Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IJCSMC, Vol. 2, Issue. 7, July 2013, pg.1 – 6

RESEARCH ARTICLE

ACTIVITY BASED PERSON IDENTIFICATION USING PARTICLE SWARM OPTIMIZATION ALGORITHM

Sruthy Sebastian
Department of Computer Science and Engineering, Calicut University, Kerala, India

sruthyksebastian@gmail.com

Abstract— This paper presents a generic non-invasive person identification method that exploits discriminative power of different activities performed by the same person. A multi-camera setup is used to capture the human body from different viewing angles. Person identification, activity recognition, and viewing angle specification results are obtained for all the available cameras independently. Utilizing a particle swarm optimization (PSO) and linear discriminant analysis (LDA) based algorithm, an unknown movement is first classified, and, then, the person performing the movement is recognized from a movement specific person classifier. Human identification performance of the proposed scheme is found to be quite good when tested on publically available databases.

Key Terms: - Person identification; activity recognition; particle swarm optimization; linear discriminant analysis

I. INTRODUCTION

Automated person identification is highly researched in recent years because of its applications, like protected access to computer systems, buildings, cellular phones, ATMs and video surveillance. Person identification is the process of associating an identity to an individual. Person identification techniques are broadly classified into three, namely knowledge based, token based, and biometric based. A knowledge-based approach depends on something an individual knows to make a personal identification, like a password or a personal identification number (PIN). Token-based approaches are based on something an individual have to make a personal identification like a passport, driver's license, ID card, credit card, or keys. These two approaches have several demerits: tokens may be stolen, lost, forgotten or misplaced. The password or PIN code can be forgotten by an authenticated person or predicted by an attacker. The biometric systems [1] use physiological or behavioural features of an individual for identification and it cannot be stolen or lost. Biometric based identification techniques include face recognition, fingerprint technology, iris recognition, hand geometry, keystroke, speech recognition etc. However these methodologies require a very controlled environment and person cooperation to identify a person properly. For example, the person should stand at a standard distance in front of a camera and look at a specific point, or have physical contact with sensors. These techniques cannot be used in automatic surveillance of people in real time situations. Gait recognition has been widely used for this purpose, as it provides a non-invasive way to recognize persons at a distance. It has gained researchers' attention in the last decade and numerous such methods have been proposed.

One disadvantage of gait recognition is the assumption that the person under investigation walks, which is not always the case. Most methods proposed in the literature would probably fail in the case where the person performs a different activity, for example if he/she bends. Thus, the activity information should be taken into account in order to provide the correct person ID. Despite the fact that gait recognition has been studied a lot, the use of other activities has not been exploited yet for person identification. In fact, “walk” can be seen as a special case of a wide range of human activities, such as “run,” “bend,” “jump,” “eat,” or “drink,” which can be used in order to reveal the identity of a person depicted in a video stream. Similar to gait, the global human body information, in the manner of human body proportions and shape, is conserved while observing a person performing other activities. In addition, dynamics observed in different activities may be very distinctive. That is, although people walk in quite a similar way, they may perform other activities, like eating, quite differently. This means that it is more probable to achieve good identification performance if we exploit several, possibly all, different activities a person performs. Indeed, other activities, besides walking, may contain more discriminant information for person identification, as execution style of one activity may uniquely describe a person.

II. Related Work

This section provides a review on a few methods that exploit activity information for person identification. A. F. Bobik [2] et al., develops a view-based approach to the representation and recognition of action that is designed to support the direct recognition of the motion itself. The basis of the representation is a temporal template which is a static vector-image where the vector value at each point is a function of the motion properties at the corresponding spatial location in an image sequence. This method uses the construction of a binary motion-energy image (MEI) which represents where motion has occurred in an image sequence. Next, generate a motion-history image (MHI) which is a scalar-valued image where intensity is a function of recency of motion. Taken together, the MEI and MHI can be considered as a two component version of a temporal template, a vector-valued image where each component of each pixel is some function of the motion at that pixel location. These view-specific templates are matched against the stored models of views of known actions. But this method assumes the same viewing angle during training and recognition phases.

