論文 / 著書情報
Article / Book Information

Title	Statistical Person Verification Using Behavioral Patterns from Complex Human Motion
Author	Felipe Gomez-Caballero, Takahiro Shinozaki, Sadaoki Furui, Koichi Shinoda
Journal/Book name	New Trends in Image Analysis and Processing ICIAP 2013, , , pp. 550-558
発行日 / Issue date	2013, 9
DOI	http://dx.doi.org/10.1007/978-3-642-41190-8_60
権利情報 / Copyright	The original publication is available at www.springerlink.com.
Note	このファイルは著者（最終）版です。 This file is author (final) version.

Statistical Person Verification using Behavioral Patterns from Complex Human Motion

Felipe Gomez-Caballero, Takahiro Shinozaki, Sadaoki Furui, and
Koichi Shinoda

Tokyo Institute of Technology, Tokyo, Japan,
`felipe@ks.cs.titech.ac.jp`

Abstract. We propose a person verification method based on behavioral patterns from complex human movements. Behavioral patterns are represented by anthropometric and kinematic features of human body motion acquired by a Kinect RGBD sensor. We focus on complex movements to demonstrate that independent and rhythmic movement of body parts carries a significant amount of behavioral information. We take a statistical approach by Gaussian mixture models to model the individual behavioral patterns. We demonstrate that subject-preferred movements are more robust against forgery attacks and variations over time than predetermined subject-independent movements. The obtained equal error rate was 15.7% when using subject-preferred movements and 27.3% when using a predefined sequence of movements.

Keywords: person verification, individuality, human movement, GMM

1 Introduction

Automatic identity verification systems provide a secure means for access control to facilities or information. Traditionally, they have required the use of keys/cards or passwords. However, these identity tokens are easily lost or stolen. This can be solved by a biometrics approach, which identifies individuals based on their physiological or behavioral traits [9]. Physiological biometrics are stable since they rely on unique and permanent physical traits, such as fingerprint or iris [14]. However, they cannot be changed if the biometric data is counterfeited. In behavioral biometrics, identity is verified through action patterns which can be repeated in a unique manner, such as voice [5] and gait [1]. Behavioral biometrics are less stable since behavior may change due to the physical state of the individual. However, they are difficult to disguise or to imitate by others.

We focus on the individuality of *human motion* as an alternative cue when other behavioral biometrics can not be obtained or when their quality is low. Some previous studies for this application have used simple movements (e.g. arm raising) as the behavioral cue [17, 16, 7, 8]. In this approach, it is easy for users to remember the movement and to repeat it in a stable manner. It is also easy for the identification systems to segment a person from the scene and track the movement. On the other hand, a psychology study has shown that such simple behavioral motions tend to be similar among different users [3], and thus, it may be difficult to be used for real authentication applications.

Therefore, the use of more complex movements is promising to increase the accuracy of person verification using behavioral patterns. However, there have been two problems in this approach. First, it may be difficult for users to do the same gesture again and again, when it is complex. If the gestures from the same user vary, they cannot be used for authentication. Second, it becomes difficult for the system to segment and track such complex movement precisely. From these two reasons, such an approach has not yet been applied until now. As for the first problem, intra-subject motion variety, Chow et al. [2] reported in a motor control study that a person can repeat the same movement precisely even if it is complex, when he/she is familiar with the movement. For example, from a biomechanics perspective, volleyball players can perform spike jump movements in the same way repeatedly [20]. The use of such subject-dependent familiar motions may solve the first problem. Recently, 3D cameras such as Kinect sensors have been often used to capture human motions [15, 4, 10]. Without using any markers, they can segment and track complex human motions. The use of 3D cameras may solve our second problem.

This paper proposes person verification based on complex behavioral motions using Kinect sensors. We take a statistical approach using Gaussian mixtures models (GMM) to robustly model individual behavioral patterns. We evaluate our method using a dataset containing a variety of complex movements.

The remainder of the paper is organized as follows. Section 2 gives a brief overview of the proposed method. Section 3 describes the features and extraction process. Section 4 describes the statistical approach used to model the person's behavioral patterns. Section 5 describes the task and classifiers used in our system. In Section 6, experimental conditions are explained and results are presented. Finally, conclusion and future work are described in Section 7.

