T2R2 東京科学大学 リサーチリポジトリ
Science Tokyo Research Repository

論文 / 著書情報
Article / Book Information

Title	A statistical approach for person verification using human behavioral patterns
Authors	Felipe Gomez-Caballero, Takahiro Shinozaki, Sadaoki Furui, Koichi Shinoda
Citation(English)	EURASIP Journal on Image and Video Processing 2013, 2013:44, , pp. 1-11
Pub. date	2013, 8
Creative Commons	次ページ参照

Powered by T2R2 (Science Tokyo Research Repository)

License

Creative Commons：表示

EURASIP Journal on Image and Video Processing
a SpringerOpen Journal

RESEARCH Open Access

A statistical approach for person verification using human behavioral patterns

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

Abstract

We propose a person verification method using behavioral patterns of human upper body motion. Behavioral patterns are represented by three-dimensional features obtained from a time-of-flight camera. We take a statistical approach to model the behavioral patterns using Gaussian mixture models (GMM) and support vector machines. We employ the maximum likelihood linear regression adaptation method to estimate GMM parameters with a limited amount of data. Experimental results show that it reduced by 28.6% the relative equal error rates from a system using the maximum likelihood estimation with 25 samples per subject. We also demonstrate that the proposed approach is robust against variations in body motion over time.

Keywords: Person verification; Behavioral biometrics; GMM; SVM; Time-of-flight camera

1 Introduction

Identity verification systems are getting popular in our daily life. They provide a secure means of controlling access to information or equipment. Traditionally, these systems have required something that *one has* or something that *one knows* (e.g., keys and password). However, these representations of identity can be easily lost, manipulated, or stolen. This problem can be solved by a biometrics approach, which identifies an individual based on his/her characteristic traits. Biometrics can be divided into two classes [1]: physiological and behavioral.

Physiological biometrics uses physical traits, such as fingerprint or the iris [2]. This type of biometrics is stable and accurate since it relies on unique and permanent physical traits. However, they cannot be changed or reissued if the biometric data are exposed or counterfeited [3]. They are also perceived as obtrusive [4]. In behavioral biometrics, a person's identity is verified through action patterns which can be repeated in a unique manner [5]. In comparison to physiological biometrics, behavioral biometrics are less stable since behavior may change due to the environment or the physical state of the individual. On the other hand, they are not obtrusive and are difficult to disguise or to be imitated by others [6]. Examples of behavioral biometrics include voice, keystroke dynamics, signature, and gait. Voice can be used for remote identity verification [7]. Keystroke verifies the identity of a user while he/she uses a computer [8]. Signature can be used in online transactions [9]. Gait has been proven to be useful for surveillance applications where data can be collected from a distance [10].

In this paper, we focus on the individuality of *upper body motion* as an alternative cue for identity verification applications where the acquisition of other behavioral biometrics is not feasible due to visual and space constraints. An example application is an automatic gatekeeper, where a person explicitly requests permission to access a secure area by performing a body gesture in front of a camera. Only a few studies have been done for this kind of application. Pratheepan et al. used arm waving motion [11] and simple actions such as sitting down, standing up, and walking away [12] for individual identification in surveillance applications. These studies used holistic features which represent characteristics from a region of interest of a person as a whole [13]. They used these features to create templates that represent each person's behavioral patterns.

The holistic features used in the above mentioned works are sensitive to visual variations such as clothing differences and view and scale changes since they

*Correspondence: felipe@ks.cs.titech.ac.jp
Tokyo Institute of Technology, 2-12-1 Ookayama, Meguro-ku, Tokyo, 152-8552, Japan

Springer

© 2013 Gomez-Caballero et al.; licensee Springer. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

rely on global information about a person's appearance [14]. To reduce the effects of visual variations, human body models describing the kinematic properties (skeleton) or the shape of the body have been used for feature extraction in gait recognition [15]. By doing this, it is possible to extract features from joint positions to alleviate the effects introduced by clothing changes or noise. However, this technique requires localization and tracking of specific human body parts, which are often erroneous.

Accurate extraction of features also depends on the correct segmentation of a person from the background scene, which can be affected by texture and illumination changes [16]. Recently, time-of-flight (ToF) cameras have been used to reduce such errors by simplifying background segmentation [17,18]. Furthermore, ToF cameras have a low dependency on lighting conditions and invariance to color and texture.

In previous approaches [11,12], extracted features were utilized to create a template characterizing a person's motion. However, templates are weak against the natural variations of individual behavioral patterns since they only encode an average representation of observed samples [19]. Statistical models such as Gaussian mixture models (GMM) and hidden Markov models (HMM) have successfully handled variations in individual behavioral patterns [20-23]. For example, Kale et al. [21] showed that HMMs were more robust than templates for gait recognition. Statistical models usually exhibit higher accuracy than templates since they model intraindividual variations well and are also able to handle variations in the sample duration. One drawback is that a relatively larger amount of data is needed to estimate their parameters. The data sparseness problem is significant when the training data are limited, which is often solved by adaptation techniques such as maximum *a posteriori* (MAP) [24], maximum likelihood linear regression (MLLR) [25], and eigenspace-based techniques [26].

In this paper, we propose a statistical method for person verification based on behavioral patterns of human upper body motion, extending our prior work published in [27] and later in [28]. We use depth information acquired from a ToF camera to extract characteristic features from specific parts of the body in the three-dimensional (3D) space. GMMs are used to robustly model individuals' behavioral patterns. To cope with the problem of data sparseness, we use the MLLR adaptation technique. For identity verification tests, we combine GMMs with support vector machines (SVM) [29] and compare its performance with the GMM log-likelihood ratio framework. Lastly, we evaluate our method using a data set containing samples collected over different sessions to demonstrate that it is able to verify the identity of a person even after a period of time.

The remainder of this paper is organized as follows. Section 2 gives a brief overview of the adopted process for person verification. Section 3 describes the features that are used for person verification in our approach. Section 4 describes the statistical modeling and adaptation techniques used to model a person's behavioral patterns. Section 5 describes the classifiers used for person verification systems. In Section 6, experimental conditions are explained and results are presented. Finally, conclusion and future work are described in Section 7.

