Decision-Level Fusion for Single-View Gait Recognition with Various Carrying and Clothing Conditions

Amer Al-Tayyan[a,*, 1], Khaled Assaleh[b], Tamer Shanableh [c]

[a] *WSP | Parsons Brinckerhoff, Sharjah-UAE*
[b] *Department of Electrical Engineering, American University of Sharjah, Sharjah-UAE*
[c] *Department of Computer Science and Engineering, American University of Sharjah, Sharjah-UAE*

***Abstract*—Gait Recognition is one of the latest and attractive biometric techniques, due to its potential in identification of individuals at a distance, unobtrusively and even using low resolution images. In this paper we focus on single lateral view gait recognition with various carrying and clothing conditions. Such a system is needed in access control applications whereby a single view is imposed by the system setup. The gait data is firstly processed using three gait representation methods as the features sources; Accumulated Prediction Image (API) and two new gait representations namely; Accumulated Flow Image (AFI) and Edge-Masked Active Energy Image (EMAEI). Secondly, each of these methods is tested using three matching classification schemes; image projection with Linear Discriminant Functions (LDF), Multilinear Principal Component Analysis (MPCA) with K-Nearest Neighbor (KNN) classifier and the third method: MPCA plus Linear Discriminant Analysis (MPCA+LDA) with KNN classifier. Gait samples are fed into the MPCA and MPCALDA algorithms using a novel tensor-based form of the gait images. This arrangement results into nine recognition sub-systems. Decisions from the nine classifiers are fused using decision-level (majority voting) scheme. A comparison between unweighted and weighted voting schemes is also presented. The methods are evaluated on CASIA B Dataset using four different experimental setups, and on OU-ISIR Dataset B using two different setups. The experimental results show that the classification accuracy of the proposed methods is encouraging and outperforms several state-of-the-art gait recognition approaches reported in the literature.**

***Keywords*—Biometrics, Gait Recognition, Decision-Level Fusion, Accumulated Prediction Image, Accumulated Flow Image, Edge-Masked Active Energy Image, Multilinear Subspace Learning.**

I. Introduction

Gait is defined as "the *coordinated* and *cyclic* combination of movements that result in human *locomotion"* [1]. As such, and based on these repeated patterns, it can be used to identify people. The use of gait traits in biometrics is increasingly attracting researchers' interest. This is mainly due to its potential in identification of individuals at a distance. It can be also applied to subjects unobtrusively, without their cooperation or awareness. The application of gait in biometrics requires specific focus on some covariates like the view angle [2], [55], [60], [61] and [62], clothing [3], [4], [5], [6], [51], [52] and [53], footwear [2] and [7], time span [2], fatigue and muscle development [1] and carrying conditions [3], [4], [5], [6] and [52].

One important decision to make before testing any gait recognition algorithm is the selection of the database. The database should be suitable to the application under study, and is used to train, test and evaluate the system.

Gait analysis methods can be generally classified into Model-Free and Model-Based. Model-based methods use the parameters taken from the human body structure or its kinematic data to build a model. Parameters can be, for instance, taken from stride length, stride speed and cadence [8]. Lee and Grimson [9] modeled the human body into seven regions, and represented each region by an ellipse. Cunado *et al.* [10] used models of the legs, as they found harmonics in the legs motion. Yoo *et al*. [11] modeled the body's parts into sticks. Yam *et al*. [12] used the pendulum-like movement of the legs as a model. Dockstader *et al*. [13] used a model of thick lines connected with points for different parts of the body and made use of the pendulum motion of the lower part for feature extraction. Bobick *et al*. [14] and BenAbdelkader *et al*. [15] have adopted the structural model of stride parameters to extract gait features. Tanawongsuwan *et al*. [16] and Wang *et al*. [8] modeled the different joints trajectories in the upper as well as the lower parts of the body. Zang *et al*. [17] used five-link biped model of the upper and lower body parts. Recently, Guasch *et al*. [18] used radar techniques to create a model of the human body based on the Doppler frequency

* Corresponding Author. Tel: +971 50 4164716
[1] Present Address: WSP | Parsons Brinckerhoff in the Middle East, email: amer.altayyan@wspgroup.ae
E-mail addresses: b00038189@aus.edu (A. Al-Tayyan), kassaleh@aus.edu (K. Assaleh), tshanableh@aus.edu (T. Shanableh).

(Doppler Signature). Luo *et al*. [54] used 3-dimensional parametric gait model reconstructed from multi-view silhouettes. 3D gait models were also used by Fernandez *et al*. in [62] combined with 2D silhouettes analysis techniques.

Model-free, or appearance-based, approaches deal with body motion holistically. Compared to model-based methods, the later generally exhibit higher computational complexity which makes them less desirable. This explains why most of the current research on gait recognition adopts model-free approach, and mainly the silhouette-based and energy-based methods.

One of the earliest model-free approaches for automatic gait recognition is dated back to 1994, and introduced by Niyogi and Adelson [19] who used spatiotemporal (XYT) patterns of gait. Little and Boyed have used phase features introducing the Shape of Motion method in [20]. BenAbdelkader *et al*. [21] implemented self-similarity between silhouettes to extract features. Later, Sarkar *et al*. [2] developed a major work in gait recognition, the HumanID Gait Challenge, in which they introduced a large database for gait as well as new techniques based mainly on silhouette similarity. Energy-based and accumulated error-based methods have then dominated the research trend with major contributions such as; The Average Silhouette by Liu and Sarkar [22], Gait Energy Image (GEI) by Han and Bhanu [23] and [24], Frame Difference Energy Image by Chen *et al*. [25], Accumulated Prediction Image (API) by Shanableh *et al*. [26], Active Energy Image (AEI) by Zhang *et al*. [3], Gait Flow Image (GFI) by Lam *et al*. [27], Shifted Energy Image (SEI) by Huang and Boulgouris [4], Chrono-Gait Image (CGI) by Wang *et al*. [6], Enhanced GEI using Joint Sparsity Model (GEI_{JSM}) by Yogarajah *et al*. [47], Averaged Gait Key-phase Image (AGKI) by Choudhury and Tjahjadi [52], Segmented GEI (SGEI) by Choudhury and Tjahjadi [55] and Gait Texture Image (GTI) by Kusakunniran *et al*. [59]. These methods have proven to be the best trade-off between computational complexity and recognition performance.

Recent research studies in biometrics have shown that single-method, or single-modality, biometric systems can be susceptible to noise, sensors' sensitivity and redundancy of features [28] and [29]. This is particularly an issue in behavioral techniques such as gait. Therefore, the typical solution would be fusing information from multiple sensors, feature extractors or classifiers. In other words, fusion at feature-level, score-level or decision-level. Fusion may also take place at the sensor level (before feature extraction). An example of this is the fusion of 2D and 3D faces using two different sensors. Fusion techniques bring to biometrics advantages like higher recognition rates, overcoming the issue of small sample size or training data, higher immunity to noise and spoof attacks and smaller FAR/FRR figures. The advantages of fusion techniques make this approach attractive in biometrics, in spite of presenting higher storage requirements, processing time and computational complexity.

Feature-level fusion can be seen in [30] where features from gait and foot pressure are concatenated. Face and gait features were fused by Chellapa *et al*. [31] and Hossain *et al*. [32]. In [33] static and dynamic gait features were fused. This fusion technique usually results into a higher dimension data, which can be then fed into a dimensionality reduction module. A more recent feature fusion technique was presented in [58], by fusing information from Euclidean and Riemannian spaces.

In score-level fusion, scores usually need to be normalized before fusion is possible, and then combined using the sum, max, min or product rules. Zhang *et al*. [34] used score-level fusion in which scores were normalized using *tanh* technique.

A third approach to fusion is the decision-level fusion in which the final decision is based on the combination of decisions from multiple classifiers or recognition modules. One advantage of decision fusion is that it picks the correct decisions made by the base classifiers and combines them producing a more accurate over-all decision. It provides a robust performance against challenges that each classifier has to deal with. Decision-level fusion becomes more attractive and useful when training samples are insufficient. There are different techniques for decision-level fusion such as; majority voting [35], Bayesian Decision Theory [36], Neural Networks [37] and the Dempster-Shafer theory of evidence [38] and [39]. Majority voting, however, seems to be simpler and easier to implement.

The main contributions in this paper are the following

1) Proposing and evaluating two novel gait representations (explained in section II); the Accumulated Flow Image (AFI) and the Edge-Masked Active Energy Image (EMAEI).
2) Further implementation and testing of multilinear subspace learning methods for gait feature extraction. The first implementation of these methods on the gait recognition problem was by Lu *et al*. [40] using the Multilinear Principal Component Analysis (MPCA) technique. However, we have tested the method on the CASIA B dataset (explained in section V-C) for the first time, introducing novel tensorial gait representations that proved to be efficient.
3) Testing the outcome of decision-level fusion on the overall performance of gait recognition systems.
4) Achieving competitive recognition rates, compared to several state-of-the-art methods.

II. Gait Representation

In this section we introduce the gait pre-processing methods used, starting with the Accumulated Prediction Image (API) method which was borrowed from [26] due to its promising performance coupled with the relatively simple concept and low computational cost.

In addition to the API, two new methods were developed and tested; the Accumulated Flow Image (AFI) which is very similar to the API in shape however based on the optical flow principles, and a variation of the AEI method from [3] based on masking the static part of the accumulated silhouettes.

A. Accumulated Prediction Image (API)

The formation of an accumulated prediction image (or

images) is analogous to the prediction error computation technique used in digital video coding. Each frame is subtracted from its immediate previous (past) frame, a process known as forward prediction. This subtraction results in an error image. A threshold is then applied to the image so as to filter out those pixels that are non-motion related. The resulted prediction error images after thresholding are then accumulated resulting in the Accumulated Prediction Image (API). For further filtration of the non-motion (or noise) pixels, a decision is made whether to use forward prediction or backward prediction (difference between each frame and the immediate future frame). This decision is based on the minimization of the Sum of Absolute Difference (SAD) between the forward prediction error image and the backward prediction error image. The image that minimizes this SAD value is taken towards the final construction of the API. Additionally, for the sake of a better representation, the process is split into two parts; positive API and negative API (by reversing the order of subtraction). This process is illustrated in Fig. 1, and the resulting accumulated images are also shown. Each video sequence ends up with two APIs that can be used for feature extraction. Note that the spatial information is maintained in the two images, an aspect that shall be of added value in feature extraction.

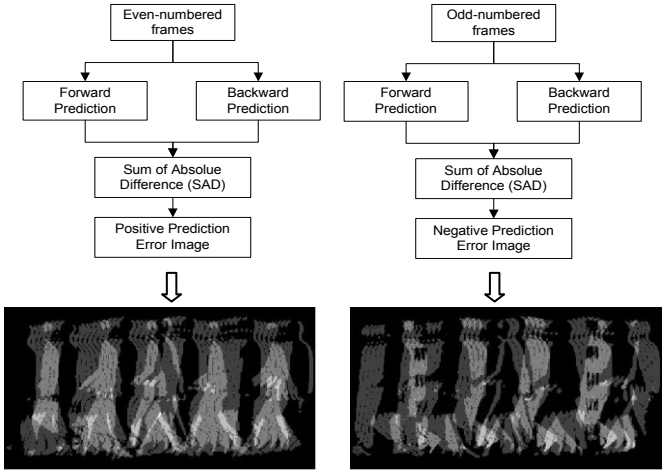


Fig. 1: API's of Subject 1, Normal Sequence 1, CASIA B Database, (a) Positive API, (b) Negative API

API method was first introduced and successfully used in [26] to report high recognition rates on a locally-collected gait database. It is noticed that this gait representation, followed with the feature extraction technique explained in section III-A (image projection + 1D-DCT), contributed in extracting linearly-separable features that are easy to input into a linear classifier, in a relatively simple approach.