2 Proposed method

Figure 1 shows an overview of the system implemented in our method, consisting of three major phases: feature extraction, model training and verification. In the feature extraction phase, the video input from a Kinect sensor is processed to extract features from skeletal joint points. In the model training phase, statistical models are created for each person using the extracted features. In the verification phase a sample video input is matched against the claimed person's model to produce a score. If the score is above a threshold the identity claim is accepted, otherwise rejected.

3 Feature extraction

We implement an image processing front-end using the Kinect SDK to locate and track the skeletal joints shown on Figure 2(a). The input consists of video stream acquired from a Kinect sensor at 30 frames per second. For each frame,

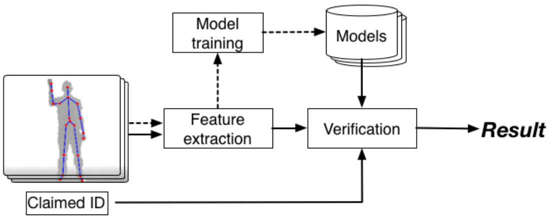

Fig. 1: Person identity verification system. (Dotted lines indicate process flow during training and solid lines indicate the flow during identity verification).

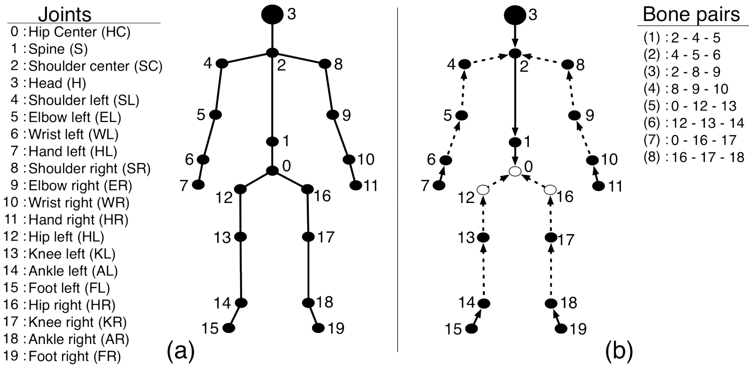

Fig. 2: (a)Skeletal joints found by Kinect SDK. (b)Hierarchical skeleton structure. Arrows indicate hierarchical dependence between joints. Dotted lines represent bones.

3D position of the 20 joints found by the Kinect SDK is extracted. The positions are normalized by following the hierarchical skeleton structure which takes the hip center joint as the origin. For each joint, velocity and acceleration values are calculated since these characteristics have been proved to be useful for recognition of self-generated actions [11]. In addition, the angle between each of eight bone pairs is calculated. A bone pair is defined by the two segments formed between three adjoining joints. Figure 2(b) illustrates the hierarchical skeleton dependence between joints and the bones considered during feature extraction.

The feature vectors created for the experiments in this paper include the 3D position, velocity and acceleration of 17 joints and the angle between eight bone pairs. By this setup, we create a 93 dimension feature vector that captures anthropometric and kinematic characteristics. The 'hip left' and 'hip right' joints are not included since these points exhibited limited motion in the captured samples. The joint 'hip center' is also not included since it is used for normalization.

4 Model training

We take a statistical approach by Gaussian mixture models (GMM) [18] to robustly model the individual behavioral patterns. A GMM is a parametric probability density function represented as a weighted sum of M component Gaussian distributions given by the equation $p(\mathbf{x}|\lambda) = \sum_{i=1}^{M} w_i g(\mathbf{x}|\mu_i \Sigma_i)$, where $\mathbf{x}$ is a D-dimensional feature vector, w_i is the mixture weight, and $g(\mathbf{x}|\mu_i \Sigma_i)$ is the component Gaussian distribution with mean vector μ_i and covariance Σ_i. A GMM represents feature vectors by its mean components, as well as their average variations by the covariance matrix. Therefore it is possible to model the variations of features that characterize individual behavioral patterns.

To robustly estimate the GMM parameters with a limited amount of data, we use the maximum likelihood linear regression (MLLR) [13] method which is often used in speaker recognition. MLLR estimates a set of transformations that can be shared by several model components, hence reducing the required amount of adaptation data [12, 6]. In MLLR, an affine transform $(\mathbf{A}, \mathbf{b})$ is applied to the Gaussian parameters (μ) of an initial model to create the parameters of a new person-dependent model $(\hat{\mu})$ by $\hat{\mu} = \mathbf{A}\mu + \mathbf{b}$, where $\mathbf{A}$ is an $n \times n$ transformation matrix (n is the dimensionality of the data) and $\mathbf{b}$ is a bias vector which maximizes the likelihood of the adaptation data. As an initial model, we use a universal background model (UBM) [19]. A UBM is a GMM trained by EM parameter estimation using the training data from all the subjects in the dataset. The UBM parameters are then adapted via MLLR to derive a person-dependent model by using the person's training data. The vector for each sample used to train the models has a dimensionality of $93 \times N$, where N is the number of frames in a given training sample and 93 corresponds to the feature vector size.