2 Overview of our person verification system

Figure 1 shows an overview of our system, consisting of three major phases: feature extraction, model training, and verification. In the feature extraction phase, the video input is processed to extract features from a human upper body motion. In the model training phase, statistical models are created for each person using the extracted features. In the verification phase, a sample 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 by the system, otherwise rejected. In our approach, we focus on capturing short image stream samples of two kinds of upper body motion: a vertical motion of either the left or right arm.

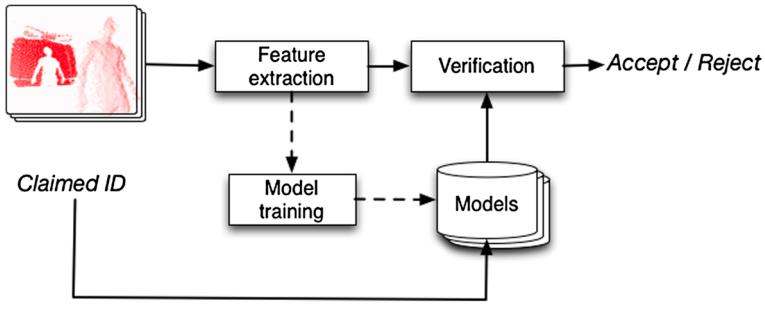

Figure 1 System overview. This figure shows an overview of our person verification system.

3 Feature extraction

We implement an image processing front-end to locate and track eight anatomical landmark points on depth image streams acquired by the ToF camera. The input consists of image streams acquired from a Swissranger SR4000 camera (MESA Imaging, Zurich, Switzerland) [30] at approximately 20 frames per second. Each image frame represents the scene depth map with a resolution of 177 × 144 pixels and a field of view of 43.6°× 34.6°. Each pixel has an accuracy higher than 1 cm within the distance measurement range of 5 m.

Figure 2 shows an overview of the feature extraction algorithm implemented in the image processing front-end. For each frame, it segments the foreground/background to find the person in the scene. For the segmentation, it employs a simplified region growing strategy [31] which aggregates pixels with similar characteristics to a cluster or region. The growing process starts from a seed pixel that initializes a reference for region building, and it decides if a pixel b is considered a part of the current region R according to the following equations:

$$\exists\, a \in R\,,\ \mathrm{Dist}(a,b) < \theta \rightarrow \{b \in R\} \tag{1}$$

$$\mathrm{Dist}(a,b) = \mid D_a - D_b \mid \tag{2}$$

where $\mathrm{Dist}(a,b)$ is the distance between pixel a and pixel b, D_a and D_b are the depth value of pixels a and b, respectively, and pixel a is the seed or a pixel already aggregated to R.

Pixel b is selected from among the four connected neighbor pixels of a, unless it is already a part of the region. The threshold θ in Equation 1 is determined empirically based on preliminary experiments. The region continues growing recursively until no neighbor pixel can satisfy the condition in Equation 1, and then only the region R, depicting the body of a person, remains.

Next, the algorithm segments the person's body into four different regions (i.e., chest, head, left, and right arms). First, it finds a chest rectangle around the center of mass of the body. Then it performs few iterations of expanding and shrinking until 15% of its total area does not contain pixels from the body region. The 15% empty area represents the space between the arms and chest due to the 'open arms' position which each subject takes at the beginning of recording or some possible holes in the body region due to noisy data. After setting the chest rectangle, the algorithm segments the head and both arm regions by region growing with a constrained search space. For the head, region growing starts from the middle point at the top edge of the chest rectangle, and the growing is constrained to the upper sector. For the right/left arm, region growing starts at the corresponding top vertex of the chest rectangle, and growing is constrained to its corresponding side. In case a previous landmark point position (i.e., elbows and head) is available, the algorithm uses it as a seed for the growing. Each region is labeled as detected if its area is larger than a threshold, which is empirically set by preliminary tests.

After body segmentation, the algorithm calculates eight landmark points. The algorithm registers the 3D correspondence of the rectangle's top vertex points as the left and right shoulder points. Also, it registers the 3D correspondence of the rectangle's center as the center of mass of the body. The head point is obtained by taking the 3D correspondence of the calculated center of mass of the head region. For the elbow and hand points, skeletonization operation [32] is applied on each arm region to find the longest single pixel line starting from the shoulder point. We assume that the hand point can be found at the end of the line produced by skeletonization and that the center of mass of this line corresponds to the elbow point. Finally, a Kalman filter [33] for each landmark point is updated to estimate its position in the next frame. If a region is not found, the estimate calculated in the previous frame is used. The Kalman filter is used for visual tracking in order to cope with ambiguities when capturing human movement.

Figure 3 shows the eight landmark points that are used to calculate static and dynamic features from the human upper body. These features are used to create a representation of the person's behavioral patterns. The static

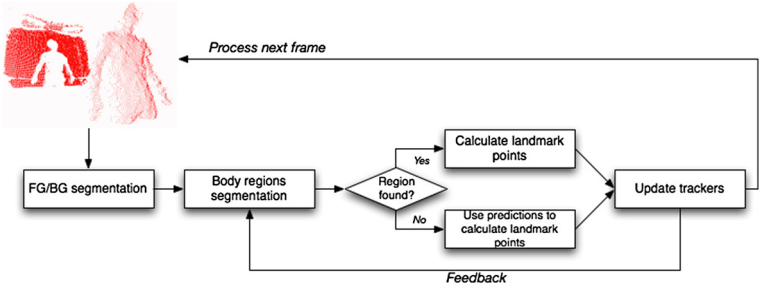

Figure 2 Landmark point location algorithm overview. This figure shows an overview of the feature extraction algorithm implemented in our image processing front-end.