A generic gait recognition approach is proposed in [3]. Here the assumptions are gait capturing is performed in environments where the background is as uniform as possible and the walking subject should be walking in a direction perpendicular to the optical axis of the capturing device since the side view of walking individuals discloses the most information about their gait. This method cannot be used for automatic person identification because it requires uniform background and specific viewing angles.

In gait recognition based on statistical shape analysis L. Wang [4] et al., proposes a simple and efficient automatic gait recognition algorithm using statistical shape analysis. For each image sequence, an improved background subtraction procedure is used to extract moving silhouettes of the walking figure from the background. Temporal changes of the detected silhouettes are then represented as an associated sequence of complex vector configurations in a common coordinate frame, and are further analysed using the Procrustes shape analysis method to obtain mean shape as gait signature. Supervised pattern classification techniques based on the full Procrustes distance measure are adopted for recognition. The main drawback of the current method is that it is view-dependent, which is analogous to the state of the art of past algorithms

Z. X. Chi [5] et al., proposes a gait recognition using radon transform and linear discriminant analysis. For each gait sequence, the transformed silhouettes are used for the computation of a template. The set of all templates is subsequently subjected to linear discriminant analysis and subspace projection. In this manner, each gait sequence is described using a low-dimensional feature vector consisting of selected Radon template coefficients. Given a test feature vector, gait recognition and verification is achieved by appropriately comparing it to feature vectors in a reference gait database. This method uses only one camera and its performance is closely related to the viewing angle. A. Kale [6] et al., proposes view invariant gait recognition algorithm. This method uses a perspective projection model to synthesize a side view (referred to as canonical view) from any other arbitrary view using a single camera.

S. Yu [7] et al., proposes a method for modelling the Effect of View Angle Variation on Appearance-Based Gait Recognition. View angle variation is a significant factor among those that affect gait recognition performance. This proposes two models, a geometrical one and a mathematical one, to model the effect of view angle variation on appearance-based gait recognition. These models will be valuable for designing robust gait recognition systems. D. Tan [8] et al., proposes a framework in an attempt to provide standard evaluation method to compare the performance of different gait recognition algorithm. The framework consists of a large gait database, a large set of well-designed experiments and some evaluation metrics. There are 124 subjects in the database, and the gait data was captured from 11 views. Three variations, namely view angle, clothing and carrying condition changes, are separately considered in the database. The framework provides a platform to evaluate gait recognition algorithms. It can promote the development of gait recognition.

A method that extends gait recognition to include running activity is presented in [9]. Temporal template matching is applied to extract the angles of lower leg rotation during the entire gait cycle. The magnitude of the Fourier transform of these rotation signals over time provides a gait representation. Classification is achieved using a k-nearest neighbor framework. While this method can incorporate running in person identification, the use of other activities in such framework is not straight forward. This is because most activities, such as “wave one hand,” “eat,” or “drink,” cannot be suitably described in this approach. Furthermore, it assumes recognition of the lower part of the human body. The gait representation is based on edge detection followed by template matching and is prone to false positives.

III. PROPOSED SYSTEM

This paper, presents a novel method for activity-based person identification. Walk is assumed to be one, but not the only, activity, which will be used for person identification. An uncalibrated multi-camera setup is used to capture the human body from different viewing angles. The person is represented by his/her body poses during activity execution. This is a low computational cost representation because it does not involve a 3-D reconstruction of the human body. The system uses static body information [Dyemes] provided by human body poses forming an activity. By calculating the similarity of each test human body pose with all dynemes, the temporal information is preserved. The classification procedure involves the projection of the final activity representation in a low-dimensional discriminant subspace using LDA. For each of the cameras, person identification, activity recognition, and viewing angle specification is performed independently. By properly combining the results produced for all the available cameras, a view-invariant and activity-independent person identification method is obtained