5 Identity verification

In this task, an unknown person claims an identity and provides a sample to be compared with a model for the person whose identity is claimed. We implement a log-likelihood ratio (LLR) scheme [19] for the decision-making process. The LLR measures how much better the claimant's model scores for a test sample compared with a non-claimant model. As shown in equation 1, the LLR is obtained by the difference in scores resulting from testing a given sample (x) against the claimed person model (λ_{pm}) and the UBM (λ_{UBM}).

$$LLR = \log p(x|\lambda_{\text{pm}}) - \log p(x|\lambda_{\text{UBM}}) \tag{1}$$

If the LLR is above a threshold the identity claim is accepted, otherwise rejected.

6 Experiments

6.1 Conditions

For evaluating the proposed method, we collected a dataset consisting of short videos depicting complex human movements recorded by a Kinect sensor over several sessions. A total of 16 subjects (4 females, 12 males) were recorded performing two different types of movements. We collected a 'predefined sequence

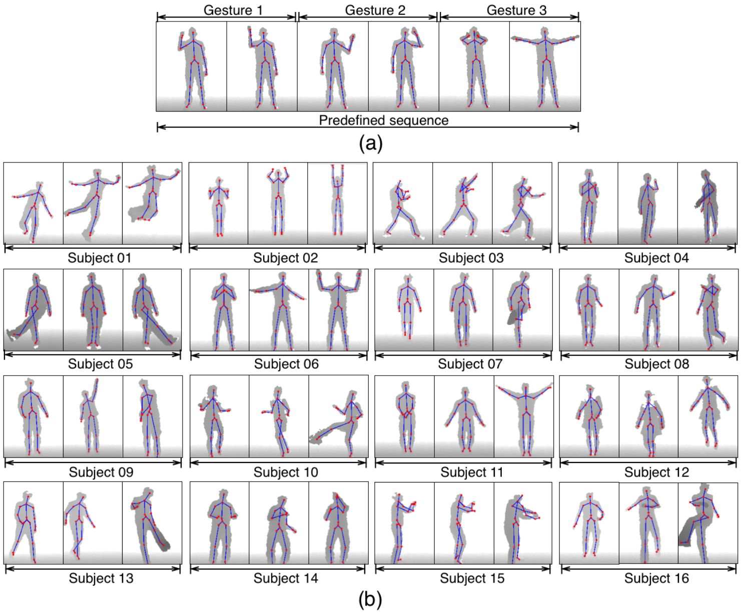

Fig. 3: (a) Example of the predefined sequence of movements. *Gesture 1*: Knock door, *Gesture 2*: Hand wave, *Gesture 3*: Open double sliding door. (b) Examples of subject-preferred movements, including: aerobics (subject 05, 06, 11), jumping (01, 03, 12), soccer kick (07, 13, 16), martial art (03), dance step (10), tennis swing (08), table tennis swing (14), badminton swing (09), batting (15) and pitch (04).

of movements' (FIXED) and a 'subject-preferred movement' (PREFERRED), in order to prove that movements familiar to the subject are more robust than other movements.

The FIXED movement consists of three consecutive gestures, knock door, hand wave and open double sliding door. Figure 3(a) shows its example. Each subject was instructed on how to perform the FIXED movement on the first recording. Figure 3(b) shows example frames of the PREFERRED movements. Subjects were asked in advance to select a preferred complex movement which they could easily repeat. Since the subjects are likely to select movements that are familiar for them, we expect that behavior patterns will be more stable. All subjects reported that they felt comfortable performing their preferred and familiar movement.

The dataset is organized in one training session and six testing sessions. Recording of the training and the first testing sessions were separated by a 28 days interval between them. The six testing sessions were recorded with an interval of seven days between them. The training session contains 20 samples per subject for each movement. Each testing session contains 10 samples per subject for each movement. Average length per sample for the FIXED and PREFERRED movement are 5.18 and 3.86 seconds respectively. Samples were recorded at 30fps.