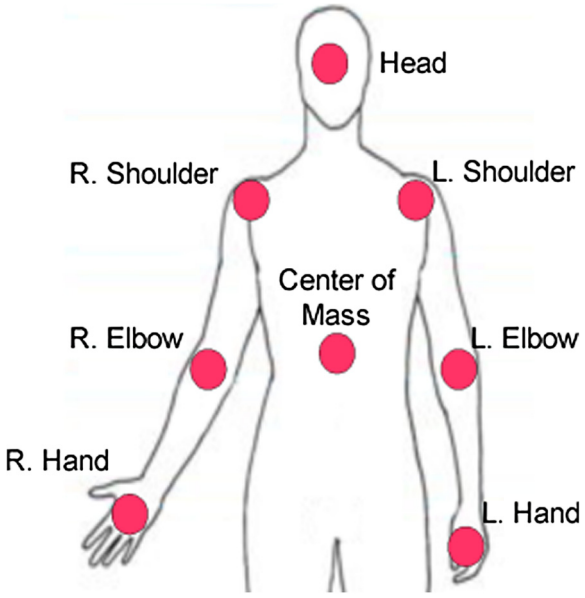

Figure 3 Body landmark points. This figure shows the eight landmark points that are used to calculate static and dynamic features from the human upper body.

features used in our approach are the relative length and orientation of body parts in the three-dimensional space. For the dynamic features, we focus on analyzing the arm motion. The extracted features were selected based on a preliminary analysis where we compared the performance of different feature sets.

For each landmark point, 3D relative position is calculated (8 points $\times$ 3 dimensions). Velocity vectors are calculated for both elbows and hands (4 points $\times$ 3 dimensions) since these points exhibit high level of motion on the image streams, and hence, they carry characteristic behavioral information of the subject. In addition, the 3D direction vectors relative to the chest region are calculated only for the shoulders and head points (3 points $\times$ 3 dimensions) since subjects tend to show a characteristic posture that is visible on these points. The extracted vectors from the posture exhibited high inter-subject differences. By this setup, we create a 45-dimension feature vector that captures the dynamic and static characteristics of each person. Figure 4 shows four frames of a sample processed by our feature extraction algorithm, where localized landmark points are represented as spheres on the body.

4 Statistical modeling based on Gaussian mixture models

In this section, we review the Gaussian mixture model and describe our motivation to use it to model behavioral patterns. A GMM is a parametric probability density function represented as a weighted sum of M component Gaussian distributions [34], as given by the equation

$$p(\mathbf{x}|\lambda) = \sum_{i=1}^{M} w_i g(\mathbf{x}|\mu_i \Sigma_i), \tag{3}$$

where $\mathbf{x}$ is a D-dimensional continuous-valued data vector (i.e., 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. The mixture weights satisfy the constraint that $\sum_{i=1}^{M} w_i = 1$. A GMM is represented by its parameter set:

$$\lambda = \{w_i, \mu_i, \Sigma_i\} \quad i = 1, \ldots, M. \tag{4}$$

The parameters of a GMM are often obtained by maximum likelihood (ML) estimation. Model parameters are iteratively estimated using the expectation maximization (EM) algorithm [35], where the original parameters (λ) are refined to increase the likelihood of the estimated model ($\bar{\lambda}$) for the observed feature vectors X, such that $p(X|\bar{\lambda}) \geq p(X|\lambda)$. Assuming that feature vectors of $X = \{\mathbf{x}_1, \ldots, \mathbf{x}_T\}$ are independent, the log-likelihood of a model (λ) for X is computed as follows:

$$\log p(X|\lambda) = \sum_{t=1}^{T} \log p(x_t|\lambda) \tag{5}$$

A GMM can represent feature vectors by its mean components as well as represent their average variations by the covariance matrix. For this reason, it is possible to model the variations of individual features that characterize a person. While other statistical models such as hidden Markov models and conditional random fields have been proven to be effective to model human motion [36], they are better suited for sequential actions characterizing an activity [37].

To robustly estimate the GMM parameters, we have to deal with the problem of data sparseness. The ML method cannot precisely estimate the model parameters when the training data are sparse and their size is small. Adaptation techniques such as MAP [24], MLLR [25], and eigenspace-based techniques [26] are often used to solve this problem. Although eigenspace-based techniques are effective when adaptation data are extremely small [38], they restrict the model to a lower dimensionality where much information might be lost [39]. On the other hand, MLLR and MAP do not impose this restriction on the models. However, MAP only updates distributions which are observable in the adaptation data, and thus, it requires more data for adaptation [40]. MLLR estimates a set of transformations that can be shared by several model components, hence reducing the amount of adaptation data required [41,42]. Therefore, we use MLLR.

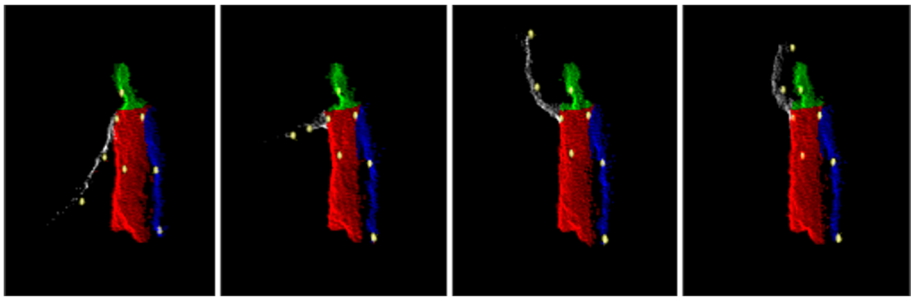

Figure 4 Sample processed by our feature extraction algorithm. This figure shows four frames from a video sample processed by our feature extraction algorithm where localized landmark points are shown as spheres.

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

$$\hat{\mu} = \mathbf{A}\mu + \mathbf{b} \tag{6}$$

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. The parameters $\mathbf{A}$ and $\mathbf{b}$ are shared among all the mixture components of a GMM. To further reduce the number of parameters, we use a diagonal transformation. As an initial model, a universal background model (UBM) [43] is often used. A UBM is a GMM trained by EM parameter estimation using the training data from all the subjects in the data set. The parameters of the UBM are adapted via MLLR to derive a person-dependent model using a specific person's training data.

5 Person 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. A match score between the claimed identity's model and the input sample is computed, and if the score is above a threshold, the identity claim is accepted, otherwise rejected. We implement a GMM log-likelihood ratio and SVM classifiers for the person identity verification system and compare their performance.