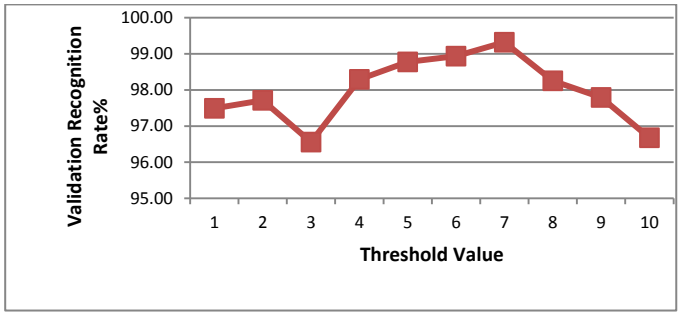


Fig. 2: Validation Recognition Rates using different threshold values for API

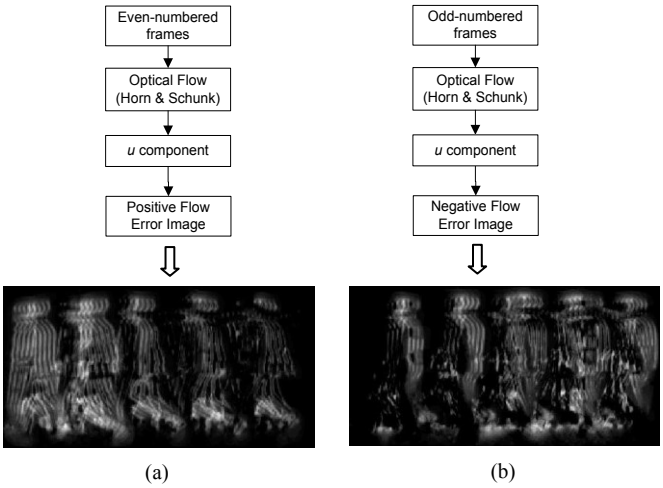


Fig. 3: AFI's of Subject 1, Normal Sequence 1, CASIA B Database, (a) Positive AFI, (b) Negative AFI

As mentioned above, this method is dependent on thresholding, which can be optimized. In principle, the threshold can be any digit between 1 and 255. Clearly, very small threshold values result in waiving some representative features. Similarly, we wish to remove the intrinsic noise. It was observed that the maximum threshold value to test for should be 10. This can be understood knowing that the intensity values 1 to 10 of the API image represent more than 90% of the non-zero pixels of the image. Beyond 10, thresholding has almost no effect. Therefore, the threshold parameter was varied between 1 and 10, with a step value of 1, and tested accordingly for this individual method and using the training data from Setup 1, which is explained further in Section V-C. These are 3 normal sequences; *nm1*, *nm2* and *nm3*. Sequences *nm1* and *nm2* are used in training for the parameters, while *nm3* is used for validation. As seen in Fig. 2, the final optimum value was 7. In [26], the threshold was set to the 75th percentile of the non-zero pixels of the prediction error image, which approximately corresponds to 3. Upon testing however, the value 7 (corresponding to the 90th percentile) yielded better performance in average.

B. Accumulated Flow Image (AFI)

Optical flow corresponds to the movement of pixels in a sequence of images [41], or in other words, the rate of change of motion (velocity) of intensity patterns. Such information is very useful for gait recognition. It can be used in objects'

segmentation and motion detection and produces a pool of features for object classification. In the literature, there are different methods for the computation of optical flow, including differentiation, correlation, feature-based and hierarchical methods. Differentiation methods are classical, yet popular. These include global differentiation techniques such as Horn and Schunk, and local differentiation techniques such as Lucas and Kanade. In the AFI representation, images are constructed using the Horn and Schunk differential method [42]. Therefore, we summarize its main equations, constraints and assumptions.

The first assumption made in the calculation of optical flow is that the image brightness, reflectivity and illumination are constant while moving in a short time interval Δt, from *t1* to *t2*. Let $I(x,y,t)$ be the image intensity at point (x,y) and time t. I is constant at a fixed point, that is

$$I_x u + I_y v + I_t = 0 \tag{1}$$

where u and v are the two components of the flow vector **v,** which we seek in the computation of the optical flow. u and v are given by:

$$u = \frac{dx}{dt} \; and \; v = \frac{dy}{dt}$$

The second constraint in Horn and Schunk method is the smoothness constraint. The idea is that the optical flow should vary smoothly everywhere over the entire image. This can be derived by minimizing the square of the magnitude of the optical flow vector **v.**

$$E_c^2 = \left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial u}{\partial y}\right)^2 + \left(\frac{\partial v}{\partial x}\right)^2 + \left(\frac{\partial v}{\partial y}\right)^2 \tag{2}$$

This equation represents the difference between the flow vector and its neighbors. Adding this to equation (1), or the constraint equation, enables us to solve for u and v. After a method of derivation explained in [42], we get two equations to iteratively update for the values of u and v:

$$u^{n+1} = \bar{u}^n - \frac{I_x\left[I_x\bar{u}^n + I_y\bar{v}^n + I_t\right]}{\alpha^2 + I_x^2 + I_y^2} \tag{3}$$

$$v^{n+1} = \bar{v}^n - \frac{I_y\left[I_x\bar{u}^n + I_y\bar{v}^n + I_t\right]}{\alpha^2 + I_x^2 + I_y^2} \tag{4}$$

where; u^{n+1} and v^{n+1} are the new optical flow components at iteration n+1, $\bar{u}^n$ and $\bar{v}^n$ are the averages of the optical flow vector components in a small neighborhood, and at previous iteration n. I_x and I_y are the spatial gradients. I_t is the temporal gradient, and α^2 is a weighing factor.

The construction of the AFI's is based on equations (3) and (4), being an iterative method. As such, the parameters that govern its behavior are the number of iterations, the smoothness factor α and the initial values of u and v. In these experiments, we assume initial values of 0 for each of u and v components.

The computation of the optical flow between every two consecutive frames produces a flow error image. Similar to the work in computing the Accumulated Prediction Image (API), error images are accumulated to produce the final AFI. In this case, however, we have to deal with two sets of variables every time, u and v. Recalling that the process is actually split into two parts, positive and negative image differences, means we will have 4 images upon the computation of each image difference. However, after the initial experiments, it was found that the horizontal component (u) results in higher recognition rates. Therefore, in all subsequent analysis we have only considered the u-based AFI's. Fig. 3 shows sample AFI's constructed from the first sequence of subject 1 in the normal condition of the CASIA B database.

The number of iterations and the smoothness factor are important parameters in defining the quality of the output gait image. As such, these two parameters were varied and the method tested accordingly, yet without optimization. Exhaustive experiments may still be required on this part.

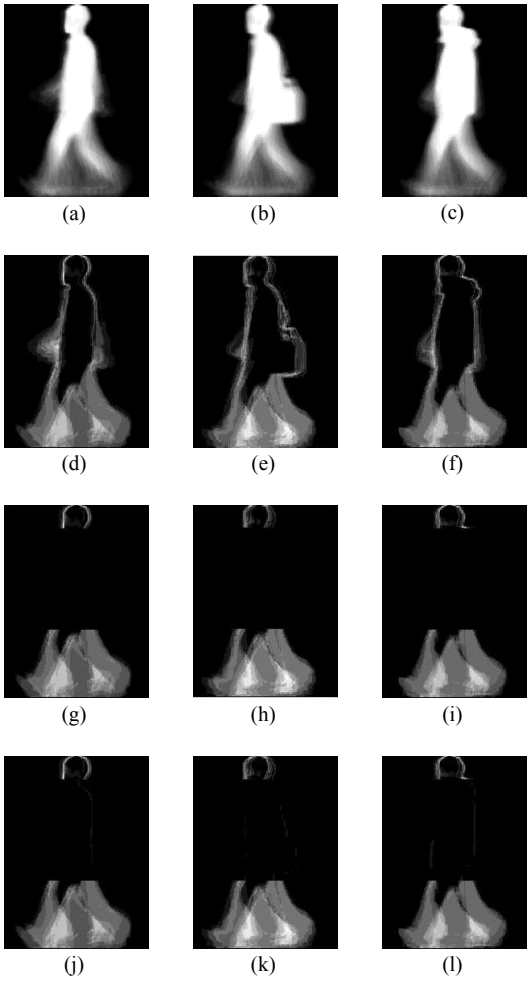


Fig. 4: (a) GEI of subject 1 under normal condition, (b) GEI of subject 1 with bag, (c) GEI of subject 1 with coat, (d) AEI of subject 1 under normal condition, (e) AEI of subject 1 with bag, (f) AEI of subject 1 with coat, images (g, h, i) are the corresponding MAEIs to images (d, e, f) using Zero Masking, images (j, k, l) are the corresponding MAEIs to images (d,e,f) using Edge Masking.

1) Number of iterations: These were varied between 1 and 100 but only at randomly selected values; 1, 5, 10, 20, 32, 64 and 100. Results were best at 64, and so it was fixed for all succeeding experiments. This selection could still be optimized by further experiments.
2) Smoothness factor: This value is responsible for the quality of the output gait image and how representative it is for the motion patterns. The higher the value, the smoother the image. However, at some relatively high values, smoothness leads to less representative features. At initial random selection of values, it was noticed that values above 20 should be avoided. The parameter was, thus, varied between 1 and 20 with a step value of 5. The value corresponding to the optimum result was 5, and so it was fixed for the succeeding experiments.

C. Edge-Masked Active energy Image (EMAEI)

Recall that the GEI method [43] produces a single gait sample image per cycle, in which the intensity at a single point is a representation of the frequency of that particular part of the body. It is noticed that GEI accumulates both dynamic and static features of the moving body, which makes it appearance-based and sensitive to the change in clothing and carrying conditions. Zhang *et al*. introduced in [3] the AEI gait representation that better highlights the dynamic features of the subject and minimizes those related to the static parts, especially the bag and the coat in our case.

This was achieved by accumulating the differences (instead of summation in GEI) between every two consecutive frames and averaging them by the number of frames *N*, or:

$$A(x,y) = \frac{1}{N}\sum_{t=0}^{N-1} D_t(x,y) \tag{5}$$

where $D_t(x,y)$ is the difference between frames given as:

$$D_t(x,y) = \begin{cases} f_{t(x,y)}, & t = 0 \\ \left\| f_t(x,y) - f_{t-1}(x,y) \right\|, & t > 0 \end{cases} \tag{6}$$

where $f_t(x,y)$ is the tth silhouette.

Example AEI images are shown in Fig. 4 and compared to GEI for the same subject (Subject 1) taken from the same dataset (CASIA B).

An observation of the resulted AEI images indicates that the mid portion of the images seem to have the least dynamic features. It was intuitive to think about tracing those areas of less contribution to the final features and masking them out, and expect this to improve the final recognition rate when testing for the bag and coat sequences. More precisely, if the human body is divided into 8 proportions, it's the area between the 1st proportion (just under the chin) and 5th proportion (mid-thigh) that is masked, or proportions 2-5. We opted to implement a simple model that imposes the mask manually during the formation of the AEI resulting in a masked AEI or MAEI. A more efficient and optimized system can be developed to trace the least significant features and mask them out automatically. This can be accomplished using filtering, wrapping or decision tree approaches. Good examples of feature subset selection techniques were presented in [44] and [55].