For the verification test, 10 subjects were randomly chosen out of 16 as the target. In order to assess the robustness of PREFERRED movements against forgery attacks, we asked the 16 subjects to act as impostors and imitate the movement of the 10 target subjects. After watching a video showing the target subject executing his/her movement, subjects tried to imitate each movement five times. In the case of FIXED movements, we randomly select five samples from non-target subjects as impostor data since all subjects perform the same movement. Verification tests are conducted for each movement category using samples from the six testing sessions and the impostor samples collected in a single session. For each genuine target subject, we conduct 60 verification trials where each trial used a single sample from the subject. For the forgery attacks, we conduct 75 verification trials per target subject where each trial used a single sample from 15 impostors. Performance is measured by the equal error rate (EER) calculated a posteriori. The EER is the value where the false acceptance and false rejection rates are equal, hence an optimal threshold can be found. The subject models were created using 32 Gaussian mixture components.

6.2 Results

Table 1 shows the false rejection (FR) and false acceptance (FA) rates per target subject and EER using global optimal threshold of the systems when forgery attacks by impostors are introduced. It can be seen that performance of the system using PREFERRED movements is higher compared to using FIXED movements. This confirms that robustness against forgery attacks of PREFERRED movements is higher than FIXED movements.

The results also suggest that the choice of PREFERRED movement affects the performance of the system, since some movements are more robust against forgery than others. For example, subject-15 movement (batting) does not provide sufficient behavioral information due to its limited motion and little change in body pose, hence making it easy to be imitated by impostors. On the other hand, the movement selected by subject-08 (tennis swing) involves a characteristic rhythmic motion of both arms and legs that was difficult to mimic by impostors. This observation is consistent with the findings in [3], demonstrating that independent and rhythmic movement of body parts carries a significant amount of behavioral information. Although FIXED movement execution time is in average longer, it can be assumed that the amount of behavioral information encoded in PREFERRED movements is higher and more stable. The Detection Error Trade-off (DET) curves for both systems are shown in Figure 4.

7 Conclusions and future work

We have extended our previously proposed approach [8] for person verification based on behavioral patterns by using complex human motion. We focused on the behavioral patterns of subject-preferred and familiar movements. By using a Kinect sensor, accurate segmentation and tracking of the human body was

Table 1: False rejection (FR) and false acceptance (FA) rates per target subject and EER for systems using predefined sequence of movements (FIXED) and subject-preferred movement (PREFERRED).

	FIXED		**PREFERRED**	
Target	**FR (%)**	**FA (%)**	**FR (%)**	**FA (%)**
01	6.7	75.0	0.0	28.0
02	15.0	54.8	0.0	21.3
03	1.7	18.2	5.0	34.7
04	100	2.1	65.0	13.3
05	0.0	31.1	1.7	6.7
06	0.0	58.8	6.7	6.7
07	96.7	0.0	13.3	1.3
08	28.3	2.3	3.3	0.0
11	16.7	7.9	51.7	0.0
15	48.3	0.0	3.3	45.3
EER (%)	27.3		15.7	

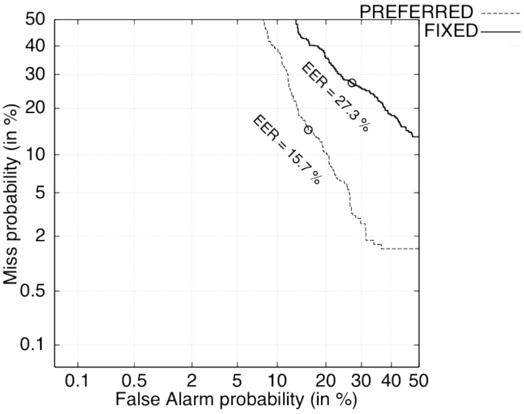

Fig. 4: DET curves for systems using FIXED and PREFERRED movements.

feasible. We have shown that our system achieves higher performance by using subject-preferred movements (EER = 15.7%) compared to using predefined sequence of movements (EER = 27.3%) when forgery attacks by impostors are introduced. Results also suggest that the choice of subject-preferred movement directly affects the performance. We have also confirmed that our system is able to verify the identity of a person even when there is a time difference of 28 to 63 days between training and testing sessions. We consider the results are encouraging and further research is worthwhile to pursue. Moreover, the proposed method can serve as support to other biometric methods.

