, 2006.
[25] J. Lu and Y.-P. Tan, Gait-based human age estimation, *IEEE Trans. Information Forensics and Security*, 5(4): 761-770, 2010
[26] D. S. Matovski, M. S. Nixon, S. Mahmoodi and J. N. Carter, The effect of time on gait recognition performance, *IEEE Trans. Information Forensics and Security*, 7(2): 543-552, 2012
[27] Y. Chai, J. Ren, W. Han and H. Li, Human gait recognition: approaches, datasets and challenges, in Proc. 4th Int. Conf. on Imaging for Crime Detection and Prevention, Surrey, U.K., Nov. 2011
[28] J. Han, G. Awad, A. Sutherland, Modelling and segmenting subunits for sign language recognition based on hand motion analysis, Pattern Recognition Letters, 30(6): 623-633, 2009
[29] J. Ren, J. Jiang and T. Vlachos, High-accuracy Sub-pixel Motion Estimation from Noisy Images in Fourier Domain, IEEE Trans. Image Processing, 19(5): 1379-1384, 2010
[30] J. Han, S. McKenna, Lattice Estimation from Images of Patterns that Exhibit Translational Symmetry, Image and Vision Computing, 32(1): 64-73, 2014
[31] J. Ren, ANN vs. SVM: Which one performs better in classification of MCCs in mammogram imaging, Knowledge-Based Systems, 26: 144-153, 2012
[32] J. Alkhateeb, J. Jiang, J. Ren and S. Ipson, Component-based segmentation of words from handwritten Arabic text, Int. J. Computer Systems Science and Engineering, 5(1), 2009.
[33] J. Ren and J. Jiang, Hierarchical modelling and adaptive clustering for real-time summarization of rush videos, IEEE Trans. Multimedia, 11(5): 906-917, 2009
[34] J. Ren and T. Vlachos, Efficient detection of temporally impulse dirt impairments in archived films, Signal Processing, 87(3): 541-551, 2007.