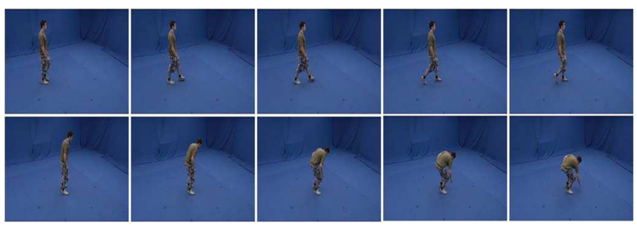


Fig. 1: Video frames depicting activity instances of a walk sequence (top) and a bend sequence (bottom) [10]

A. Preprocessing Phase

In this phase firstly divide the whole video representing the activities performed by the user into frames. Compute the difference between the 2 consecutive image frames and apply fixed threshold to the difference image. Set zeros to those pixels whose values are below the threshold. This outputs the binary representation for the entire video frames. Then divide the image area into rectangular blocks. If the difference of image frames is one within the block, it contains motion. This procedure depicts a distinct posture of the movement for each frame. From these body posture masks, the body posture regions of interest (ROIs) are extracted, centered in respect to the centroid of the body postures along the whole movement sequence, and scaled to the same dimension using bicubic interpolation. A ROI is scanned column wise to produce the so called posture vector X $\in \Re^F$, where F is the number of pixels in the ROI.

Fig. 2. Posture frames of human poses during the execution of seven activities depicted from various viewing angles. From left to right, “walk,” “run,” “jump in place,” “jump forward,” “wave right hand,” “eat,” and “drink.”

B. Video sequence classification

Let W be an annotated database of movement videos belonging to one of $q = 1, \ldots, Q$ different classes, where q declares a specific person or a specific movement, depending on the recognition task. The i^{th} video sequence of the q^{th} class with length L_i is represented as a set of posture vectors $\{X^{(q)}_{i,1,\ldots,\ldots}, X^{(q)}_{i}, L_i\}$, and, similarly, the whole database with the set $\{X^{(1)}_{i,1,\ldots,\ldots}, X^{(1)}_{i}, L_{i,\ldots} X^{(Q)}_{Oq,1,\ldots} \ldots, \ldots X^{(Q)}_{Oq,L}\}$

where Oq is the number of sequences in the q-th class. Each class is modelled as a mixture density, where the mixture components are represented by their centers, called dyneme vectors, {v1, . . . , vC}. Considering unlabelled data and assuming that the number of dynemes C and the fuzzification parameter m are known, the

K-Means clustering[11] algorithm is used to compute the dynemes and fuzzy logic applied to compute the quantized posture vectors . After relabelling the posture vectors, the arithmetic mean of the quantized postures is taken to represent the i^{th} video sequence of the q^{th} class. Therefore, each video in the database is represented by a single vector. The labeling information can be further exploited to reduce the dimensionality of the feature vectors using linear discriminant analysis [LDA]. The q^{th} class (movement type or human id) can then be represented by the mean of all feature vectors belonging to this class. Hence, in order to classify a test movement video we first retrieve the feature vector of the video. Particle swarm optimization algorithm is used for feature extraction and classification. The test video is assigned to the class represented by the prototype that gave the maximum probability value in Bayesian classifier.

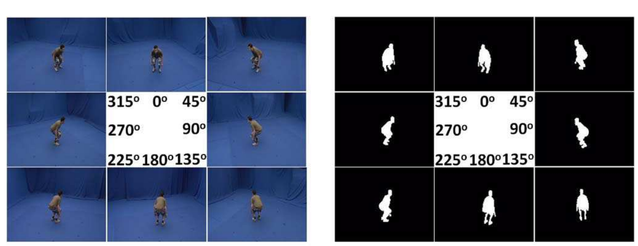
C. Person Identification from real time activities

Let U be an annotated movement video database that contains P persons performing R different movements, i.e., each movement video has two labels, r ∈ 2 [1,R] and p ∈ 2 [1, P] regarding the movement type and the person it belongs respectively. Using all the movement videos of the database and utilizing only the movement type labelling information r, the procedure described in section III-B is used to train a movement type classifier. Then, we break the movement video database to R distinct subsets Ur, r = 1, . . . ,R, i.e., Ur subset contains only movement videos of the rth movement type. Each subset, Ur, is then used to train a movement specific, person classifier. The training of each classifier within each subset is done using the algorithm of section III-B where now only the person specific labeling information p is exploited.