5.1 Log-likelihood ratio

We use a hypothesis-testing framework to decide if the input sample x belongs to the claimed person or not. Under this framework, the hypothesis H_0 corresponds to the case where x belongs to the person whose identity is claimed, and the hypothesis H_1 corresponds to the case where x is not from the claimed person. The hypothesis H_0 is represented by a person-dependent model, λ_{pd}, that characterizes the claimed person's identity. For the alternative hypothesis H_1, we use a UBM, λ_{UBM}, to characterize the entire space of possible alternatives to the claimed person's identity. The UBM characterizes the person-independent distribution of features of the subject population expected during recognition. For an input sample x, a claimed person's identity with corresponding model λ_{pd}, and the model of possible non-claimant persons λ_{UBM}, the likelihood ratio is

$$\frac{p(\,x \text{ is from the claimed person})}{p(\,x \text{ is not from claimed person})} = \frac{p(x|\lambda_{\text{pd}})}{p(x|\lambda_{\text{UBM}})}. \tag{7}$$

Then we can obtain the log-likelihood ratio (LLR) [44] by

$$\text{LLR} = \log p(x|\lambda_{\text{pd}}) - \log p(x|\lambda_{\text{UBM}})\,, \tag{8}$$

where $p(x|\lambda_{\text{pd}})$ is the likelihood that the sample belongs to the claimed person, and $p(x|\lambda_{\text{UBM}})$ is the likelihood that the sample does not belong to the claimed person. If the LLR is above a threshold, the identity claim is accepted, otherwise rejected. The LLR scheme is often used for identity verification using probabilistic models since it measures how much better the claimant's model scores for a test sample compared with a non-claimant model (i.e., the UBM). The likelihood normalization provided by the UBM in this scheme helps to minimize person-independent variations that would be common between target and non-target subjects. Figure 5 shows a diagram of this verification scheme.

5.2 Support vector machine

A support vector machine is a binary classifier which focuses on modeling of the boundary between two classes [29,45]. This makes it suitable for identity verification since we can model the boundary between the target person and impostors. It shows good performance especially when the amount of training data is small. A SVM is constructed as a sum of kernel functions $K(\cdot,\cdot)$,

$$f(x) = \sum_{i=1}^{L} \alpha_i t_i K(x, x_i) + d\,, \tag{9}$$

where t_i is an ideal output (either $+1$ or -1, depending on whether the corresponding support vector is a positive

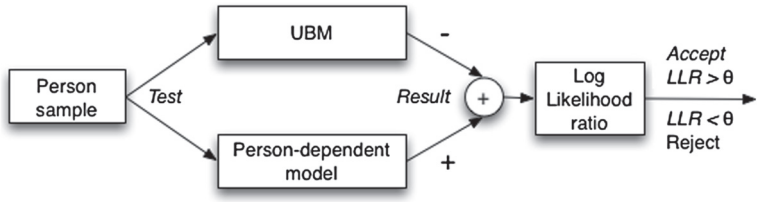

Figure 5 Log-likelihood ratio-based verification. This figure shows the log-likelihood ratio-based verification phase using person-dependent and UBM models.

or negative example of a given class), and α_i and d are the SVM parameters set during the training step. The vector x_i is a support vector obtained from a training set by an optimization method. The data points from the training set lying on the boundaries are the support vectors.

We utilize two methods, SVM-S and SVM-T, where different input features are used. SVM-S employs GMM supervectors as input feature vectors [46,47]. A GMM supervector is formed by concatenating the mean vectors (μ) of GMM mixture components into a single vector. One GMM is created by adaptation using one sample as adaptation data. For training, positive feature vectors are made from the GMMs of the target subject, and negative feature vectors are made from the GMMs of non-target subjects.

SVM-T employs MLLR transformation parameters as input feature vectors [48,49]. The elements of the matrix **A** and the vector **b** in Equation 6 are concatenated to form a single supervector. By doing this, it is possible to model the difference between the subject GMM and the UBM instead of modeling each subject's characteristics. One supervector is obtained by performing MLLR adaptation using one sample as adaptation data. For training, positive feature vectors are made from the transformation parameters of the target subject, and negative feature vectors are made from the transformation parameters of non-target subjects.

6 Experiments

6.1 Conditions

We collected a new data set to evaluate our method since there is no public database containing human upper body movements recorded with a ToF camera over several sessions. The data set used in our approach consists of short image streams of 12 subjects (three females and nine males), where each subject performed two different upper body movements separately. The movements are classified as 'raising left arm' and 'raising right arm'. The data were organized in five sessions recorded with an interval of 3 to 5 days between them. The first session (session 0) contains 25 samples per user for each movement and was used only for the model training phase. Each of the four remaining sessions contains eight samples per person for each movement and was used only as testing samples. The average length per sample is 2.93 s, and the average frame number per sample is 70.5 frames.

Identity verification tests were conducted for each of the four sessions available for testing in the data set. In each session, we conducted eight verification trials per person where each trial used a single sample per subject. System performance was measured by the equal error rate (EER) calculated *a posteriori* for the optimal decision threshold. The EER is the value where the false acceptance rate and false rejection rate are equal. The obtained optimal threshold was used for all subjects. Detection error trade-off (DET) curves were also plotted to assess system behavior in the full range of operating points.

For the LLR system, we used person-dependent models adapted from a UBM by MLLR using a diagonal transformation and compared its performance with models obtained by ML estimation using the EM algorithm. Models used in LLR systems were created with 16 Gaussian mixtures since this setup exhibited the best performance in our preliminary experiments. The Hidden Markov Model Toolkit [50] was used to train GMMs. Each of the SVM-S and SVM-T systems used 2, 4, 8, 16, and 32 Gaussian mixture variants. For the sake of simplicity, only configurations which exhibit the best result are presented in this paper. The SVM-light toolkit [51] was used to train the SVMs. Based on preliminary experiments, we chose linear kernel for SVM training.

The SVMs were trained using the target person's feature vectors as positive examples (25 samples for training) and the non-target persons' feature vectors as negative examples (275 samples for training). To deal with the problem of an unbalanced training data set, we use a cost factor [52] to penalize classification errors on positive examples stronger than errors on negative examples by setting a higher cost for false-positives compared to false-negatives. The feature vector dimensionality for the SVM-S system was $45 \times N$, where N is the number of Gaussian mixtures in the GMM, and 45 corresponds to the number of mean components per Gaussian mixture. For the SVM-T, the MLLR transforms resulted in 2,070-dimensional feature vectors ($45 \times 45 + 45$, including the bias vector **b**).

6.2 Results

First, we show the accuracy results for a landmark point localization test. Then the results for person verification task are presented.

