ACTIVITY-BASED HUMAN IDENTIFICATION

Tzu-Yi Hung[1], *Jiwen Lu*[2], *Junlin Hu*[1], *Yap-Peng Tan*[1], *and Yongxin Ge*[3]

[1]School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore
[2]Advanced Digital Sciences Center, Singapore
[3]School of Software Engineering, Chongqing University, China

ABSTRACT

We investigate in this paper the problem of activity-based human identification. Different from most existing gait recognition methods where only human walking activity is considered and utilized for person identification, we aim to identify people from various activities such as eating, jumping, and weaving. For each video clip, we first extract binary human body masks by using background substraction, followed by computing the average energy image (AEI) features to represent each video clip. Then, a mapping is learned by applying an adaptive discriminant analysis (ADA) method to project AEI features into a low-dimensional subspace, such that the intra-class (activities performed by the same person) variations are minimized and the interclass (activities performed by different persons) are maximized, simultaneously. Moreover, interclass samples with large similarity difference are deemphasized and those with small difference are emphasized, such that more discriminative information can be used for recognition. Experimental results on three publicly available databases show the efficacy of our proposed approach.

Index Terms— Human identification, gait recognition, human activity analysis.

1. INTRODUCTION

Gait recognition [1, 2, 3, 4] has been widely studied in computer vision and pattern recognition over the past decades due to its promising signatures such as non-invasive and non-contactable characteristics. While many gait recognition methods have been proposed in the literature, few of them consider recognizing people based on other activities besides walking. In many real applications, people may perform other activities such as eating and drinking rather than walking in the scene. A natural question is whether we can identify person from his/her activities rather than walking? In this paper, we investigate this problem and propose an appearance-based learning approach to address this problem.

The research on activity-based human identification is relative new, and to our best knowledge, there are few related works providing efforts on it [5, 6]. In their work, several activities of the same person are used for human identification. Firstly, binary human body masks are extracted, followed by a fuzzy c-means method to quantize each binary mask, and the labelling information is exploited as well to reduce the dimensionality of the feature vectors. Then, different activities are recognized by an activity classifier, and then an activity-specific person classifier is adopted to recognize human identity. While some promising results have been provided, their method requires obtaining correctly activity classification results or the following activity-specific person recognition may fail to work. Moreover, some discriminative information may be lost due to the quantization errors cause by fuzzy c-means.

In this paper, we propose a new approach to activity-based human identification. Similar for the existing work, for each video clip, we extract binary human body masks by using background substraction. Instead of using fuzzy c-means as quantization method, we compute the average energy image (AEI) feature for feature representation which is the similar idea from the gait energy image (GEI), but extends to multiple activities. Then, we learn a mapping to project AEI features into a low-dimensional subspace by applying an adaptive discriminant analysis (ADA) [7] method, such that the intra-class (activities performed by the same person) variations are minimized and the inter-class (activities performed by different persons) are maximized, simultaneously, such that more discriminative information can be exploited for recognition. Figure 1 shows the flow chart of our proposed approach. Experimental results on three publicly available action databases show the efficacy of our proposed approach.

2. PROPOSED APPROACH

2.1. Feature Extraction

Each activity video clip contains one activity of a person, such as waving, boxing, drinking with a cup, and so on. We firstly extract human body silhouettes by using background subtraction, and then align each body mask into 64×48 to obtain the region of interest (ROI) as shown in the first six columns of Figure 2. Subsequently, we obtain the ROI images from video clips. Let $F = \{f_1, \cdots, f_i, \cdots, f_n\}$ be a set of the ROI images from a video clip, where f_i is the ith ROI frame of the binary human body mask, and n is the number of the frames.

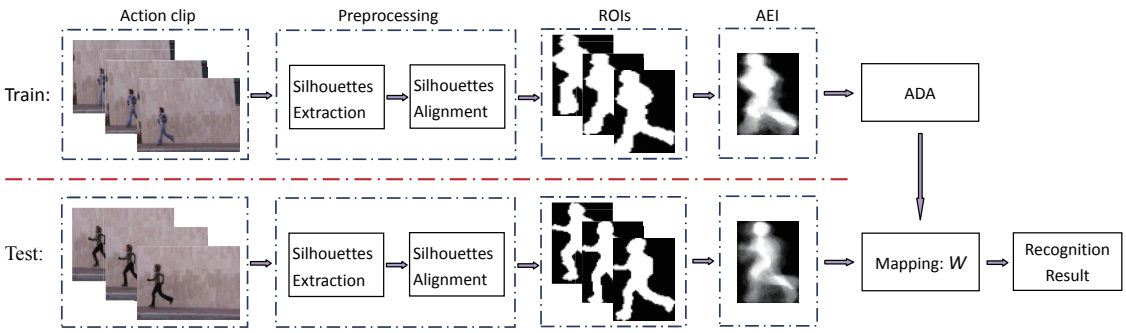

Fig. 1. Flow chart of our proposed approach.