Two masking versions have been tested in this work:

1) Zero Masking: This simply replaces all values that lie between two predetermined lines by zeroes. The two lines were at rows 30 (Proportion 2 of the body) and 150 (Proportion 5 of the body). Knowing that all AEI images are of fixed normalized size 240x180 pixels, means this fixed selection should generalize well as a primary mask. Example masked AEI images using this approach are shown in Fig. 4, images g, h and i.
2) Edge Masking: Instead of zeroing out all pixels between the two lines (at rows 30 and 150), we applied an edge detection technique to the same area. The intention is to minimize the contribution of the static features related to the bags and coats, while still maintaining some discrimination power in them. Edge detection, using 'Sobel' edge operator, is applied to every difference image prior to image accumulation. This method performed better compared to Zero Masking, based on the average recognition rates in both cases. We notice from Fig. 4 that although the Edge-Masked AEI's (or EMAEI's), j, k and l, show very little information in the mid portion, are still non-zero, indicating those areas with motion difference.

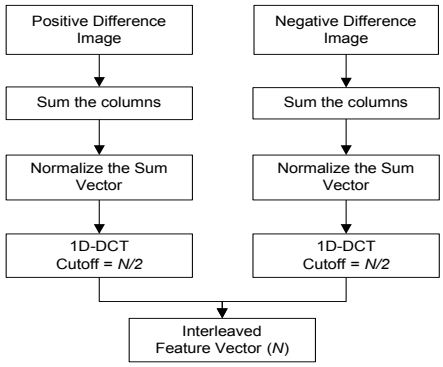


Fig. 5: Illustration of Image Projection + 1D-DCT Method

III. Feature Extraction

A. Image Projection and 1D-DCT

This simple approach takes the horizontal projection of each gait image, followed by one-dimensional Discrete Cosine Transform (DCT) to smooth and reduce the size of the projected values. This produces one feature vector which is truncated using a pre-selected cutoff value, the value of which is determined empirically. After testing the code through a range of dimensionality values between 20 and 200, the value 100 was selected. When applying this approach to the API's and AFI's, being two images per sample, we use half the dimensionality value as the cutoff for each image. Then the two feature sub-vectors are interleaved to produce one combined vector of 100 variables. This method is illustrated in Fig. 5. When used with the edge-masked AEI gait image, the

whole image produces one single feature vector of dimensionality 100.

B. Multilinear Principal Component Analysis (MPCA)

MPCA is an unsupervised Multilinear Subspace Learning (MSL) method that implements the Tensor-to-Tensor Projection (TTP) technique, to project high-order tensor objects into lower-dimensional tensors. In other words, it applies dimensionality reduction directly on objects in their tensorial form. Vector-based subspace learning methods have been used in gait recognition several times in the literature; recently in [56] and [57]. However, the main advantage of using this method, which is introduced in [40], is that it is generalized to work on higher-order tensors. Therefore, it introduces systematic procedure and provides tensor representation to tensorial objects, compared to other heuristic methods or those that use vector or matrix forms.

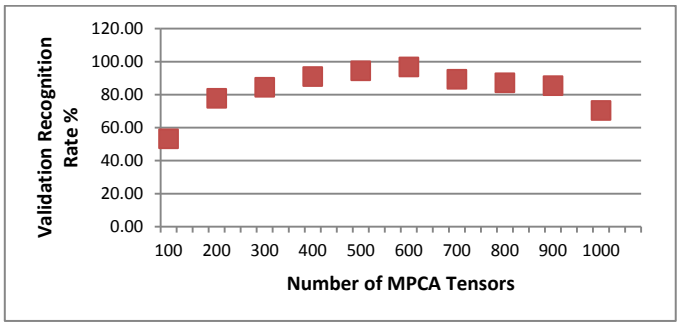


Fig. 6: Effect of varying the number of tensors in the MPCA algorithm

MPCA was implemented on gait recognition, tested on the USF database and compared to the HumanID algorithm introduced in [2]. It was, thus, a good benchmark to develop our method and compare our results.

The objective of the MPCA algorithm is to define N multilinear projection matrices$\{U^n \in R^{I_n \times P_n}, n = 1, \dots, N\}$, each of which is to map an input tensor object $\mathcal{X}_m \in R^{I_1 \times I_2 \times \dots \times I_N}$ into a lower-dimensional tensor object $\mathcal{Y}_m \in R^{P_1 \times P_2 \times \dots \times P_N}$ (where $P_n < I_n$ for n=1, 2,…, N).

The MPCA objective function is defined as:

$$\{U^n, n = 1, \dots, N\} = argmax_{U^1, U^2, \dots, U^N}(\psi_{\mathcal{Y}}) \quad (7)$$

Where $\psi_{\mathcal{Y}}$ is the total tensor scatter, defined as:

$$\psi_{\mathcal{Y}} = \sum_{m}^{M} \|\mathcal{Y}_m - \bar{\mathcal{Y}}\|_F^2 \quad (8)$$

Where M is the number of samples and $\bar{\mathcal{Y}}$ is the mean tensor calculated as:

$$\bar{\mathcal{Y}} = \frac{1}{m} \sum_{m}^{M} \mathcal{Y}_m \quad (9)$$

Equation 7 means that the objective of MPCA is to determine the N projections that maximize the total tensor scatter $\psi_{\mathcal{Y}}$.

Below is a general MPCA algorithm:

Input: $\mathcal{X}_m \in R^{I_1 \times I_2 \times \dots \times I_N}, m = 1, \dots, M$

Algorithm:

1. Center the input samples,$\tilde{\mathcal{X}}_m = \mathcal{X}_m - \bar{\mathcal{X}}$, where $\bar{\mathcal{X}} = \frac{1}{m}\sum_{m=1}^{M} \mathcal{X}_m$ is the sample mean.
2. Calculate the eigen decomposition of $\phi^n = \sum_{m=1}^{M} \tilde{X}_{m(n)} \cdot \tilde{X}_{m(n)}^T$
3. Set $\tilde{U}^n$ to consist of the eigenvectors corresponding to the most significant P_n eigenvalues.
4. Calculate cumulative distribution of eigenvalues, $\lambda_{i_n} = \lambda_{i_{n-1}} + \lambda_{i_n}$
5. Determine the dimension of the projected space P_n
6. For k=*1:K* (K: number of iterations)
   - For n=*1:N*
     Set $\tilde{U}^n$ to consist of the P_n eigenvectors of ϕ^ncorresponding to the largest P_n eigenvalues.
7. Calculate the weight tensor (to be used in classification):

$$\mathcal{W}(p_1, p_2, \dots, p_N) = \sqrt{\prod_{n=1}^{N} \lambda_{p_n}^{(n)}}$$

8. Calculate the projection of tensor samples:

$$\mathcal{Y}_m = \mathcal{X}_m \times_1 \tilde{U}^{(1)T} \times_2 \tilde{U}^{(2)T} \dots \times_N \tilde{U}^{(N)T}, m = 1, \dots, M$$

Output: $\mathcal{Y}_m \in R^{P_1 \times P_2 \times \dots \times P_N}, m = 1, \dots, M$

Finally, the output tensor is vectorised and sorted by discriminality. The number of vector components kept for the analysis is determined empirically. We have tested values ranging between 100 and 1000, with a step value of 100, as seen in Fig. 6. These are similar to the values range used in [40]. The value that yielded the best average classification rates using different gait inputs was 600. Training and validation data is same from Section II-A

In an attempt to optimize the iterative solution, the following techniques were used in [40] and adopted in our work:

1. Initialization by Full Projection Truncation (FPT): This is done by starting iterations in each mode n by assuming $P_n = I_n$. In each mode, we get I_n number of eigenvalues. These are sorted in descending order, each two successive eigenvalues accumulated and then normalized by dividing by the sum of all eigenvalues. The resulted accumulated and normalized eigenvalues are then used to determine the projection dimensionality reduction. This method proved to yield quick convergence, using only one iteration.
2. Determination of the tensor subspace dimensionality: using the Q-based method. Q is a user-defined number, determined empirically, that represents the number of accumulated eigenvalues (or energy) kept. If Q=97, for example, P_n is identified as the index corresponding to the normalized eigenvalue of less than or equal to 0.97.

In this work Linear Discriminant Analysis (LDA) is also applied after the MPCA output to maximize the ratio of the between-class scatter matrix to the within-class scatter matrix, in an attempt to produce higher class discriminability. Therefore, we have actually tested two methods; MPCA and MPCA+LDA. In both sets of experiments the data setup and classifier are the same. In the case of the LDA algorithm, the dimension of the feature vector is fixed to C-*1*, where C is the number of classes. In our data C=124, yielding an LDA feature vector of dimension 123. In this case, the number of vector components kept in the analysis, and subsequent experiments on different values, is 200.

The MPCA method was applied to our pre-processed gait images (explained in Section II) so as to extract the gait features. With three gait representation methods, namely API, AFI and edge-masked AEI, we shall have three models for feature extraction using the MPCA algorithm and three more using the MPCA+LDA algorithm. For this purpose, our data need to be arranged first in tensorial form. Compared to the work in [40], the gait cycle determination module was dropped in our work, since our accumulated gait images present these intrinsically. As such, we chose the gait samples to be the accumulated images mentioned earlier. With these arranged in tensorial form and fed to the MPCA algorithm, a novel feature representation method is introduced; namely, Accumulated Gait Tensor (AGT). Primary comparative experiments have yielded better results using this representation compared to using single silhouettes or cycles. However, further analysis on this part, and using different gait databases and experimental set-ups, is necessary before generalizing this finding. One advantage of the AGT's is reducing the dimensionality of the tensorial input dramatically while preserving discriminative features. Additionally, we reduce the input tensors from being three-mode (3rd-order) tensors, spatial and temporal modes, into two-mode (2nd-order) tensors since the time-mode is embedded in the accumulated gait images. Below is a break-up of the tensorial data representation and dimensionality of each of the three gait representations:

a) API Tensorial Representation

The API approach applies to the original frames, and these are of dimension 240x320 pixels. In our preprocessing, the images are cropped to 168x320 pixels, where the cropped portion has no information. Moreover, and for the sake of using them in the MPCA algorithm, it was intuitive to reduce the resolution. We tested using 25%, 50% and 75% of the original resolution. The differences in the recognition rates were negligible, thus we opted to use the maximum reduction, 75%, yielding images of 42x80 pixels. Since in API there are two accumulated images per subject, positive and negative, the final tensor dimension would be 42x80x2, which is still a 3rd-order tensor. However, compared to using the individual silhouettes, which average to approximately 60 per sequence, in the latter case we get a tensor dimension of 42x80x60. And using each gait half-cycle as a sample, with average gait half-cycle of approximately 10 silhouettes, the tensor dimension in this case is 42x80x6. The reduction of the input tensor dimensionality is therefore apparent. Applying the AGT concept to the data, the term Accumulated Prediction Tensors (APT's) shall be used in our experimental analysis later to indicate the APIs in tensorial form.

b) AFI Tensorial Representation

AFI also uses the 240x320 frames. Therefore, the produced tensors have the same three-mode of 42x80x2. This data is called Accumulated Flow Tensors (AFTs).

c) EMAEI Tensorial Representation

Unlike in the two previous gait representations, edge-masked AEI uses the gait silhouettes, and not the frames, as inputs. These are also found in the CASIA B dataset with the resolution of 240x320. In preprocessing, all silhouettes were cropped to produce images of 240x180 pixels, in which silhouettes were centered and resized to occupy the full frame. This way, we ensure that all silhouettes are aligned. In preparation for using them in the MPCA algorithm, the resolution is reduced by 75% to 60x45 pixels. And knowing that we have one image to represent each gait sequence means that this problem is actually reduced into a two-mode tensor problem, or two-dimensional (2D) PCA. For the sake of consistency with the other two tensorial representations, however, we will call the tensor data here Edge-Masked Active Energy Tensors (EMAETs).