At the testing phase, assuming that a test video depicts the same person performing sequentially the R different movements, the test video is segmented to produce R different movement videos. Then, each movement video is classified from the movement classifier, and it is directed to the respective movement specific, person classifier. Therefore, for each classifier a different feature vector is computed. The movement specific classifiers have been trained using different training sets, and, thus, we may assume that the feature vectors of the test movement videos, are conditionally independent. In this case, the sum rule proposed in [12] can be applied to combine the results of the individual classifiers.

IV. EXPERIMENTS

This section presents the experiments conducted in order to evaluate the proposed method. A multi-view activity recognition database [13] is used to demonstrate the effectiveness of the proposed method in dealing with all open issues discussed in the literature. The database has been created using a convergent eight camera setup to produce high definition multi-view videos, where each video depicts one of eight persons performing one of twelve different human motions. Various types of motions have been recorded, i.e., scenes where one person performs a specific movement, scenes where a person executes different movements in a succession and scenes where two persons interact with each other. Moreover, the subjects have different body sizes, clothing and are of different sex, nationalities, etc.. The multi-view videos have been further processed to produce a 3D mesh at each frame describing the respective 3D human body surface. The studio where the database was recorded is equipped with eight Thomson Viper cameras, equally spaced in a ring of 8m diameter at a height of 2m above the studio floor. An even ambient illumination of around 4000 lux is provided by an array of KinoFlo fluorescent tubes on the ceiling, with flicker less operation and a consistent color spectrum. The cameras were positioned above the capture volume and were directed downward to exclude the lighting from the field-of view. The cameras have a wide 45° baseline to provide 360° coverage of a capture volume of 4m × 4m × 2m. A blue screen backdrop was used to facilitate foreground segmentation. Human actions are captured in HD-SDI 20-bit 4:2:2 formats with 1920 × 1080 resolution at 25Hz progressive scan. Synchronized videos from all eight cameras are recorded uncompressed direct to disk with eight dedicated PC capture boxes using DVS HD capture cards



(a) (b)

Fig. 3. (a) Eight video frames depicting a person of the i3DPost multi-view activity recognition database from different viewing angles and (b) the corresponding binary body images

The cross-validation procedure has been applied to the i3DPost eight-view database. Multi-period videos were manually split to activity videos depicting one activity instance. Four instances of each activity class were used. In each fold of the cross-validation procedure, activity videos depicting three instances of each activity class performed by all the persons in the database have been used for training. Activity videos depicting the remaining activity instances were used for evaluation. Fig. 4 illustrates the mean person identification rates obtained for different numbers of dynemes and classifiers combination strategies. It can be seen that the Sum rule outperforms both the Product rule and ML estimation. Furthermore, it can be seen that, by increasing the number of dynemes, the identification rate increases. This experiment illustrates the ability of the proposed method to correctly identify a person performing an arbitrary activity is 94.37%

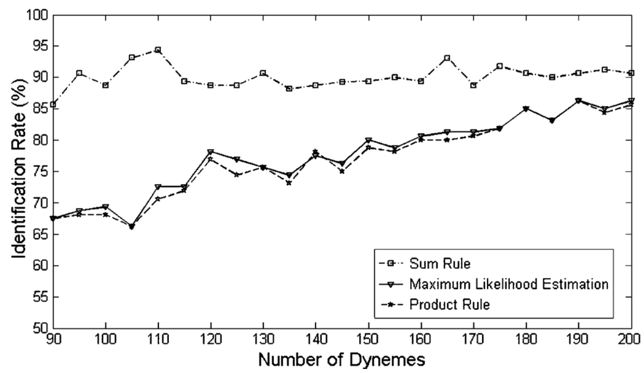


Fig. 4. Mean person identification rates applying the cross-validation procedure to the i3DPost 8-view activity recognition database versus the number dynemes