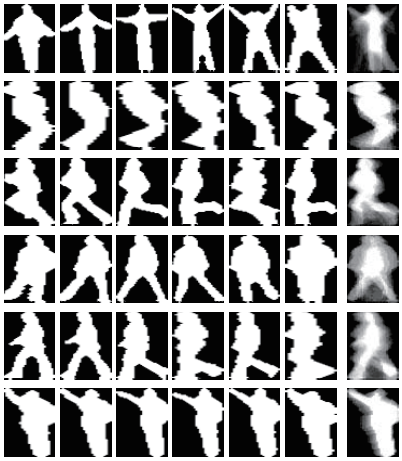

Fig. 2. Six sample ROI sequences and the corresponding AEI images. The rightmost images in each row are the AEIs. From top to bottom, action types are jack, jump, run, side, walk and wave1 (wave-one-hand), respectively.

2.2. Feature Representation

Having obtained a set of the ROI images, we represent each video clip as a feature vector by an average energy image (AEI) which is an extension of the gait energy image (GEI) [8] as follows:

$$AEI(x,y) = \frac{1}{n}\sum_{i=1}^{n} f_i(x,y) \tag{1}$$

where x and y are the coordinates of f_i. The last column of Figure 2 shows AEIs of different activity sequences from the same person. AEI can reflect dynamic information of the activity based on the value of pixels. The higher intensity of a pixel has, the more frequently human activity occurs at this position. Then, we represent each video clip as a vector $x \in R^d$ by concatenating the column pixels of AEI with feature dimension d.

2.3. Adaptive Discriminant Analysis

For activity-based human identification, we want to minimize the intra-variance (i.e., activities performed by the same person) and maximize the inter-variance (i.e., activities performed by different persons). Hence, to enhance the discriminative power, we adopt adaptive discriminant analysis (ADA) [7] to learn a set of projection axes to maximize the Fisher criterion, the ratio of between-class scatter to within-class scatter. Different from the conventional linear discriminant analysis method [9], ADA provides a penalty weight to the classes with high similarity. Let $X_i = [x_{i1}, x_{i2}, \cdots, x_{iz_i}] \in R^{d\times z_i}$ be the samples in the ith class where z_i is the number of samples in this ith class. Let $X = [X_1, X_2, \cdots, X_c]$ be the training set where c is the number of classes. In our scenario, c denotes the number of persons and X_i denotes the various activity samples of the same person. Hence, $N = \sum_{j=1}^{c} z_j$ is the number of the samples in the training set. The ADA subspace learning method is defined as follows:

$$\begin{aligned} \max_{w} \quad & w^T S w \\ s.t. \quad & w^T w = 1 \end{aligned} \tag{2}$$

where $w^T w = 1$ is an orthogonal constraints to well-posed the above maximization problem with respect to w, and $S = S_B - S_W$. S_W is the within-class scatter calculated by the samples and their class mean $\bar{X}$ defined as follows:

$$S_W = \frac{1}{N}\sum_{i=1}^{c}\sum_{j=1}^{z_i}(x_{ij} - \bar{X}_i)(x_{ij} - \bar{X}_i)^T \tag{3}$$

where S_B is the between-class scatter with penalty function $a(i,j)$ to impose different weights to characterize the relation of the ith and jth classes:

$$S_B = \frac{1}{c^2}\sum_{i=1}^{c}\sum_{j=1}^{c} a(i,j)(\bar{X}_i - \bar{X}_j)(\bar{X}_i - \bar{X}_j)^T \tag{4}$$

with a correlation function $a(i,j)$:

$$a(i,j) = \frac{< \bar{X}_i, \bar{X}_j >}{\|\bar{X}_i\|_2 \cdot \|\bar{X}_j\|_2} \quad (5)$$

where operation $\|b\|_2$ denotes the L_2 norm of b. Generally, the larger penalty, $a(i,j)$, should be imposed, when $\bar{X}_i$ and $\bar{X}_j$ are more similar. The problem can then be solved by using Lagrange multipliers, $Q(w) = w^T S w + \lambda(1 - w^T w)$. Compute the gradient of $Q(w)$ with respect to w, and then, ADA can be solved as the eigenvalue equation: $Sw = \lambda w$. Hence, the ADA subspace to be sought is the projection matrix $W = [w_1, w_2, \cdots, w_m]$ where $w_1, w_2, \ldots, w_m$ are the column vectors ordered according to their eigenvalues.