IV. Classification and Fusion techniques

We opted to use two simple classifiers in our tests; Linear Discrinimant Functions (LDF) and K-Nearest Neighbor (KNN, K=1). Some more advanced classifiers, such as the Polynomial Networks, Support Vector Machines (SVM) and Neural Networks might be tested to compare the performance and final results. The distance measure we used in the 1-NN classifier is the Modified (or weighted) Angle Distance (MAD) from [40], given by:

$$d(a,b) = -\frac{\sum_{h=1}^{H} a(h).b(h)}{w(h)\sqrt{\sum_{h=1}^{H} a(h)^2 \sum_{h=1}^{H} b(h)^2}} \tag{13}$$

Where H represents tensors and $w(h)$ represents the weight tensor computed in the MPCA algorithm, and defined as:

$$\mathcal{W}(p_1, p_2, \dots, p_N) = \sqrt{\prod_{n=1}^{N} \lambda_{p_n}^{(n)}} \tag{14}$$

Where $\lambda_{p_n}^{(n)}$ represents the p_nth n-mode eigenvalue corresponding to the projection matrix $U^{(n)}$.

We used the following voting schemes for decision-level fusion:

1) Unweighted Voting (UWV)

This is the simplest and most straightforward voting method for decisions' combination. It simply counts the number of decisions for each class and assigns the sample to the class that received the highest number of votes. In this case the final decision D is given by:

$$D = arg \max_{C_j} \sum_{1}^{k} \delta(D_k, C_j) \tag{15}$$

Where C_j is the target class,
D_k is the decision of the base classifier C_k, and

$$\delta = \begin{cases} 1, if\ D_k = C_j \\ 0\ otherwise \end{cases} \tag{16}$$

2) Weighted Voting (WV)

By introducing weights, we aim to give more significance to those base classifiers that perform better individually.

Most of the weighted voting methods are derived from evaluating the decisions of all base classifiers. They are sorted according to their estimated accuracy (using the training dataset), and proportional weights are assigned. Accuracy is estimated by validation. For this purpose, the training data used in Setup 1 (*Tr1*) (explained in section V-C) is split into two parts; 2 *normal* sequences for training and 1 *normal* sequence to test for the classifiers' accuracy. We test here one simple weighted voting method.

After estimating the error of each base classifier e_k based on the validation data, we evaluate the authority value a_k, which is equivalent to $1\text{-}e_k$. The weight w_k is given by:

$$w_k = \frac{a_k}{\sum_i a_i} \tag{17}$$

After assigning the weights to all base classifiers, the final decision will be evaluated using similar equation to 15, by only adding the weights w_k:

$$D = arg \max_{C_j} \sum_{1}^{k} w_k \delta(D_k, C_j) \tag{18}$$

V. Experimental Results and Analysis

A. Final proposed scheme

Fig. 7 illustrates a simplified block diagram of our proposed gait recognition system. A decision-level fusion scheme takes as inputs the decision labels from nine different gait recognition sub-systems. The sub-systems are designed in such a way to utilize three different gait representations (pre-processing techniques) with three different matching methods. The system is made-up of the following components:

Gait Representation (Pre-processing) Techniques:

1) Accumulated Prediction Image (API)
2) Accumulated Flow Image (AFI).
3) Edge-Masked Active Energy Image (EMAEI)

Each of the above-listed gait images is used three times; implementing the following methods:

1) Image projection followed by LDF classifier. We shall label this as Method 1 or (*Mthd1*).
2) MPCA, followed by KNN classifier (1-NN). We shall label this as Method 2 or (*Mthd2*).
3) MPCA+LDA, followed by KNN classifier (1-NN). We shall label this as Method 3 or (*Mthd3*).

Note that inputs to both *Mthd2* and *Mthd3* are the tensorial gait representations, APT, AFT and EMAET.

B. Testing Methodology

The next section lists-up the four different experimental setups used to evaluate the proposed method. In each setup we carry-out 11 experimental groups. Each group comprises a number of experiments equivalent to the number of testing probes as detailed for each case. The groups are as follows:

1) Single-level gait recognition:

a. API + *Mthd1*, labeled Recognizer 1 (*R1*)
b. AFI + *Mthd1*, labeled Recognizer 2 (*R2*)
c. Edge-masked AEI + *Mthd1*, labeled Recognizer 3 (*R3*)
d. APT + *Mthd2*, labeled Recognizer 4 (*R4*)
e. AFT + *Mthd2*, labeled Recognizer 5 (*R5*)
f. Edge-masked AET + *Mthd2*, labeled Recognizer 6 (*R6*)
g. APT + *Mthd3*, labeled Recognizer 7 (*R7*)
h. AFT + *Mthd3*, labeled Recognizer 8 (*R8*)
i. Edge-masked AET + *Mthd3*, labeled Recognizer 9 (*R9*)

2) Decision-level fusion schemes:

a. Three-decision Fusion Schemes detailed as follows:
(1) *3Fus123*: 3-decision fusion of *R1*, *R2* and *R3*.
(2) *3Fus456*: 3-decision fusion of *R4*, *R5* and *R6*.
(3) *3Fus789*: 3-decision fusion of *R7*, *R8* and *R9*.
(4) *3Fus147*: 3-decision fusion of *R1*, *R4* and *R7*
(5) *3Fus258*: 3-decision fusion of *R2*, *R5* and *R8*

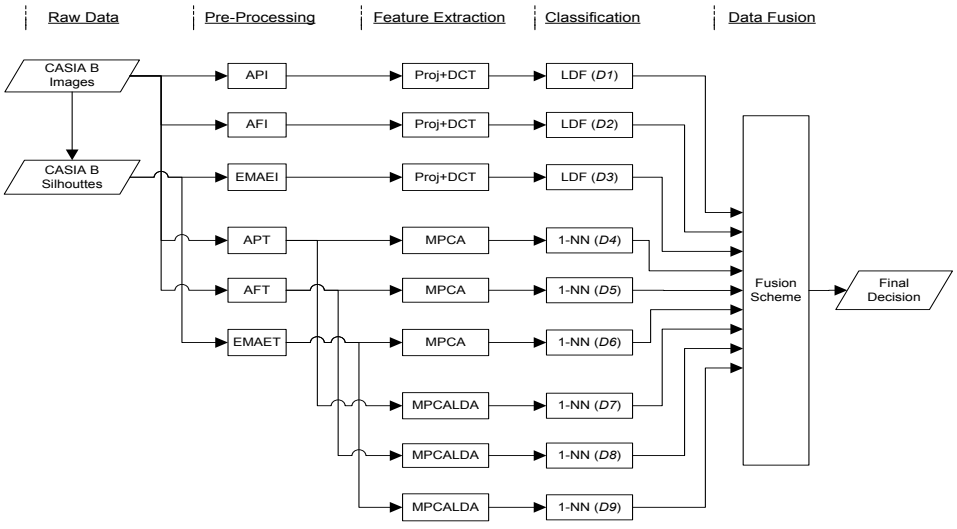


Fig. 7: Block Diagram of the proposed Gait Recognition System

TABLE I
RECOGNITION RATES (%) OF EXPERIMENTS USING SETUP 1 [3] AND [46]

	R1	R2	R3	R4	R5	R6	R7	R8	R9	9Fus (UWV)	9Fus (WV)	Zhang *et al*. [3]	Zeng *et al*. [46]
Ts11	**98.66**	98.39	65.59	93.55	93.28	90.59	89.25	76.61	90.32	**99.19**	**99.19**	98.40	98.40
Ts12	87.90	84.68	41.53	75.81	82.26	52.02	68.95	53.23	44.76	**96.77**	95.97	91.95	93.50
Ts13	40.32	51.21	42.74	42.34	51.21	81.05	39.92	41.94	76.61	83.87	88.31	72.19	**90.30**
Average	75.63	78.09	49.95	70.57	75.58	74.55	66.04	57.26	70.56	93.28	**94.49**	87.51	94.07

Ts11 (*nm1*, *nm2*&*nm3), Ts12 (bg1 & bg2)* and *Ts13 (cl1 & cl2)*

TABLE II
WEIGHTS GIVEN TO BASE RECOGNITION MODULES

	R1	R2	R3	R4	R5	R6	R7	R8	R9
e_k	**0.081**	**0.057**	0.629	0.081	0.089	0.105	0.177	0.210	**0.032**
a_k (WV)	0.919	0.944	0.371	0.919	0.911	0.895	0.823	0.790	0.968
w_k (WV)	0.122	0.125	0.049	0.122	0.121	0.119	0.109	0.105	0.128

(6) *3Fus369*: 3-decision fusion of *R3*, *R6* and *R9*

Each of these fusion experiments is repeated twice, using Unweighted Voting (UWV) and Weighted Voting (WV) techniques.

b. Nine-decision Fusion (*9Fus*) Scheme. Similarly, repeated twice using the UWV and WV techniques.

C. Experimental Setups

As mentioned previously, in this work the CASIA B database is used. This database consists of 124 subjects, each subject walked 10 times resulting in 10 video sequences; 6 sequences in normal condition, 2 sequences carrying different shapes and sizes of bags, and 2 sequences wearing different types of coats. With 11 cameras distributed at 11 different view angles, ranging between 18^o and 180^o, we end up with a total of 13,640 video sequences. As we only focused on testing for the carrying condition and clothing, we used sequences corresponding to only single view angle, which is 90^o or lateral view. This corresponds to using only 1240 sequences. The database is divided into training dataset (gallery) data and testing dataset (probe). Number of sequences in each of the two data groups has been varied based on similar setups taken from nine recently published papers that studied the same topic and used the same database. Total number of sequences is obviously the same in all the 4 setups and these are:

a. Six sequences for normal-condition walking, labeled *nm1*, *nm2*, *nm3*, nm4, *nm5* and *nm6*
b. Two sequences for walking carrying bag, labeled *bg1* and *bg2*.
c. Two sequences for walking wearing coat (cloth), labeled *cl1* and *cl2*.

The data breakdown in each of the setups is as follows:

1) Setup 1 as reported in [3] and [46]

a. Training Data (*Tr1*): 3 normal sequences; *nm1*, *nm2* and *nm3*
b. Testing Data : 3 sets for 3 experiments;
(1) *Ts11*: 3 normal sequences; *nm4*, *nm5* and *nm6*
(2) *Ts12*: 2 bag sequences; *bg1* and *bg2*
(3) *Ts13*: 2 coat sequences; *cl1* and *cl2*

Zhang *et al*. [3] used the Active Energy Image (AEI) method for gait representation, followed by two-dimensional Locality Preserving Projection (2DLPP) for dimensionality reduction and feature extraction, and finally nearest neighbor classifier. Zeng *et al*. propose in [46] a gait approach based on silhouette features via deterministic learning.