6.2.1 Body landmark point localization accuracy test

We examined the accuracy of the landmark point localization implemented on the image processing front-end. The measure was an average accuracy per landmark point for all subjects. We provided hand-labeled ground truth data and compared them to the landmark positions inferred by our method. If a landmark point was found within D centimeters from the ground truth position, it was considered as correctly localized, otherwise it was considered as incorrectly localized. We set $D = 10$ cm. This value is the same as the one used in the previous human pose recognition research [53]. The hand-labeled data consisted of image streams depicting left arm movements of five subjects (ten samples per person), recorded on a separate session under the same conditions as the rest of the data set.

Table 1 shows the average accuracy results of each landmark point as well as the average distance from the ground truth position for both correct and incorrect localizations. These results show that our method is able to accurately localize the body, head, shoulders, right elbow, and right hand points. However, the localization accuracy decreases for the left side elbow and hand points since the left arm exhibits quick motion.

Despite the low accuracy exhibited for the elbow and hand compared to other landmark points, the tracked points still follow the motion path thanks to the Kalman filter implementation. The measurement errors are smoothed by tuning the parameters of the Kalman filter to achieve a balance between the responsiveness of the tracker and estimate variance. By relying on the Kalman filter, the tracking results are consistent across samples for each subject, which results in a tracker that exhibits low variance - high bias for the elbow and hand points. For this reason, we assume that the motion pattern is preserved to some degree even when the estimated position differs with respect to the ground truth position. Furthermore, by combining features from all landmark points, the effects of inaccurate estimations of elbow and hand points are minimized since the remaining landmark points are estimated and tracked more accurately.

6.2.2 Person verification results

Table 2 shows the average EER over four testing sessions of ML and MLLR estimation using the LLR classifier. We compare identity verification performance when either a full feature set or an arm feature set is used in both training and testing phases. The arm feature set consists of an 18-dimension feature vector including 3D position of the shoulder, elbow, and hand points; velocity vectors of the elbow and hand points; and direction vector of the shoulder point. This feature vector represents only the traits observed on the moving arm.

By using the full feature set, the systems exhibit overall higher performance compared to using only features from the arm in motion. The reason for this result is that the whole upper body takes part in the execution of the analyzed gestures. For example, subjects assume a slightly different posture when raising their arms, and a characteristic motion is observed on the arm that its not raised. The combination of these perceptible features allows the creation of better representations of individual behavioral patterns.

By using the full feature set, the ML estimation yielded an average EER of 9.1% for the left arm samples. The MLLR adaptation reduced the average EER from 9.1% to

Table 1 Landmark localization accuracy

	Average accuracy (%)	Correct localization average distance (cm)	Incorrect localization average distance (cm)
Landmark point			
Head	100	1.8	N/A
Body	100	3.6	N/A
Right shoulder	99.9	1.8	21.8
Left shoulder	98.6	3.4	10.5
Right elbow	98.7	1.5	45.7
Left elbow	52.8	3.5	61.7
Right hand	82.1	3.3	54.6
Left hand	31.8	3.9	63.7
Total average	83.01	2.9	32.2

This table shows the average accuracy results for body landmark localization on left arm movement samples.

Table 2 Average equal error rate for the LLR system

Models	EER(%)	
	Full feature set	Arm feature set
ML (16)	9.1	15.9
MLLR (16)	6.5	17.4

The data presented are the average EER over four sessions for the LLR system with models created by ML and MLLR using left arm movement samples. A full feature set and an arm-only feature set were used to compare verification performance. The number within parentheses is the number of Gaussian mixtures in the GMM used for a given system configuration.

6.5%. The relative reduction in EER by 28% confirmed that MLLR adaptation was effective.

Table 3 shows the average EER over four testing sessions for our proposed LLR, SVM-S, and SVM-T systems. The accuracies in the left arm motion samples were higher than those in the right arm motion samples. Statistical McNemar's test [54] showed that differences between using the left arm and the right arm motion samples are statistically significant (P value < 0.001). The higher accuracy in the left hand movement might be because this movement is more difficult to control intentionally. Thus, the difference of individual behavioral patterns is more perceptible. It is worth noting that most of the subjects included in the data set are right handed.

The two SVM systems achieved the same EER of 8.9% by using different numbers of Gaussian components for the GMM models used to construct the supervectors. However, contrary to our expectations, the LLR system with MLLR adaptation achieved 27% relative reduction in EER compared to the SVM system. McNemar's test confirmed that performance difference between the systems is statistically significant (P value < 0.001).

The reason why the SVM systems did not achieve a better performance might be the size of the data used to derive GMM models from which the supervectors were created. The number of frames per sample used for GMM model adaptation on the SVM-S and SVM-T systems was relatively small (70.5 frames in average), and thus, adaptation was less effective compared to the case of LLR system where person-dependent models were derived using all training samples of the target person.

Table 4 shows the EER per session for the LLR and SVM systems, and Figure 6 shows DET curves for each of these systems. It can be seen that the EER and the performance of each system vary over the four sessions due to the natural change of behavior patterns of human body motion. However, EER does not increase considerably even though there is a time difference of 3 to 20 days between the training and testing sessions. We consider that the proposed method is robust against variations of behavioral patterns over time. Therefore, our approach is promising for general verification tasks.

For comparison purposes, we measured the training and testing time for each system. We used a PC with 8 GB RAM and a double-core Intel(R) Xeon(R) CPU running at 1.86 GHz. For the LLR system, the average training time for the UBM is 34 and 0.33 s for each person-dependent model by MLLR adaptation. For the SVM systems, the average training time is 21.91 s. The average testing time per sample for the LLR and SVM systems is 0.04 and 0.17 s, respectively.

7 Conclusions

We have proposed a statistical approach for person identity verification using behavioral patterns observed on human body motion. In particular, we used behavioral patterns from left arm and right arm movements. By using a ToF camera, we simplified the segmentation of the human body to track specific human body parts in the 3D space. Since we extract static and dynamic features of human motion directly from identified landmark points, the effects of appearance changes were reduced. By taking a statistical approach, we effectively modeled the natural variation in features observed on behavioral patterns. To deal with the problem of data sparseness, we used the MLLR adaptation method along with a UBM to estimate parameters for person-dependent GMMs. In addition, we used GMM components and MLLR transformation parameters as features to create supervectors in the context of SVMs.