For the future work, we plan to increase the number of subjects in the dataset and include a wider variety of complex human movements. In order to further improve the performance, we would like to implement a Hidden Markov Model (HMM) based framework, taking advantage of the temporal information of complex movements. We also would like to improve the feature set to include richer information about each individual behavioral patterns.

References

1. Aqmar, M., Shinoda, K., Furui, S.: Robust Gait-Based Person Identification against Walking Speed Variations. IEICE on Information and Systems 95, 668–676 (2012)
2. Chow, J.Y., Davids, K.W., Button, C., Rein, R.: Dynamics of movement patterning in learning a discrete multiarticular action. Motor Control: The Int. Journal for the Multidisciplinary Study of Voluntary Movement 12(3), 219–240 (2008)
3. Daprati, E., Wriessnegger, S., Lacquaniti, F.: Kinematic cues and recognition of self-generated actions. Experimental Brain Research 177(1), 31–44 (2007)
4. Fernandez-Baena, A., et al.: Biomechanical validation of upper-body and lower-body joint movements of kinect motion capture data for rehabilitation treatments. In: Intelligent Networking and Collaborative Systems 2012. pp. 656–661 (2012)
5. Furui, S.: 40 Years of Progress in Automatic Speaker Recognition. In: Advances in Biometrics, vol. 5558, pp. 1050–1059. Springer (2009)
6. Gales, M.J.F., Woodland, P.C.: Mean and Variance Adaptation within the MLLR Framework. Computer Speech Language 10(4), 249–264 (1996)
7. Gomez-Caballero, F., Shinozaki, T., Furui, S.: User Identification Using Time-of-Flight Camera Image Streams. IPSJ Technical Report Vol.2, 615–616 (2010)
8. Gomez-Caballero, F., Shinozaki, T., Furui, S., Shinoda, K.: Person Authentication Using 3D Human Motion. In: Proc of the 2011 joint ACM workshop on Human gesture and behavior understanding. pp. 35–40. J-HGBU '11, ACM (2011)
9. Jain, A., Ross, A., Prabhakar, S.: An Introduction to Biometric Recognition. Circuits and Systems for Video Technology, IEEE Transactions on 14(1), 4–20 (2004)
10. Kato, K., et al.: Exploring analysis of football actions considering correlation of joints in the body. In: Networked Sensing Systems 2012 Int. Conference on (2012)
11. Knoblich, G., Prinz, W.: Recognition of Self-Generated Actions From Kinematic Displays of Drawing. Journal of Experimental Psychology: Human Perception and Performance 27(2) (2001)
12. Leggetter, C.J., Woodland, P.C.: Flexible Speaker Adaptation Using Maximum Likelihhod Linear Regression. Proc. ARPA Spoken Language Technology Workshop 9, 110–115 (1995)
13. Leggetter, C.J., Woodland, P.C.: Maximum Likelihood Linear Regression for Speaker Adaptation of Continuous Density hidden Markov models. Computer Speech & Language 9(2), 171–185 (1995)
14. Lin, J.L., Hsu, H.L., Jong, T.L., Hsu, W.H.: Biometric Authentication. In: Pattern Recognition, Machine Intelligence and Biometrics, pp. 607–631. Springer (2011)
15. Munsell, B.C., Temlyakov, A., Qu, C., Wang, S.: Person identification using full-body motion and anthropometric biometrics from kinect videos. In: Proc of the 12th int. conference on Computer Vision. pp. 91–100. Springer-Verlag (2012)
16. Pratheepan, Y., et al.: Body Language Based Individual Identification in Video Using Gait and Actions. In: Image and Signal Processing. Lecture Notes in Computer Science, vol. 5099, pp. 368–377. Springer (2008)
17. Pratheepan, Y., et al.: Style of Action Based Individual Recognition in Video Sequences. In: Systems, Man and Cybernetics. pp. 1237–1242 (2008)
18. Reynolds, D.: Automatic Speaker Recognition using Gaussian Mixture Speaker Models. The Lincoln Laboratory Journal pp. 173–192 (1995)
19. Reynolds, D.A., Quatieri, T.F., Dunn, R.B.: Speaker Verification Using Adapted Gaussian Mixture Models. Digital Signal Processing 10(1-3), 19 – 41 (2000)
20. Tokuyama, M., et al.: Individuality and reproducibility in high-speed motion of volleyball spike jumps by phase-matching and averaging. Journal of Biomechanics 38(10), 2050 – 2057 (2005)