2) Setup 2 as reported in [4], [44], [45], [47] and [48]

a. Training Data (*Tr2*): 4 normal sequences; *nm1*, *nm2*, *nm3* and *nm4*
b. Testing Data : 3 sets for 3 experiments;
(1) *Ts21*: 2 normal sequences; *nm5* and *nm6*
(2) *Ts22*: 2 bag sequences; *bg1* and *bg2*
(3) *Ts23*: 2 coat sequences; *cl1* and *cl2*

Huang *et al*. [4] implemented feature-level fusion of two gait signatures; Shifted Energy Image (SEI) and Gait Structural Profile (GSP). Dupuis *et al*. [44] implemented the Random Forest algorithm for feature ranking, and produced a Masked GEI, followed by Canonical Discriminant Analysis (CDA). Gait Pal and Pal Entropy (GPPE) Image is used by Jeevan *et al*. in [45] as the gait representation, followed by PCA and SVM classifier. Experiments in [45] were carried out on 98 subjects only. In [47], Yogarajah *et al*. implement the Joint Sparsity Model, to enhance the GEI. While Jiménez *et al*. use tracklet-based gait and SVM classifier in [48]

3) Setup 3 as reported in [6]

In [6], Wang *et al*. implemented a temporal gait template, named Chrono-Gait Image (CGI), followed by PCA+LDA for feature extraction and 1-NN classifier.

a. Training Data (*Tr3*): 5 normal sequences; *nm1*, *nm2*, *nm3*, *nm4*&*nm5*.
b. Testing Data : 5 sets for 5 experiments;
(1) *Ts31*: normal sequence *nm6*
(2) *Ts32*: bag sequence *bg1*
(3) *Ts33*: bag sequence *bg2*
(4) *Ts34*: coat sequence *cl1*
(5) *Ts45*: coat sequence *cl2*

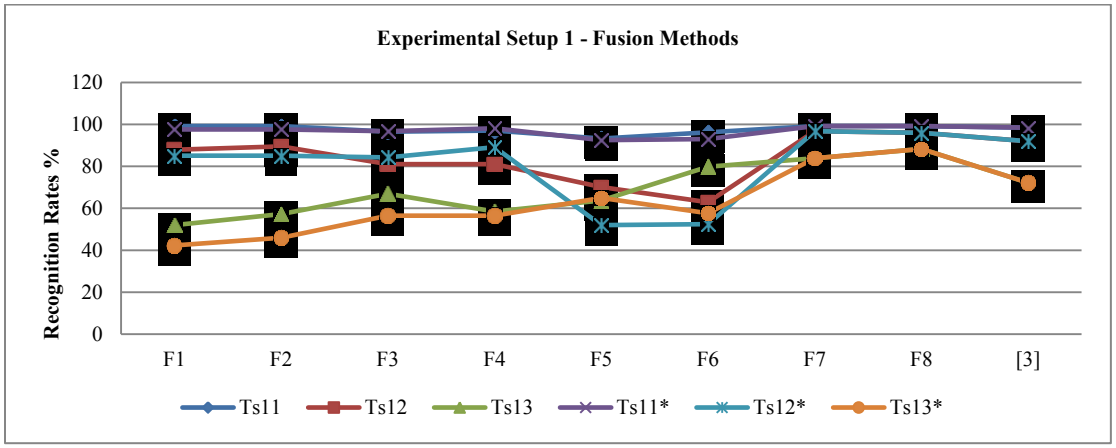


Fig. 8: Recognition Rates (%) of the experiments groups *Ts11* (*nm1*, *nm2&nm3), Ts12 (bg1 & bg2)* and *Ts13 (cl1 & cl2),* using different methods of fusion. *Ts11**, *Ts12** and *Ts13** are the experimental results for fusion schemes *F1**-*F6**

4) Setup 4 as reported in [5]

In this paper, Kumar *et al*. used the concept of 'Axis of Least Inertia' to produce gait features.

a. Training Data (*Tr4*): 3 normal sequences; *nm1*, *nm2* and *nm3* plus 1 bag sequence *bg1* and 1 coat sequence *cl1*.
b. Testing Data : 5 sets for 5 experiments;
   (1) *Ts41*: 1 normal sequence; *nm4*
   (2) *Ts42*: 1 normal sequence; *nm5*
   (3) *Ts43*: 1 normal sequence; *nm6*
   (4) *Ts44*: 1 bag sequence; *bg2*
   (5) *Ts45*: 1 coat sequence; *cl2*

D. Results and Analysis

1) Results using experimental setup 1

Table I summarizes the recognition rates (%) obtained for the experiments carried out using Training Data *Tr1* and Testing Data *Ts11*, *Ts12* and *Ts13*. Note that in this arrangement sequences of *bag* and *coat* are unseen during training. We compare our results with those obtained by Zhang *et al*. in [3] and Zeng *et al*. propose in [46]

We notice that it is possible to achieve high recognition rates when testing *normal* sequences (*Ts11*) even without fusion, using a simple linear classifier, as in *R1* and *R2*. This would be observed in all the succeeding results of other setups, when testing for *normal* sequences, which would prove that the implemented gait representation techniques, especially the API, are efficient when covariates are controlled.

We notice also that experiments using API with Projection and LDF were good enough to produce very good recognition rates when testing for *bag* sequences (*Ts12*). They are yet not good enough compared to those reported in [3] or [46]. 9-decision fusion (*9Fus*) method, however, outperforms Zeng *et al*. by almost 3%. Edge-masked AEI is still not efficient representation for the two cases of *normal* and *bag* sequences, especially in *R3*. However, it works well in the case of *coat* sequences (*Ts13*), even without optimization. We notice that although Zeng *et al*. achieved very good results, especially in the *Ts13* test, their average recognition rates from the three tests is 94.07 compared to 94.49 from our *9Fus* (WV) method.

TABLE III
RECOGNITION RATES (%) OF FUSION TECHNIQUES USING SETUP 1 [3]

Fusion Scheme	Ts11	Ts12	Ts13	Average
3Fus123/ (UWV) – F1	**99.19**	87.90	52.02	79.70
3Fus123 / (WV) – F2	**99.19**	89.52	57.26	81.99
3Fus456 / (UWV) – F3	96.51	81.05	66.94	81.50
3Fus456 / (WV) – F4	97.04	81.05	58.47	78.85
3Fus789 / (UWV) – F5	93.28	70.16	63.71	75.72
3Fus789 / (WV) – F6	96.24	62.90	79.84	79.66
3Fus147/ (UWV) – F1*	97.58	85.08	42.34	75.00
3Fus147 / (WV) – F2*	97.58	85.08	45.97	76.21
3Fus258 / (UWV) – F3*	96.77	84.27	56.45	79.16
3Fus258 / (WV) – F4*	98.12	89.11	56.45	81.23
3Fus369 / (UWV) – F5*	92.47	52.02	64.92	69.80
3Fus369 / (WV) – F6*	93.01	52.42	57.66	67.70
9Fus / (UWV) – F7	99.19	**96.77**	83.87	93.28
9Fus / (WV) – F8	**99.19**	95.97	88.31	**94.49**
Zhang *et al*. [3]	98.40	91.95	72.19	87.51
Zeng *et al*. [46]	98.40	93.50	**90.30**	94.07

We show herein a comparison between Unweighted and Weighted fusion techniques at the two levels; 3-decision fusion and 9-decision fusion. In order to do that, we need first to estimate the base classifiers' accuracies and assign them authority values, and eventually weights. For this purpose, we split the training data used in Setup 1 (*Tr1*) into two parts; 2 *normal* sequences (*nm1* and *nm2*) for training and 1 *normal* sequence (*nm3*) to test for the classifiers' accuracy. This gives us a general indication on each classifier's relative performance. Table II lists the errors (e_k), authority values (a_k) and weights (w_k) given to each of the nine base recognizers using the WV (Equation 17) method. These weights are used throughout the different experimental setups. However, we will demonstrate them in details here using Setup 1.

These weights are used in the voting scheme for fusion, and consequently we will compare the results obtained for the presented fusion techniques. After applying this to the experiments in Setup 1, we get the fusion recognition rates (%) as in Table III and Fig. 8. Note that schemes F1-F6 combine recognizers of the same [feature extraction +

TABLE IV
RECOGNITION RATES (%) OF EXPERIMENTS USING SETUP 2 [4], [44], [45], [47] AND [48]

	R1	R2	R3	R4	R5	R6	R7	R8	R9	9Fus (UWV)	9Fus (WV)	Huang *et al.* [4]	Dupuis *et al.* [44]	Jeevan *et al.* [45]*	Yogarajah *et al* [47]	Jiménez *et al.* [48]
Ts21	99.19	**99.60**	77.82	94.76	95.57	91.13	94.36	93.95	96.77	**99.60**	**99.60**	99	98.79	93.36	97.20	99.50
Ts22	43.15	54.03	50.00	46.77	54.03	85.08	37.50	46.77	88.71	87.50	91.93	64	**92.74**	22.44	63.30	78.80
Ts23	88.71	88.71	51.21	79.03	84.27	53.23	74.19	72.18	60.48	**98.39**	**97.58**	72	77.82	56.12	91.90	96.80
Avg.	77.02	80.78	59.68	73.52	77.96	76.48	68.68	70.97	81.99	**95.16**	**96.37**	78.33	89.78	57.31	84.13	77.02

*Experiments carried out on 98 subjects only
Ts21 (*nm1*, *nm2), Ts22 (cl1 & cl2)* and *Ts23 (bg1 & bg2)*

TABLE V
FUSION RECOGNITION RATES (%) OF EXPERIMENTS USING SETUP 2 [4], [44], [45], [47] AND [48]

	3Fus123 (UWV)	3Fus123 (WV)	3Fus456 (UWV)	3Fus456 (WV)	3Fus789 (UWV)	3Fus789 (WV)	9Fus (UWV)	9Fus (WV)	Huang *et al.* [4]	Dupuis *et al.* [44]	Jeevan *et al.* [45]	Yogarajah *et al* [47]	Jiménez *et al.* [48]
Ts21	**99.60**	**99.60**	97.98	97.18	98.79	98.79	**99.60**	**99.60**	99	98.79	93.36	97.20	99.50
Ts22	57.26	60.89	64.52	62.50	69.35	89.52	87.50	91.93	64	**92.74**	22.44	63.30	78.80
Ts23	90.60	91.94	82.66	85.08	79.44	78.63	**98.39**	97.58	72	77.82	56.12	91.90	96.80
Avg.	82.49	84.14	81.72	81.59	82.53	88.98	95.16	**96.37**	78.33	89.78	57.31	84.13	91.70

TABLE VI
RECOGNITION RATES (%) OF EXPERIMENTS USING SETUP 3 [6]

	R1	R2	R3	R4	R5	R6	R7	R8	R9	9Fus (UWV)	9Fus (WV)	Wang *et al.* [6]
Ts31	99.19	**100**	78.23	96.77	97.58	93.55	96.77	98.39	98.39	**100**	**100**	100
Ts32	89.52	87.90	52.42	79.03	82.26	55.65	79.03	82.26	68.55	**99.19**	97.58	68.52
Ts33	91.94	90.32	58.06	83.07	85.48	57.26	78.23	77.42	61.29	**99.19**	98.39	75.00
Ts34	41.94	54.84	50.81	41.94	50.81	87.90	38.71	46.77	**92.74**	90.32	90.32	49.07
Ts35	44.35	60.48	54.84	54.84	62.10	88.71	46.77	53.23	**94.36**	91.94	91.94	44.44
Average	73.39	78.71	58.87	71.13	75.65	76.61	67.90	71.61	83.07	**96.13**	95.65	67.41

Ts31 (*nm6), Ts32 (bg1), Ts33 (bg2), Ts34 (cl1)* and *Ts35 (cl2)*

TABLE VII
FUSION RECOGNITION RATES (%) OF EXPERIMENTS USING SETUP 3 [6]

	3Fus123 (UWV)	3Fus123 (WV)	3Fus456 (UWV)	3Fus456 (WV)	3Fus789 (UWV)	3Fus789 (WV)	9Fus (UWV)	9Fus (WV)	Wang *et al.* [6]
Ts31	99.19	**100**	98.39	98.39	**100**	**100**	**100**	**100**	100
Ts32	90.32	91.94	82.26	85.48	83.87	83.87	**99.19**	97.58	68.52
Ts33	91.94	95.16	82.26	86.29	86.29	82.26	**99.19**	98.39	75.00
Ts34	60.48	60.48	71.77	62.10	73.39	**90.32**	**90.32**	**90.32**	49.07
Ts35	63.71	65.32	75.81	71.77	74.19	**94.35**	91.94	91.94	44.44
Average	81.13	82.58	82.10	80.81	83.55	90.16	**96.13**	95.65	67.41

classifier] pair, while varying gait representation techniques; whereas schemes F1*-F6* combine recognizers of the same representation techniques while varying [feature extraction + classifier] pairs.