We have shown that by using a model adaptation method in the training phase, the average EER of the LLR system was reduced to 6.5%, a relative reduction of 27% compared with our SVM systems. The reason why the SVM systems did not exhibit better performance might be because the model adaptation used to derive GMM models for creating supervectors was not as effective as

Table 3 Average EER over four sessions for verification task using left and right arm movements

System	Left arm EER (%)	Right arm EER (%)
LLR-MLLR (16)	6.5	17.7
SVM-S (8)	8.9	12.2
SVM-T (4)	8.9	35.7

The number within parentheses is the number of Gaussian mixtures in the GMM used for a given system configuration.

Table 4 EER per session for verification task using left arm movement (%)

System	Session 1	Session 2	Session 3	Session 4	Average
LLR-MLLR (16)	4.2	9.4	5.2	7.3	6.5
SVM-S (8)	7.3	14.6	6.3	7.3	8.9
SVM-T (4)	7.3	7.3	9.4	11.5	8.9

The number inside the parenthesis is the number of Gaussian mixtures in the GMM used for a given system configuration.

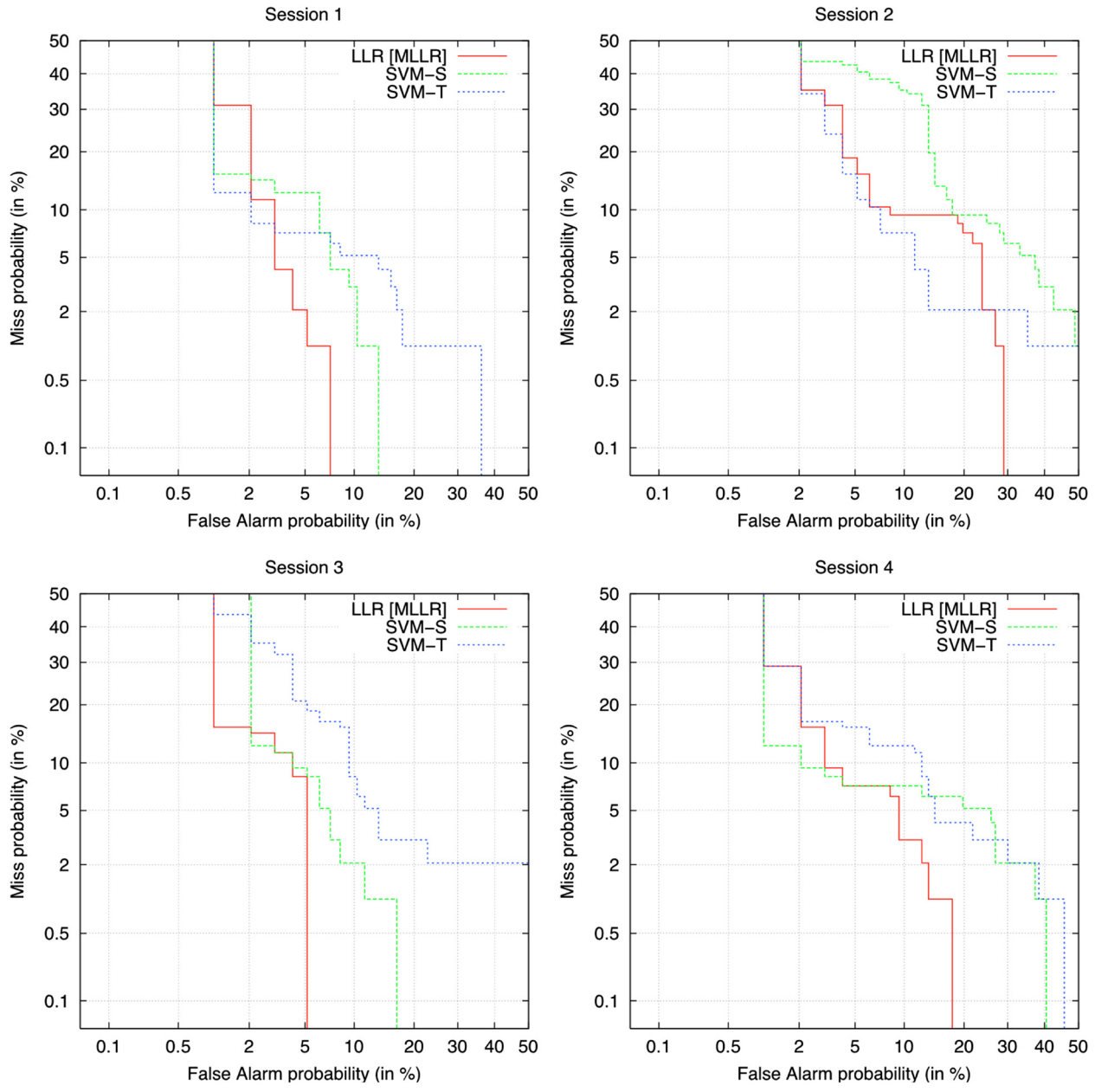

Figure 6 Detection error trade-off curves. This figure shows the detection error trade-off curves at each of the four sessions for the LLR, SVM-S, and SVM-T systems.

the LLR system. While the verification performance did not improve by using SVM classifiers, we consider that providing a comparison against the LLR system is useful for future improvement of such an approach. We found that features extracted from the left arm motion samples exhibit an overall higher degree of distinctiveness compared to those from the right arm motion samples. Furthermore, experimental results showed that our system is able to verify the identity of a person even after a period of time. Hence, our approach is promising for person verification tasks even when natural variations in behavioral patterns exist. Although we used a vertical arm motion in our experiments, the approach presented in this paper is suitable for any other upper body movements or gestures as well.

For future work, we plan to increase the size of the data set, in both the number of users and sessions, to perform further analysis using a wider range of body movements. We would also like to measure the discriminative degree of different upper body movements, especially for

cases when each subject performs a personal movement. We also plan on implementing a more robust landmark point localization and tracking method in order to minimize errors introduced by ambiguities and noisy data. We are interested in implementing a HMM-based framework where more complex movements are used as a cue for identity verification, taking advantage of the temporal information of such movements. We would also like to explore the use of the Kinect image sensor since it can acquire depth images with higher resolutions.