V. CONCLUSIONS

This paper presents a unified framework aiming at activity-based person identification. A generic view-invariant person identification method has been proposed, by exploiting the information provided by a multi-camera setup and incorporating several activity classes in the person identification procedure. The adopted activity representation scheme exploits the global human body information, in the form of binary body masks. Primitive human body poses, “dynemes” are determined in the training phase. Test activity videos, consists of human body poses, and are represented by their similarity with the dynemes. Particle swarm optimization algorithm is used for feature selection; fuzzy logic is used for classifying these features to different activity classes. By properly combining classification results coming from all the available cameras in the person identification (test) phase using a Bayesian approach, the proposed method achieves high identification rates.

REFERENCES

[1] A. K. Jain, A. Ross, S. Prabhakar, "An Introduction to Biometric Recognition", IEEE Trans. on Circuits and Systems for Video Technology, Vol. 14, No. 1, pp 4-19, January 2004
[2] A. Bobick and J. Davis, “The recognition of human movement using temporal templates,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 23, no. 3, pp. 257–267, Mar. 2001.
[3] N. Boulgouris, D. Hatzinakos, and K. Plataniotis, “Gait recognition: A challenging signal processing technology for biometric identification,” IEEE Signal Process. Mag., vol. 22, no. 6, pp. 78–90, Nov. 2005
[4] L. Wang, T. Tan, W. Hu, and H. Ning, “Automatic gait recognition based on statistical shape analysis,” IEEE Trans. Image Process., vol. 12, no. 9, pp. 1120–1131, Sep. 2003.
[5] N. Boulgouris and Z. Chi, “Gait recognition using radon transform and linear discriminant analysis,” IEEE Trans. Image Process., vol. 16, no. 3, pp. 731–740, Mar. 2007.
[6] A. Kale, A. Chowdhury, and R. Chellappa, “Towards a view invariant gait recognition algorithm,” in Proc. IEEE Conf. Advanced Video and Signal Based Surveillance, 2003, pp. 143–150.
[7] S. Yu, D. Tan, and T. Tan, “Modeling the effect of view angle variation on appearance-based gait recognition,” in Proc. Computer vision (ACCV 2006), 2006, pp. 807–816
[8] S.Yu, D.Tan, and T. Tan, “A framework for evaluating the effect of viewangle, clothing and carrying condition on gait recognition,” in Proc. IEEE 18th Int. Conf. Pattern Recognition, 2006, vol. 4, pp. 441–444.
[9] C. Yam, M. Nixon, and J. Carter, “Extended model-based automatic gait recognition of walking and

running," in Audio-and Video-Based Biometric person authentication. Newyork: Springer, 2001, pp. 278-283

[10] N. Gkalelis, H. Kim, A. Hilton, N. Nikolaidis, and I. Pitas, "The i3DPost multi-view and 3d human action/interaction database," in Proc. 6th Conf. Visual Media Production, Nov. 2009, pp. 159–168

[11] A. Webb, Statistical Pattern Recognition. London, U.K.: Hodder (1999)

[12] J. Kittler, M. Hatef, R. P. W. Duin, J. Matas: On combining classifiers IEEE Trans. Pattern Anal. Mach. Intell., vol. 20, no. 3, PP. 226-239 (1998)

[13] N. Gkalelis, H. Kim, A. Hilton, N. Nikolaidis, and I. Pitas, "The i3DPost multi-view and 3d human action/interaction database," in Proc. 6th Conf. Visual Media Production, Nov. 2009, pp. 159–168