Based on the fusion experiments, the following is concluded:

1) In the case of 3-decision fusion; fusing decisions of different representations using the same [feature extraction + classifier] pair performed better than fusing decisions from different classifiers. We will make use of this in the succeeding experimental setups.
2) In average, the Weighted Voting (WV) method performed better than the Unweighted Voting (UWV) method, as seen in 5 out of the 7 experiments in Table III.
3) In the 9-decision fusion schemes (*9Fus*), each of the two tested schemes (unweighted and weighted) performed well, while WV gives higher average classification rate.

2) Results using experimental setup 2

Table IV summarizes the recognition rates (%) obtained for the experiments carried out using Training Data *Tr2* and Testing Data *Ts21*, *Ts22* and *Ts23*. Sequences with *bag* and *coat* are still unseen during training. The difference here is one additional normal sequence (*nm4*) in the training data. We compare our results with those obtained by Huang *et al*. in [4], Dupuis *et al*. in [44], Jeevan *et al*. in [45], Yogarajah *et al*. in [47] and Jiménez *et al.* in [48]

TABLE VIII
RECOGNITION RATES (%) OF EXPERIMENTS USING SETUP 4 [5]

	R1	R2	R3	R4	R5	R6	R7	R8	R9	9Fus (UWV)	9Fus (WV)	Kumar *et al*. [5]
Ts41	96.77	**98.39**	66.94	94.36	93.55	95.97	90.32	94.35	**98.39**	**99.19**	**99.19**	93.33
Ts42	**100**	99.19	58.06	95.97	92.74	95.16	87.90	90.32	98.39	**100**	**100**	96.66
Ts43	**99.19**	96.77	66.94	94.36	93.55	91.94	87.10	91.94	97.58	**100**	**100**	92.50
Ts44	98.39	**99.19**	63.71	93.55	90.32	87.10	84.68	87.90	93.55	**100**	**100**	76.66
Ts45	85.48	95.16	56.45	84.68	85.48	95.97	70.97	77.42	**97.58**	99.19	**100**	79.16
Average	95.97	97.74	62.42	92.58	91.13	93.23	84.19	88.39	97.10	99.68	**99.84**	87.66

Ts41 (*nm4), Ts42 (nm5), Ts43 (nm6), Ts44 (bg2)* and *Ts45 (cl2)*

TABLE IX
FUSION RECOGNITION RATES (%) OF EXPERIMENTS USING SETUP 4 [5]

	3Fus123 (UWV)	3Fus123 (WV)	3Fus456 (UWV)	3Fus456 (WV)	3Fus789 (UWV)	3Fus789 (WV)	9Fus (UWV)	9Fus (WV)	Kumar *et al*. [5]
Ts41	**99.19**	**99.19**	97.58	98.39	95.97	97.58	**99.19**	**99.19**	93.33
Ts42	99.19	99.19	97.58	97.58	96.77	98.39	**100**	**100**	96.66
Ts43	97.58	96.77	98.39	98.39	97.58	97.58	**100**	**100**	92.50
Ts44	99.19	99.19	96.77	97.58	96.77	97.58	**100**	**100**	76.66
Ts45	92.74	95.97	95.97	95.97	93.55	99.19	99.19	**100**	79.16
Average	97.58	98.06	97.26	97.58	96.13	98.06	99.68	**99.84**	87.66

TABLE X
RECOGNITION RATES (%) FROM FUSION OF R1, R2 AND R9 USING SETUP 1 [3] AND [46]

	3Fus129 (WV)	9Fus(WV)	Zhang *et al*. [3]	Zeng *et al*. [46]
Ts11	98.93	**99.19**	98.40	98.40
Ts12	86.29	**95.97**	91.95	93.50
Ts13	79.84	88.31	72.19	**90.30**
Average	88.35	**94.49**	87.51	94.07

TABLE XI
RECOGNITION RATES (%) FROM FUSION OF R1, R2 AND R9 USING SETUP 2 [4], [44], [45], [47] AND [48]

	3Fus129 (WV)	9Fus(WV)	Huang *et al*. [4]	Dupuis *et al*. [44]	Jeevan *et al*. [45]	Yogarajah *et al* [47]	Jiménez *et al*. [48]
Ts21	**99.60**	**99.60**	99	98.79	93.36	97.20	99.50
Ts22	86.69	91.93	64	**92.74**	22.44	63.30	78.80
Ts23	92.73	**97.58**	72	77.82	56.12	91.90	96.80
Average	**93.01**	**96.37**	78.33	89.78	57.31	84.13	91.70

It is still noticeable that high recognition rates are possible without fusion using the two recognition approaches of *R1* and *R2*, in both cases of *normal* sequences (*Ts21*) and *bag* sequences (*Ts23*). Again, edge-masked AEI gait representation has worked well for *coat* sequences (*Ts22*), especially with the MPCA+LDA algorithm *R9*. The rate in [44] still exceeds this one by more than 4%. But it is higher than the rate obtained in *9Fus (WV)* by less than 1%.

General behavior of all base methods is similar to that in Setup 1. Average 9-decision fusion results (WV) are approximately 6% higher than the highest of the other methods.

Table V shows the fusion results using Setup 2. It can be seen that 3-decision fusion techniques are more effective here compared to the more onerous Setup 1. Again, in average, WV performs better than UWV.

3) Results using experimental setup 3

Table VI summarizes the recognition rates obtained for the experiments carried out using Training Data *Tr3* and Testing Data *Ts31*, *Ts32*, *Ts33,Ts34* and *Ts35*. Note that one more normal sequence (*nm5*) is added to the training data. We compare our results with those obtained by Wang *et al*. in [6].

Most of the base methods have performed well. It can be noticed that one of these methods, *R2* has outperformed the target in all the five testing experiments. Fusion has improved this very dramatically. We can see in the case of *coat* (*Ts34* and *Ts35*) our fusion rates are approximately 50% higher than those reported in [6]. Table VII shows the fusion results using Setup 3. It can be seen that all 3-decision fusion techniques outperforms the target results and score very well. As expected, 9-decision fusion improves this further.

TABLE XII
RECOGNITION RATES (%) FROM FUSION OF R1, R2 AND R9 USING SETUP 3 [6]

	3Fus129 (WV)	9Fus (WV)	Wang *et al*. [6]
Ts31	**100**	**100**	100
Ts32	91.94	**97.58**	68.52
Ts33	95.97	**98.39**	75.00
Ts34	**90.32**	**90.32**	49.07
Ts35	**91.94**	**91.94**	44.44
Average	**94.03**	**95.65**	67.41

TABLE XIII
RECOGNITION RATES (%) FROM FUSION OF R1, R2 AND R9 USING SETUP 4 [5]

	3Fus129 (WV)	9Fus (WV)	Kumar *et al*. [5]
Ts41	**100**	99.19	93.33
Ts42	99.19	**100**	96.66
Ts43	99.19	**100**	92.50
Ts44	99.19	**100**	76.66
Ts45	97.58	**100**	79.16
Average	**99.03**	**99.84**	87.66

TABLE XIV
TOTAL PROCESSING TIMES OF THE PROPOSED WEIGHTED VOTING FUSION METHODS

Method	3Fus123	3Fus456	3Fus789	3Fus129	9Fus
Time (sec)	10.537	47.975	96.177	49.953	179.927

4) Results using experimental setup 4

Table VIII summarizes the recognition rates obtained for the experiments carried out using Training Data *Tr4* and Testing Data *Ts41*, *Ts42*, *Ts43,Ts44* and *Ts35*. This is the only setup in which training data has samples from all three groups, *normal*, *bag* and *coat*. It is expected that this arrangement should enhance the final overall recognition rate, especially when testing for *bag* and *coat* sequences. We compare our results with those obtained by Kumar *et al*. in [5].

Exactly as expected, recognition rates have topped high due to training for the two variables of carrying condition (*bag*) and clothing (*coat*). Most of our methods, especially *R1*, *R2* and *R9*, have outperformed those in [5]. This is especially noticeable in the two cases of *Ts44* (*bg2*) and *Ts45* (*cl2*). Edge-masked AEI is apparently not so efficient when combined with image projection and LDF (*R3*). However, it performs very well with MPCA+LDA (*R9*). Table IX shows the fusion results using Setup 4. 3-decision fusion techniques also perform very well, and they all outperform the target results. Almost all of the recognition rates achieved by 9-decision fusion are at top 100%. This table also shows, through all average values, that Weighted Voting (WV) performs better compared to Un-Weighted Voting (UWV).

In general, the unweighted and weighted 9-decision voting schemes outperform all the recently published methods that use state-of-the-art algorithms. While this is the case, fusion may not be always needed. In most of the tests presented above, two of the base methods, *R1* and *R2*, have also individually exceeded the target results. To recall, these methods use the API and AFI gait representations respectively, using the simple image projection for feature extraction and linear classifier. The edge-masked AEI gait representation has performed well for the *coat* sequences, particularly when combined with the MPCA+LDA algorithm. As can be noticed, experimental setups gradually decrease in difficulty moving from Setup 1 to Setup 4. 3-decision fusion techniques perform very well in Setup 4. They also outperform the target results in Setup 3, and with less percentage for Setup 2. In Setup 1, however, we need 9-decision fusion to achieve our targets.

It should be also noted that the fusion results in Setups 2-4 were achieved using the weights from Setup 1. These might be optimized further, and results improved, by utilizing the additional training dataset for validation.

As a final test, and based on the findings of our experiments above, we will evaluate another 3-decision fusion combination using base recognizers *R1*, *R2* and *R9*. These 3 base methods yielded the smallest errors from Table II, and the best average individual recognition rates in the four setups. We will also concentrate on the Weighted Voting (WV) scheme only. The method is labeled 3Fus129. Tables X-XIII list these final results and compare them to the 9-decision fusion rates and the published rates. We can see this 3-decision fusion method outperforms almost all the target results in the four setups. Obviously, the 9-decision fusion enhances the rates further.

As said earlier in Section I, there are also drawbacks of the fusion methods, generally higher processing time and computational complexity. To further highlight this point we used the Matlab functions *tic* and *toc* to simulate the processing power of each of the proposed fusion method. The PC used during the test has: Intel Core i5 processor, 8GB Memory and 5400RPM HDD. In this case, we report the total processing time required for pre-processing, feature extraction

and classification of one single object, being bag sequence 1 (*bg1*). Table XIV summarizes these processing times. We notice that the tensor-based multilinear subspace learning methods (using MPCA and MPCALDA) are more computationally complex than other methods, which is expected. We suggest that the method giving the best trade-off between recognition rates and computational complexity is *3Fus129*.

E. Robustness Experiments using OU-ISIR database

1) OU-ISIR Treadmill Dataset B

As an additional robustness test, we carried out several experiments using the OU-ISIR database, particularly the 'Treadmill' database. The Treadmill database tests for different variations including; view variation (Dataset A), cloth variation (Dataset B), speed variation (Dataset C) and gait fluctuations (Dataset D). We chose to use Dataset B, which is more relevant to this paper. The dataset, explained in [49], comprises 68 subjects captured from side view, walking on treadmill, with up to 32 different combinations of clothes as shown in Fig. 9. This dataset was also used in some of the recent experiments on gait recognition as in [50], [52] and [53].

The dataset is divided into three parts:

1. Training Subset: contains 20 subjects with 15-28 cloth variations. We did not use this subset, as we wish to test our pre-tuned system using the same parameters from previous experiments
2. Gallery Subset: contains 48 subjects with one cloth type (9).
3. Probe Subset: contains 48 subjects with the rest cloth types, resulting into 856 samples.

2) Experiments and results

The first observation on the dataset is that it comprises normalized silhouettes of the subjects only. To recall, two of the three gait representations we used in this paper; API and AFI, are based on the original frames as inputs, in order to accumulate differences producing strip-like images that preserve the spatial information. These frames are not available in the dataset, thus the two mentioned methods could not be used in this test. The only method that could be used is the non-optimized EMAEI. It was still worth testing how the system would generalize and see the effect of fusion as well.