Competing interests
The authors declare that they have no competing interests.

Received: 25 September 2012 Accepted: 2 August 2013
Published: 8 August 2013

References

1. A Jain, A Ross, S Prabhakar, An introduction to biometric recognition. Circuits Syst. Video Technol., IEEE Trans. **14**, 4–20 (2004)
2. JL Lin, HL Hsu, TL Jong, WH Hsu, in *Pattern Recognition, Machine Intelligence and Biometrics*. Biometric authentication (Springer Berlin, 2011), pp. 607–631
3. L O'Gorman, Comparing passwords, tokens and biometrics for user authentication. Proc. IEEE. **91**(12), 2021–2040 (2003)
4. E Vildjiounaite, SM Makela, M Lindholm, R Riihimaki, V Kyllonen, J Mantyjarvi, H Ailisto, in *Pervasive Computing*. Lecture Notes in Computer Science, vol 3968. Unobtrusive multimodal biometrics for ensuring privacy and information security with personal devices (Springer Berlin, 2006), pp. 187–201
5. RV Yampolskiy, V Govindaraju, Behavioural biometrics: a survey and classification. Int. J. Biometrics. **1**, 81–113 (2008)
6. R Moskovitch, C Feher, A Messerman, N Kirschnick, T Mustafic, A Camtepe, B Lohlein, U Heister, S Moller, L Rokach, Y Elovici, in *IEEE International Conference on Intelligence and Security Informatics, 2009*. Identity theft, computers and behavioral biometrics (IEEE Piscataway, 2009), pp. 155–160
7. S Furui, in *Advances in Biometrics*. Lecture Notes in Computer Science, vol 5558. 40 years of progress in automatic speaker recognition (Springer Berlin, 2009), pp. 1050–1059
8. K Revett, ST De Magalhães, HMD Santos, in *Progress in Artificial Intelligence*. 13th Portuguese Conference on Artificial Intelligence, EPIA'07. On the use of rough sets for user authentication via keystroke dynamics (Springer Berlin, 2007), pp. 145–159
9. G Bailador, C Sanchez-Avila, J Guerra-Casanova, A de Santos Sierra, Analysis of pattern recognition techniques for in-air signature biometrics. Pattern Recognit. **44**, 2468–2478 (2011)
10. S Sarkar, P Phillips, Z Liu, I Vega, P Grother, K Bowyer, The humanID gait challenge problem: data sets, performance, and analysis. Pattern Anal. Mach. Intell., IEEE Trans. **27**(2), 162–177 (2005)
11. Y Pratheepan, G Prasad, J Condell, in *IEEE International Conference on Systems, Man and Cybernetics, 2008. SMC 2008*. Style of action based individual recognition in video sequences (IEEE Piscataway, 2008), pp. 1237–1242
12. Y Pratheepan, P Torr, J Condell, G Prasad, in *Image and Signal Processing*. Lecture Notes in Computer Science, vol 5099. Body language based individual identification in video using gait and actions (Springer Berlin, 2008), pp. 368–377
13. J Davis, in *Proceedings of the IEEE Workshop on Detection and Recognition of Events in Video, 2001*. Hierarchical motion history images for recognizing human motion (IEEE Piscataway, 2001), pp. 39–46
14. N Li, Y Xu, XK Yang, in *2010 International Conference on Machine Learning and Cybernetics (ICMLC)*. Part-based human gait identification under clothing and carrying condition variations (Piscataway, 2010), pp. 268–273
15. DK Wagg, MS Nixon, in *Sixth IEEE International Conference on Automatic on Face Gesture Recognition, 2004*. On automated model-based extraction and analysis of gait (IEEE Piscataway, 2004), pp. 11–16
16. L Lee, W Grimson, in *Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition, 2002*. Gait analysis for recognition and classification (IEEE Piscataway, 2002), pp. 148–155
17. R Jensen, R Paulsen, R Larsen, in *Dynamic 3D Imaging*. Lecture Notes in Computer Science, vol 5742. Analysis of gait using a treadmill and a time-of-flight camera (Springer Berlin, 2009), pp. 154–166
18. MO Derawi, H Ali, FA Cheikh, in *BIOSIG. LNI*, vol. 191. Gait recognition using time-of-flight sensor (GI Bonn, 2011), pp. 187–194
19. N Boulgouris, D Hatzinakos, K Plataniotis, Gait recognition: a challenging signal processing technology for biometric identification. Signal Process. Mag., IEEE. **22**(6), 78–90 (2005)
20. A Sundaresan, A RoyChowdhury, R Chellappa, in *Proceedings of the 2003 International Conference on Image Processing, ICIP 2003*, vol. 2. A hidden Markov model based framework for recognition of humans from gait sequences (IEEE Piscataway, 2003), p. II–93–6 vol. 3
21. A Kale, A Sundaresan, AN Rajagopalan, NP Cuntoor, AK Roy-chowdhury, V Krüger, Identification of humans using gait. IEEE Trans. Image Process. **13**, 1163–1173 (2004)
22. MR Aqmar, K Shinoda, S Furui, Robust gait-based person identification against walking speed variations. IEICE Trans. Inf. Syst. **95**, 668–676 (2012)
23. D Reynolds, R Rose, Robust text-independent speaker identification using gaussian mixture speaker models. Speech Audio Process., IEEE Trans. **3**, 72–83 (1995)
24. J Luc, C Gauvain, H Lee, Maximum a posteriori estimation for multivariate gaussian mixture observations of Markov chains. IEEE Trans. Speech Audio Process. **2**, 291–298 (1994)
25. CJ Leggetter, PC Woodland, Maximum likelihood linear regression for speaker adaptation of continuous density hidden Markov models. Comput. Speech Lang. **9**(2), 171–185 (1995)
26. R Kuhn, P Nguyen, JC Junqua, L Goldwasser, N Niedzielski, S Fincke, K Field, M Contolini, in *International Conference on Spoken Language Processing*. Eigenvoices for speaker adaptation (ASSTA Camberra, 1998), pp. 1771–1774
27. F Gomez-Caballero, T Shinozaki, S Furui, User identification using time-of-flight camera image streams. Inf. Process. Soc. Japan Tech Rep. **2**, 615–616 (2010)
28. F Gomez-Caballero, T Shinozaki, S Furui, K Shinoda, in *Proceedings of the 2011 Joint ACM Workshop on Human Gesture and Behavior Understanding, J-HGBU '11*. Person authentication using 3D human motion (ACM New York, 2011), pp. 35–40
29. N Vapnik, *Statistical Learning Theory*. (Wiley, New York, 1998)
30. T Oggier, M Lehmann, R Kaufmann, M Schweizer, M Richter, P Metzler, G Lang, F Lustenberger, N Blanc, An all-solid-state optical range camera for 3D real-time imaging with sub-centimeter depth resolution (SwissRanger). Proc. SPIE Proceedings. **5249**, 534–545 (2004)
31. L Bianchi, R Gatti, L Lombardi, P Lombardi, in *Advances in Image and Video Technology*. 3rd Pacific Rim Symposium on Advances in Image and Video Technology. Lecture Notes in Computer Science, vol 5414. Tracking without background model for time-of-flight cameras (Springer Berlin, 2009), pp. 726–737
32. D Eberley, *Skeletonization of 2D binary images*. (Geometric Tools, Tumwater, 2008). http://www.geometrictools.com/Documentation/Skeletons.pdf. Accessed 06 Aug 2013
33. RE Kalman, A new approach to linear filtering and prediction problems. J. Basic Eng. **82**(1), 35–45 (1960). doi:10.1115/1.3662552. http://fluidsengineering.asmedigitalcollection.asme.org/article.aspx?articleid=1430402
34. D Reynolds, Automatic speaker recognition using gaussian mixture speaker models. Lincoln Lab. J. **8**(2), 173–192 (1995). http://www.ll.mit.edu/publications/journal/journalarchives08-2.html#4
35. AP Dempster, NM Laird, DB Rubin, Maximum likelihood from incomplete data via the EM algorithm. J. R. Stat. Soc. **39**(Series B), 1–38 (1977)
36. M Mendoza, N Pérez De La Blanca, Applying space state models in human action recognition: a comparative study. Articulated Motion and Deformable Objects. **5098**, 53–62 (2008)
37. T Starner, J Weaver, A Pentland, Real-time american sign language recognition using desk and wearable computer based video. IEEE Trans. Pattern Anal. Mach. Intell. **20**, 1371–1375 (1998)
38. R Kuhn, JC Junqua, P Nguyen, N Niedzielski, Rapid speaker adaptation in Eigenvoice space. IEEE Trans. Speech Audio Process. **8**(6), 695–707 (2000)