2.4. Identification

For each test video clip, we recognize the identity information of the person from his/her activity by using a nearest neighbor classifier

$$s = \arg\min_i d(y_{x_t}, y_{x_i}) \quad (6)$$

where $y_{x_t} = W^T x_t$ and $y_{x_i} = W^T x_i$ are the low-dimensional representations of the testing sample x_t and each training sample x_i, respectively, d is the Euclidean distance between the low-dimensional representations, and s denotes x_s, the nearest neighbor of the testing sample x_t. Hence, x_t and x_s are expected to be the same person. No doubt, a more sophisticated classifier could be employed to further improve the recognition accuracy. However, the main interest here is to evaluate the genuine discriminatory ability of the extracted feature (AEI feature representation and ADA feature extraction) in our approach.

3. EXPERIMENTAL RESULTS

3.1. Database and Experimental Settings

We performed activity-based human identification experiments on three publicly available human activity databases, weizmann [10], MOBISERV-AIIA [6], and KTH [11]. The weizmann database, as shown in Figure 3, contains 9 persons, and each person performed 10 different activities, including bend, run, skip, walk, gallop-sideways, jumping-jack, jumping-forward-on-two-legs, jump in place-on-two-legs, wave-one-hand, and wave-two-hands, respectively. There are totally 90 video sequences in this database. Since some videos contain two or more cycles of a specific activity performed by some subjects, we break up such videos into several single period activity videos to generate an activity database. In our experiments, we remove the videos of the bend activity class because there is only one activity period in these videos and it is not enough to obtain training/testing samples for this activity. Subsequently, we construct a database of 215 videos in total. For each activity class, one video is randomly selected for training, and the remaining videos for testing.

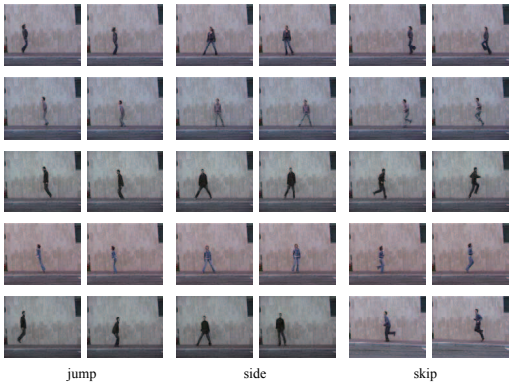

Fig. 3. Three activity examples of five persons. From left to right are image frames of the jump, side, and skip activities, respectively. Images in each row are three different activities of the same person.

Fig. 4. Sample frames of the MOBISERV-AIIA database. The first two columns are the activity of drinking with a cup, and the last two columns are the activity of eating with a fork. Each of the activities contains 2 scenarios.

The MOBISERV-AIIA database, as shown in Figure 4, contains 12 persons, and each person performed 2 different activities, drinking with a cup and eating with a fork, respectively, in four different days. There are 2 scenarios for each person in each day, one wearing a short sleeve top and one wearing a long sleeve top, respectively. We randomly choose one day sequences as testing samples, and the rest sequences are as training samples. The clips are broken up into several single period activity resulting 776 clips in total.

Fig. 5. Sample frames of the KTH database. Each column denotes different activities, and each raw denotes different scenarios.

Table 1. CRR comparison (%) with the existing activity-based human identification method for the weizmann database.

Method	CRR (%)
Method in [5]	55.22
Our method	**92.54**

Table 2. CRR comparison (%) with the existing activity-based human identification method for the MOBISERV-AIIA database.

Method	CRR (%)
Method in [5]	60.14
Our method	**89.00**

Table 3. CRR comparison (%) with the existing activity-based human identification method for the KTH database.

Method	CRR (%)
Method in [5]	31.91
Our method	**35.04**

Table 4. CRR comparison (%) with other subspace learning methods for the weizmann database.

Method	CRR (%)
PCA [12]	89.55
LDA [9]	91.43
LPP [13]	90.86
Our method	**92.54**

The KTH dataset, as shown in Figure 5, contains 25 persons, and each person performed 6 different activities, including boxing, handclapping, handwaving, jogging, running, and walking, respectively. There are 4 scenarios under each activity, including outdoors, outdoors with scale variation, outdoors with different clothes and indoors, respectively, resulting 599 video sequences in total. Due to the diversity of the data, we randomly choose 3 scenarios as training examples for each activity where only the frontal 60 frames of the video sequences are used, and the rest sequences are as testing samples. Experiments are repeated ten times with different randomly selected training and testing samples, and the final results are shown as the mean of the correct recognition rate.