Besides, there is only one sample per subject in the gallery, which means that the number of samples (48) is less than the number of features (100) in the linear feature matrix. Therefore, we could not use the LDF classifier and had to replace it with the KNN (N=1) classifier.

Fig. 9: OU-ISIR Treadmill Dataset B sample clothing types

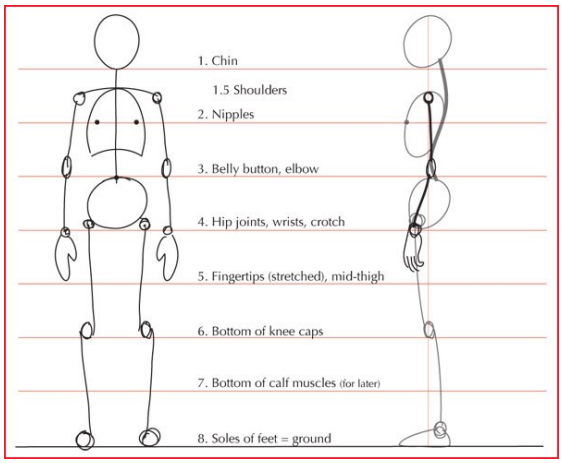


Fig. 10: Human Body Proportions (Courtesy www.tutsplus.com)

With this, the experiments we have undertaken in this section are as follows:

a) Experiments using EMAEI with mask that extends from parts 2 to 5 of the body proportions

An edge-mask similar to the one described in section II-C is implemented here, after dividing each silhouette image into 8 equal proportions (Fig 10). The lines that mark the beginning of 2nd proportion (chin) and the end of 5th proportion (mid-thigh) are 16 and 80 respectively, and thus are used in this experiment. See Fig. 11-a for example. This experimental set is labeled *EMAEI-25*, and it comprises three experiments, namely:

(1) EMAEI + Image Projection + KNN Classifier; labeled R3* to differentiate it from R3 used in previous experiments (replacing LDF with KNN for classification)

(2) EMAET + MPCA + KNN Classifier - (R6)

TABLE XIV
RECOGNITION RATES (%) FROM EXPERIMENTS ON OU-ISIR DATASET B USING 1 GALLERY SEQUENCE

	R3*	R6	R9	3FusWV	3FusWV*	3FusUWV	Arora *et al.* [50]
EMAEI-15	27.78	53.34	2.11	22.74	53.34	34.11	61.20
EMAEI-25	29.07	60.84	1.64	23.80	60.84	36.81	
AEI	29.43	**66.94**	1.64	24.97	**66.94**	38.57	
Average	28.76	60.37	1.80	23.84	60.37	36.50	-

TABLE XV
RECOGNITION RATES (%) FROM EXPERIMENTS ON OU-ISIR DATASET B USING 5 GALLERY SEQUENCES

	R3*	R6	R9	3FusWV	3FusWV*	3FusUWV
EMAEI-15	35.25	62.48	25.87	42.21	62.48	47.81
EMAEI-25	37.52	71.10	21.79	44.78	71.10	52.50
AEI	45.99	**83.21**	30.26	54.92	**83.21**	60.51
Average	39.59	72.26	25.97	47.30	72.26	53.61

We also reduced the silhouettes resolution similar to the work in section III-C, resulting into a gallery matrix of size 64x44x48 and probe matrix of size 64x44x856

(3) EMAET + MPCALDA + KNN Classifier - (R9)
The same resolution reduction also applies here.

b) Experiments using EMAEI with mask that extends from parts 1 to 5 of the body proportions

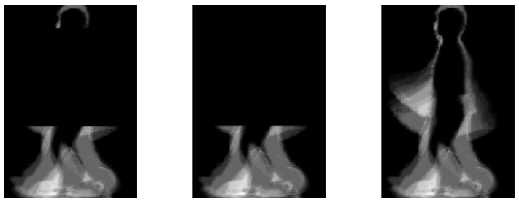
It was noticed that some of the cloth types in the OU-ISIR dataset B include various types of head cover in different combinations. Trying to eliminate the effect from this part of the body, we tested to extend the mask to include proportions 1 to 5, or pixels 1 to 80 in the corresponding silhouette image, keeping only the legs swing. See Fig. 11-b for example. This experimental set is labeled *EMAEI-15*.

c) Experiments using AEI

This is the basis method for the EMAEI, described by Zhang *et al*. in [3]; we repeated the same experiments on AEI for comparison. See Fig. 11-c for example.

The results from these experiments are listed in Table XIV, and compared to the results reported by Arora *et al*. in [50], who used Gait Information Image with Sigmoid Feature (GII-SF) method, which is based on the information set theory.

We notice that R6, which is based on tensorial data and the non-linear MPCA algorithm, has yielded the best results among the three individual methods. And this is in line with the results on CASIA B database, where the average results using MPCA were the highest compared to the rest methods. Method EMAEI-25 was slightly better than EMAEI-15, which means that there are still discriminating features in the head part. Recall that EMAEI-25 is the same method used in CASIA B experiments. The results using the basis method AEI are still better. As noted earlier, the Masked AEI methods implemented in this paper are based on a simple approach of manual masking and must be improved, by using more advanced feature selection criteria, such as those introduced in [44] and [55].



(a) (b) (c)

Fig. 11: (a) EMAEI-25 of subject 94, cloth type 9, (b) EMAEI-15 of the same subject and cloth type, (c) AEI of the same subject and cloth type.

d) Decision Fusion Experiments

We tested here three different fusion methods:

(1) Weighted Decision Fusion using weights from Table II

These weights are 0.049, 0.119 and 0.128 corresponding to methods R3*, R6 and R9 respectively. Recall that these weights are calculated from testing the different methods at normal walking conditions, using Setup 1, on CASIA B Dataset. At those controlled and limited conditions, R9 scored the highest rate, which was not the case on OU-ISIR Treadmill B dataset. This led to poor fusion results, as seen in Table XIV under 3FusWV, which indicates the need for more experiments on different conditions in order to better estimate the fusion weights.

(2) Weighted Decision Fusion using new weights

In these experiments we use a different set of weights estimated from the performance on the OU-ISIR Treadmill B dataset. These are 0.3, 0.6 and 0.1 corresponding to R3*, R6 and R9 respectively. And the fusion results using these weights are shown in Table XIV under 3FusWV*. We notice that these weights have dramatically improved the rates, matching the individual rate of the MPCA-based method, R6.

(3) Unweighted Decision Fusion

Finally, fusion of the three decisions without weights yielded the rates under 3FusUWV, which as seen are between those yielded from the two previous fusion methods.

e) Experiments using five gallery sequences (cloth types)

As additional experiments to test the impact of adding gallery samples, we added four more sequences to the gallery set, selecting those that are common to all subjects; cloth types A, B, C, 5 and 9 resulting into a linear feature matrix of size 240x100. The rest of sequences go in the probe set with feature matrix of size 664x100. And the tensorial data, after resolution reduction, become 64x44x240 for gallery and 64x44x664 for probe.

After repeating exactly the same experiments from *a* through *d* above, we get the results in Table XV. We notice that the behavior is still the same, with higher rates in all experiments resulted from the additional samples.

VI. Conclusions and Future Work

Major part of the research work on gait recognition is dedicated to overcome the challenges associated with covariates such as carrying and clothing conditions. In the outcome of our literature review, and based on the results of the experimental analysis, it is concluded that model-free gait approaches, and particularly spatio-temporal (accumulated error) and energy-based methods can perform well.

The focus of this work has been on single-view gait recognition with varying clothing and carrying conditions. Hence, the use of this part of the CASIA-B dataset and OU-ISIR Dataset B. In future work other applications should be explored where multi-view is needed.

Nine different gait approaches have been proposed and evaluated. Three gait representation methods are used; Accumulated Prediction Image (API) and two new representation techniques namely, Accumulated Flow Image (AFI) and Edge-Masked Active Energy Image (EMAEI). It was shown that the first two methods are effective in producing distinctive features. The EMAEI method, although requires further optimization, has performed well when tested with *coat* sequences on CASIA B dataset as it removes the parts of the body that correspond to static features. Therefore, it contributes in the final fusion scheme.

Three methods have been evaluated for feature extraction. Image projection with 1D DCT, and two multilinear techniques namely MPCA and MPCA+LDA. The two multilinear techniques are compatible with the tensorial nature of the gait features, being 3-mode data. The advantage was to keep data in its raw nature. Each of the three gait representation images was tested with each of the feature extraction methods, yielding a model of nine different base recognizers.

As future work, the different parameters obtained empirically in this work may still be optimized by further testing and experimenting. More advanced methods of feature selection could be tested that may reduce the effect of static features in gait images.

It is shown that fusion can combine the base classifiers' discriminating power and boost the overall system's performance. Decision-level fusion is simple and straightforward compared to feature-level and score-level fusion techniques. It is concluded that fusion is recommended for gait recognition. Based on a comparison between unweighted and weighted voting fusion schemes, it was found that weighting can improve the classification accuracy, provided previous knowledge of the individual methods' performance. As for the two voting schemes tested in this work; Unweighted Voting (UWV) and Weighted Voting (WV), the WV method has shown improvements in majority of the fusion tests carried out, while UWV is still an easy and safe choice if individual performances are ignored or unknown. On the other hand, further testing would be required to improve the computational power of the proposed fusion methods.

Acknowledgment

The authors would like to thank: the Chinese Academy of Science and Institute of Automation for making their CASIA B dataset available for researchers, Makihara *et al*. [49] for providing the OU-ISIR Treadmill B dataset, and Lu *et al*. for providing their MPCA algorithm, as explained and implemented in [40].

References

[1] J. Little and J. Boyd, "Biometric Gait Recognition," in *Advanced Studies in Biometrics*, *Lecture Notes in Computer Science,* vol.3161, pp. 19–42, 2005

[2] S. Sarkar, P. Phillips, Z. Liu, I. Vega, P. Grother, and K. Bowyer, "The HumanID gait challenge problem: Data sets, performance and analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 27, no. 2, pp. 162–177, Feb. 2005.

[3] E. Zhang,Y. Zhao and W. Xiong, "Active energy image plus 2DLPP for gait recognition," *Journal of Signal Processing*, vol. 90, no. 7, pp. 2295–2302, 2010

[4] X. Huang and N. Boulgouris, "Gait Recognition With Shifted Energy Image & Structural Feature Extraction," *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 2256-2268, 2012

[5] H. Kumar and H. Nagendraswamy, "Gait Recognition- An Approach Based on Interval Valued Features," in *Proc. International Conference on Computer Communication and Informatics (ICCCI)*, Coimbatore, India, pp. 1-6, Jan. 04-06, 2013

[6] C. Wang, J. Zhang, L. Wang, J. Pu and X. Yuan. "Human Identification Using Temporal Information Preserving Gait Template," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 11, Nov. 2012

[7] V. Von Tscharner and B. Goepfert, "Gender dependent EMGs of runners resolved by time/frequency and principal pattern analysis," *Journal of Electromyography and Kinesiology,* vol.13, no. 3, pp 253–272, 2003

[8] L. Wang, T. Tan, H. Ning and W. Hu, "Fusion of static and dynamic body biometrics for gait recognition," *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 14, no. 2, pp. 149–158, Feb. 2004.