39. O Thyes, R Kuhn, P Nguyen, JC Junqua, in *Sixth International Conference on Spoken Language Processing*. Speaker identification and verification using eigenvoices (ISCA, Baixas, 2000), pp. 1–3
40. CH Lee, CH Lin, BH Juang, in *1990 International Conference on Acoustics, Speech, and Signal Processing*. A study on speaker adaptation of continuous density HMM parameters (Albuquerque, 3–6 April 1990)
41. CJ Leggetter, PC Woodland, Flexible speaker adaptation using maximum likelihhod linear regression. Proc. ARPA Spoken Lang. Technol. Workshop. **9**, 110–115 (1995)
42. MJF Gales, PC Woodland, Mean and variance adaptation within the MLLR framework. Comput. Speech Lang. **10**(4), 249–264 (1996)
43. DA Reynolds, TF Quatieri, RB Dunn, Speaker verification using adapted gaussian mixture models. Digit. Signal Process. **10**(1–3), 19–41 (2000)
44. DA Reynolds, WM Campbell, *Text-independent Speaker Recognition*. (Springer, Berlin, 2008), pp. 763–782
45. N Cristianini, J Shawe-Taylor, *Support Vector Machines*. (Cambridge University Press, Cambridge, 2000)
46. WM Campbell, in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 1. Generalized linear discriminant sequence kernels for speaker recognition (IEEE Piscataway, 2002), pp. I-161–I-164
47. W Campbell, D Sturim, D Reynolds, A Solomonoff, in *IEEE International Conference on Acoustics, Speech and Signal Processing, 2006*, vol. 1. SVM based speaker verification using a GMM supervector kernel and NAP variability compensation (IEEE Piscataway, 2006), p. 97
48. A Stolcke, L Ferrer, S Kajarekar, E Shriberg, A Venkataraman, in *Proceedings of the 9th European Conference on Speech Communication and Technology*. MLLR transforms as features in speaker recognition (ISCA, Baixas, 2005), pp. 2425–2428
49. M Ferras, CC Leung, C Barras, JL Gauvain, in *Odyssey-2008*. MLLR techniques for speaker recognition. (2008), paper 023
50. SJ Young, D Kershaw, J Odell, D Ollason, V Valtchev, P Woodland, *The HTK Book Version 3.4*. (Cambridge University Press, Cambridge, 2006)
51. T Joachims, in *Making Large-scale Support Vector Machine Learning Practical*, ed. by B Schölkopf, CJC Burges, and AJ Smola. Advances in kernel methods (MIT Press Cambridge, 1999), pp. 169–184
52. K Morik, P Brockhausen, T Joachims, in *Proceedings of the 16th International Conference on Machine Learning (ICML-99)*. Combining statistical learning with a knowledge-based approach - a case study in intensive care monitoring (Morgan San Francisco, 1999)
53. J Shotton, A Fitzgibbon, M Cook, T Sharp, M Finocchio, R Moore, A Kipman, A Blake, in *IEEE Conference on Computer Vision and Pattern Recognition*. Real-time human pose recognition in parts from single depth images (IEEE Washington, D.C., 2011)
54. Q McNemar, Note on the sampling error of the difference between correlated proportions or percentages. Psychometrika. **12**(2), 153–157 (1947)

doi:10.1186/1687-5281-2013-44
Cite this article as: Gomez-Caballero *et al.*: **A statistical approach for person verification using human behavioral patterns.** *EURASIP Journal on Image and Video Processing* 2013 **2013**:44.

Submit your manuscript to a SpringerOpen journal and benefit from:

► Convenient online submission
► Rigorous peer review
► Immediate publication on acceptance
► Open access: articles freely available online
► High visibility within the field
► Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com