3.2. Results and Analysis

Comparisons with the Existing Activity-Based Human Identification Method: We compare our approach with the method proposed in [5]. We implemented their method ourselves and the number of dynemes is empirically set as 30 in our implementation for the weizmann, 200 for the MOBISERV-AIIA databases and 450 for the KTH databases. The fuzzy parameter $m = 1.1$ was employed to compute the dynemes for all activities. Table 1 compares the correct recognition rates (CRR) of different methods for the weizmann database. As can be seen, our approach significantly outperforms their method with gains in average accuracy of 37.32%. For the MOBISERV-AIIA database, the results are as shown in Table II where our approach significantly outperforms their method as well. For the KTH database, the results of the proposed method are better than those of their method as shown in Table 3, but both of them are low due to the diversity and complexity of the database. Furthermore, as we mentioned before, there is one vector quantization step used in the method in [5], and such quantization will result in some errors such that discriminative information may be lost in their feature representation. Moreover, the dynamic information of human activity is totally ignored in their method and our AEI feature can reflect both dynamic and static shape information. Hence, significantly better recognition performance can be obtained.

Comparisons with Other Subspace Learning Methods: To further investigate the effectiveness of our approach, we also compare our method with three conventional subspace learning algorithms: PCA [12], LDA [9], and locality preserving projections (LPP) [13]. Table 4 tabulates the correct recognition rates (CRR) of different methods. As can be seen from this table, our approach also outperforms PCA, LDA and LPP with gains in average accuracy of 2.99%, 1.11%, and 1.68% respectively, which further indicates the effectiveness and advantage of the used ADA method for our activity-based human identification task.

4. CONCLUSIONS

In this paper, we have proposed an activity-based human identification approach by using average energy image and adaptive discriminant analysis. Different from most existing gait-based human identification methods which only consider human walking as a biometric characteristic, we exploit the discriminative power of different actions for human identification. Our experimental results have clearly demonstrate the feasibility of using different actions to recognize people at a distance and the efficacy of our proposed method.

5. ACKNOWLEDGEMENT

Jiwen Lu is supported by the research grant for the Human Sixth Sense Program at the Advanced Digital Sciences Center from the Agency for Science, Technology and Research (A*STAR) of Singapore.

6. REFERENCES

[1] 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," *PAMI*, vol. 27, no. 2, pp. 162–177, 2005.

[2] C. Wang, J. Zhang, J. Pu, X. Yuan, and L. Wang, "Chrono-gait image: a novel temporal template for gait recognition," in *ECCV*, 2010, pp. 257–170.

[3] J. Lu and E. Zhang, "Gait recognition for human identification based on ica and fuzzy svm through multiple views fusion," *Pattern Recognition Letters*, vol. 28, no. 16, pp. 2401–2411, 2007.

[4] M.S. Nixon and J.N. Carter, "Automatic recognition by gait," *Proceedings of the IEEE*, vol. 94, no. 11, pp. 2013–2024, 2006.

[5] N. Gkalelis, A. Tefas, and I. Pitas, "Human identification from human movements," in *ICIP*, 2009, pp. 2585–2588.

[6] A. Iosifidis, A. Tefas, and I. Pitas, "Activity based person identification using fuzzy representation and discriminant learning," *TIFS*, vol. 7, no. 2, pp. 530–542, 2012.

[7] J. Lu and Y.-P. Tan, "View recognition of human gait sequences in videos," in *ICIP*, 2010, pp. 2457–2460.

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

[9] P. N. Belhumenur, J. P. Hepanha, and D. J. Kriegman, "Eigenfaces vs. fisherface: recognition using class specific linear projection," *PAMI*, vol. 19, no. 7, pp. 711–720, 1997.

[10] L. Gorelick, M. Blank, E. Shechtman, M. Irani, and R. Basri, "Action as space-time shapes," *PAMI*, vol. 29, no. 12, pp. 224–2253, 2007.

[11] C. Schuldt, I. Laptev, and B. Caputo, "Recognizing human actions: a local svm approach," in *ICPR*, 2004, vol. 3, pp. 32–36.

[12] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.

[13] X. He, S. Yan, Y. Hu, P. Niyogi, and H. J. Zhang, "Face recognition using laplacianfaces," *PAMI*, vol. 27, no. 3, pp. 328–340, 2005.