[9] L. Lee, and W. Grimson, "Gait Analysis for Recognition and Classification," in *Proc. Fifth IEEE International Conference on Automatic Face and Gesture Recognition,* Washington, DC, USA, pp. 148–155, 2002

[10] D. Cunado, M. Nixon, and J. Carter, "Using Gait as a Biometric, via Phase-Weighted Magnitude Spectra," in *Proc. First International*

Conference on Audio- and Video-Based Biometric Person Authentication(AVBPA), London, UK, pp. 95-102, 1997

[11] J. Yoo, and M. Nixon, "Markerless Human Gait Analysis via Image Sequences," in *Proc. International Society of Biomechanics XIXth Congress,* Dunedin, New Zealand, 2003

[12] C. Yam, M. Nixon, and J. Carter, "Automated Person Recognition by Walking and Running via Model-based Approaches," *Pattern Recognition Journal,* vol. 37, no. 5, pp. 1057–1072, 2004

[13] S. Dockstader, M. Berg and A. Tekalp, "Stochastic Kinematic Modeling and Feature Extraction for Gait Analysis," *IEEE Transactions on Image Processing,* vol. 12, no. 8, pp. 962–976, 2003

[14] A. F. Bobick, and A. Y. Johnson, "Gait recognition using static, activity-specific parameters," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition,* Kauai, HI, USA, vol. 1, pp. 423-430, 2001

[15] C. BenAbdelkader, R. Cutler, and L. Davis, "Stride and cadence as a biometric in automatic person identification and verification," in *Proc. Fifth IEEE International Conference on Face and Gesture Recognition,* Washington, DC, USA, *pp. 372–377*, 2002

[16] R. Tanawongsuwan, and A. Bobick, "Gait recognition from time-normalized joint-angle trajectories in the walking plane," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition,* Kauai, HI, USA, vol. 2,pp. 726–731, 2001

[17] R. Zhang, C. Vogler, and D. Metaxas, "Human gait recognition," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition,* Washington, DC, USA,vol. 1, pp. 18, 2004

[18] J. Fortuny-Guasch, P.F. Sammartino, and J. Petit, "Radar techniques for human gait automatic recognition," in *Proc. International Carnahan Conference on Security Technology,* Zurich, Switzerland, *pp. 221-226, 2009*

[19] S. Niyogi and E. Adelson, "Analyzing and Recognizing Walking Figures in XYT," in *Proc.IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Seattle, Wash, USA,* pp. 469–474, 1994

[20] J. Little, and J. Boyd, "Recognizing People by Their Gait: The Shape of Motion," *Videre,* vol. 1, no. 2, pp. 1-33, 1998.

[21] C. BenAbdelkader, R. Cutler, H. Nanda and L. Davis, "Eigen Gait: motion-based recognition of people using image self-similarity, Audio and Video-based Biometric Person Authentication," *Lecture Notes in Computer Science,* vol. 2091, pp. 284-294, 2001.

[22] Z. Liu, and S. Sarkar, "Simplest representation yet for gait recognition: Averaged silhouette," in *Proc. 17th International Conference on Pattern Recognition,* Cambridge, England, UK, vol. 4, pp. 211–214, 2004

[23] J. Han, and B. Bhanu, "Statistical feature fusion for gait-based human recognition," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition,* Washington, DC, USA, vol. 2, pp. 842–847, 2004

[24] J. Han, and B. Bhanu, "Individual Recognition Using Gait Energy Image," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 28, no. 2, pp. 316–322, 2006.

[25] C. Chen, J. Liang and H. Zhao, 'Frame difference energy image for gait recognition with incomplete silhouettes," *Pattern Recognition Letters*, vol. 30, pp. 977-984, 2009

[26] T. Shanableh, K. Assaleh, L. Al-Hajjaj and A. Kabani, "Gait Recognition System Tailored for Arab Costume of The Gulf Region," in *Proc. IEEE International Symposium on Signal Processing and Information Technology (ISSPIT),* Ajman, UAE, pp. 544-549, 2009

[27] T. H.W. Lam, K.H. Cheung and J.K. Liu, "Gait flow image - A silhouette-based gait representation for human identification," *Pattern Recognition Journal*, vol. 44, pp. 973-987, 2011

[28] A. Jain, and A. Ross, "Multibiometric systems," in *Communications of the ACM - Multimodal interfaces that flex, adapt, and persist*, vol. 47, no. 1, pp. 34-40, Jan 2004.A. Ross, and A. K. Jain, "Information fusion in biometrics," *Pattern Recognition Letters*, vol. 24, pp. 2115-2125, Sep 2003.

[29] A. Ross, and A. K. Jain, "Information fusion in biometrics," *Pattern Recognition Letters*, vol. 24, pp. 2115-2125, Sep 2003.

[30] S. Zheng, K. Huang, T. Tan, and D. Tao. "A cascade fusion scheme for gait and cumulative foot pressure image recognition," *Pattern Recognition Journal*, vol. 45, no. 10, pp. 3603-3610, 2012.

[31] R. Chellapa, A. K. Roy-Chowdhury and A. Kale, "Human Identification using Gait and Face," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition,* Minneapolis, MN, USA, pp. 1-2, 2007

[32] E. Hossain, and G. Chetty, "Multimodal Face-Gait Fusion for Biometric Person Authentication," in *Proc. IFIP 9th International Conference on Embedded and Ubiquitous Computing*, Melbourne, Australia, pp. 332-337, 2011

[33] L. Wang, H. Ning, T. tan and W. Hu, "Fusion of Static and Dynamic Body Biometrics for Gait Recognition," *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 14, no. 2, pp. 149-158, 2003

[34] Y. Zhang, S. Jiang, Z. Yang, Y. Zhao and T. Guo, "A Score Level Fusion Framework for Gait-based Human Recognition," in *Proc.15th IEEE International Workshop on Multimedia Signal Processing* (*MMSP),* Pula, Italy, pp. 189-194, 2013

[35] L. Lam and C. Suen, "Application of majority voting to pattern recognition: an analysis of its behavior and performance," *IEEE Transactions on System, Man and Cybernetics*, vol. 27, no. 5, pp. 553-568, 1997

[36] L. Xu, A. Krzyak and C.Y. Suen, "Methods of combining multiple classifiers and their applications to handwriting recognition," *IEEE Transactions on System, Man and Cybernetics*, vol. 22, no. 3, pp. 418-435, 1992

[37] R. Maclin, and J. Shavlik, "Combining the Prediction of Multiple Classifiers: Using Competitive Learning to Initialize Neural Networks," in *Proc. 14th International Joint Conference on Artificial Intelligence*, Montreal, Quebec, Canada, 1995

[38] G. Rogova and R. Menon," Decision Fusion for Learning in Pattern Recognition," in *Proc.FUSION'98, First International Conference on Multisource-Multisensor Information Fusion,* Las Vegas, Nevada, USA, 1998

[39] G. Rogova, "The Dempster-Shafer theory of evidence for classification of medical images," in *Proc. the 6th IEEE Dual-Use Technologies and Application Conference*, Syracuse, NY, USA, 1996

[40] H. Lu, K. Plataniotis and A. Venetsanopoulos, "MPCA: Multilinear Principal Component Analysis of tensor Objects," *IEEE Transactions on Neural Networks*, vol. 19, no. 1, pp. 18-39, Jan 2008.

[41] P. O'Donovan, "Optical Flow Techniques and Applications," *The University of Saskatchewan*, Apr 6, 2005

[42] B. Horn, and B. Schunck, "Determining Optical Flow," *Artificial Intelligence Journal*, vol. 17, no. 1-3, pp. 185-203, 1981

[43] J. Han and B. Bhanu, "Individual recognition using gait energy image," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 2, pp. 316-322, Feb. 2006

[44] Y. Dupuis, X. Savatier and P. Vasseur, "Feature subset selection applied to model-free gait recognition," *Image and Vision Computing Journal*, vol. 31, no. 8, pp. 580-591, 2013

[45] M. Jeevan, N. Jain, M. Hanmandlu, and G. Chetty. "Gait Recognition Based on Gait Pal and Pal Entropy Image," in *Proc. 20thIEEE International Conference on Image Processing (ICIP),* Melbourne, Australia, pp. 4195-4199, 2013

[46] W. Zeng, C. Wang and F. Yang, "Silhouette-based gait recognition via deterministic learning," *Elsevier Pattern Recognition Journal*, vol. 47, pp. 3568–3584, 2014

[47] P. Yogarajah, P. Chaurasia, J. Condell and G. Prasad, "Enhancing gait based person identification using joint sparsity model and l1-norm minimization," *Elsevier Information Sciences Journal*, vol. 308, pp. 3-22, 2015

[48] M. Jiménez, F. Castro, Á. Poyato and N. Guil, " On how to improve tracklet-based gait recognition systems," *Elsevier Pattern Recognition Letters*, vol. 68, pp. 103–110, 2015

[49] Y. Makihara, H. Mannami, A. Tsuji, Md. Hossain, K. Sugiura, A. Mori and Y. Yagi, " The OU-ISIR Gait Database Comprising the Treadmill Dataset," *IPSJ Transactions on Computer Vision and Applications*, vol. 4, pp. 53-62, 2012.

[50] P. Arora, M. Hanmandlu and S. Srivastava, "Gait based authentication using gait information image features," *Elsevier Pattern Recognition Letters*, vol. 68, pp. 336-342, 2015

[51] A. Nandy, R. Chakraborty and P. Chakraborty, "Cloth invariant gait recognition using pooled segmented statistical features," *Elsevier Neurocomputing Letters*, vol. 196, pp. 117-140, 2016

[52] S. D. Choudhury and T. Tjahjadi, "Clothing and carrying condition invariant gait recognition based on rotation forest," *Elsevier Pattern Recognition Letters*, vol. 80, pp. 1-7, 2016

[53] M. A. Hossain, Y. Makihara, J. Wang and Y. Yagi, "Clothing-invariant gait identification using part-based clothing categorization and adaptive weight control," *Elsevier Pattern Recognition Journal*, vol. 43, pp. 2281-2291, 2010

[54] J. Luo, J. Tang, T. Tjahjadi and X. Xiao, "Robust arbitrary view gait recognition based on parametric 3D human body reconstruction and virtual posture synthesis," *Elsevier Pattern Recognition Journal*, vol. 60, pp. 361-377, 2016

[55] S. D. Choudhury and T. Tjahjadi, "Robust view-invariant multiscale gait recognition," *Elsevier Pattern Recognition Journal*, vol. 48, pp. 798-811, 2016

[56] W. Xu, C. Luo, A. Ji and C. Zhu, "Coupled locality preserving projections for cross-view gait recognition," *Elsevier Neurocomputing Journal*, 2016, in press.

[57] G. Ma, Y. Wang and L. Wu, "Subspace ensemble learning via totally-corrective boosting for gait recognition," *Elsevier Neurocomputing Journal*, 2016, in press.

[58] D. Kastaniotis, I. Theodorakopoulos, G. Economou and S. Fotopoulos, "Gait based recognition via fusing information from Euclidean and Riemannian manifolds," *Elsevier Pattern Recognition Letters*, vol. 84, pp. 245-251, 2016

[59] W. Kusakunniran, Q. Wu, J. Zhang, Y. Ma and H. Li, "A New View-Invariant Feature for Cross-View Gait Recognition," *IEEE Transactions on Information Forensics and Security*, vol. 8, no 10, 2013

[60] M. Deng, C. Wang and Q. Chen, "Human gait recognition based on deterministic learning through multiple views fusion," *Elsevier Pattern Recognition Letters*, vol. 78, pp. 56-63, 2016

[61] T. Connie, K. Goh and A. Teoh, "Multi-view gait recognition using a doubly-kernel approach on the Grassmann manifold," *Elsevier Neurocomputing Journal*, vol. 216, pp. 534-542, 2016

[62] D. L. Fernandez, F. J. M. Cuevas, A. C. Poyato, R. M. Salinas and R. M. Carnicer, "A new approach for multi-view gait recognition on unconstrained paths," *Elsevier Journal of Visual Communication & Image Representation*, vol. 38, pp. 396-406, 2016