ON THE POTENTIAL OF ACTIVITY RELATED RECOGNITION

A. Drosou[1,2], K. Moustakas[1], D. Ioannidis[1] and D. Tzovaras[1]
[1]*Informatics and Telematics Institute, P.O. Box 361, 57001 Thermi-Thessaloniki, Greece*
[2]*Imperial College London, SW7 2AZ, London, U.K.*

Keywords: Biometric authentication, Biometrics, Activity recognition, Motion analysis, Body tracking, Hidden Markov models, HMM.

Abstract: This paper proposes an innovative activity related authentication method for ambient intelligence environments, based on Hidden Markov Models (HMM). The biometric signature of the user is extracted, throughout the performance of a couple of common, every-day office activities. Specifically, the behavioral response of the user, stimuli related to an office scenario, such as the case of a phone conversation and the interaction with a keyboard panel is examined. The motion based, activity related, biometric features that correspond to the dynamic interaction with objects that exist in the surrounding environment are extracted in the enrollment phase and are used to train an HMM. The authentication potential of the proposed biometric features has been seen to be very high in the performed experiments. Moreover, the combination of the results of these two activities further increases the authentication rate. Extensive experiments carried out on the proprietary ACTIBIO-database verify this potential of activity related authentication within the proposed scheme.

1 INTRODUCTION

It is well-known that biometrics can be a powerful cue for reliable automatic person identification and authentication. As a result, biometrics have recently gained significant attention from researchers, while they have been rapidly developed for various commercial applications ranging from surveillance and access control against potential impostors to smart interfaces. However, established physical biometric (Jain et al., 2004) identification techniques like fingerprints, palm geometry, retina and iris, and facial characteristics demonstrate a very restricted applicability to controlled environments. Thus, behavioral biometric characteristics (Jain et al., 2004), using shape based activity signals (gestures, gait, full body and limb motion), of individuals as a means to recognize or authenticate their identity have come into play.

1.1 Current Approaches

Emerging biometrics can potentially allow the non-stop (on-the-move) authentication or even identification in an unobtrusive and transparent manner to the subject and become part of an ambient intelligence environment. Previous work on human identification using activity-related signals can be mainly divided in two main categories. a) sensor-based recognition (Junker et al., 2004) and b) vision-based recognition. Recently, research trends have been moving towards the second category, due to the obtrusiveness of sensor-based recognition approaches.
Additionally, recent work and efforts on human recognition have shown that the human behavior (e.g. extraction of facial dynamics features) and motion (e.g. human body shape dynamics during gait), when considering activity-related signals, provide the potential of continuous authentication for discriminating people ((Ioannidis et al., 2007),(Boulgouris and Chi, 2007) and (Kale et al., 2002)).

Some first attempts that showed the potential of video-based recognition according to the face dynamics are usually categorized as follows: a) in the holistic method(displacements or the pose evolution) (Li et al., 2001). b)in the feature-based method (Chen et al., 2001), c) in the hybrid methods (Colmenarez et al., 1999) and d) in probabilistic frameworks (Liu and Chen, 2003). he most known example of activity-related biometrics is gait recognition (Boulgouris and Chi, 2007). On the other hand, shape identification using behavioral activity signals has recently started to attract the attention of the research community. Behavioral biometrics are related to specific actions and the way that each person executes them.

Earlier, in ((Kale et al., 2002) & (Bobick and Davis, 2001)) person recognition has been carried out

Drosou A., Moustakas K., Ioannidis D. and Tzovaras D. (2010).
ON THE POTENTIAL OF ACTIVITY RELATED RECOGNITION.
In *Proceedings of the International Conference on Computer Vision Theory and Applications*, pages 340-348
DOI: 10.5220/0002832703400348

using shape-based activity signals, while in (Bobick and Johnson, 2001), a method for human identification using static, activity-specific parameters was presented.

1.2 Motivation - The Proposed Approach

As a version of a multi-model concept for human authentication, a highly novel, innovative visual-based approach of a multi-activity scheme is proposed in the current work. Activity-related biometric features and the potential they show to increase the overall performance of an unobtrusive biometric system and to emerge the development of new algorithms for human recognition based on their daily activities, are investigated in the proposed framework.

The novelty of the approach lies in the fact that the measurements used for authentication will correspond to the response of the person to specific stimuli, while the biometric system is integrated in an ambient intelligence infrastructure. The proposed approach presents features that implicitly encode the behavioral and anthropometric characteristics of the individuals and is therefore proved to be extremely resistive against spoofing attempts.

Further, the integration of the activity-related identification capacities of different activities in an ambient intelligence environment and the combination of their results towards an improved identification possibility introduce a completely new concept in biometric authentication.

Specifically, biometric signatures, based on the user's response to specific stimuli, generated by the environment, are extracted, while the user performs specific work-related everyday activities, with no special protocol. Since the user is expected to remain sit, the most interesting characteristics to provide the biometric signature are the head and hands. The proposed approach is the first step in the exploration of such activity-related signals and their potential use in real applications. The proposed algorithm has been tested and evaluated in a large proprietary database.

The rest of the paper is organized as follows. In Section 2, an overview of the proposed system is presented. The upper-body tracker as a whole and its consisting image processing methods are described in Section 3, while in Paragraph 4.1 the signal processing methods, used to calculate final biometric signature of each user for a certain activity are briefly discussed. The Hidden Markov Model algorithm, mobilized in our approach to the problem of user identification, is presented in Paragraph 3.4, while a short description of the database used, follows in Section 5. Finally, the results of our work are presented and thoroughly discussed in Section 6.

2 SYSTEM OVERVIEW

The overall block diagram of the system is depicted in Figure 1. The user is expected to act with no constraints in an ambient intelligence office environment, in the context of a regular day at work. Meanwhile, some events, such as the ringing of the telephone or an instant message for online chatting, trigger specific reaction from the user. The user's movements are recorded by two of cameras and the raw captured images are processed, in order to track the user's head and hands. Specifically, the user's face is detected, based on haar-like features (Viola and Jones, 2001), motion history images (Bobick and Davis, 2001), skin color detection (Gomez and Morales, 2002), depth information extraction from disparity images as well as body-metric based restrictions support analysis of the raw pictures. The incorporation of these techniques in the in the first step of the proposed framework will be analyzed in the sequel.

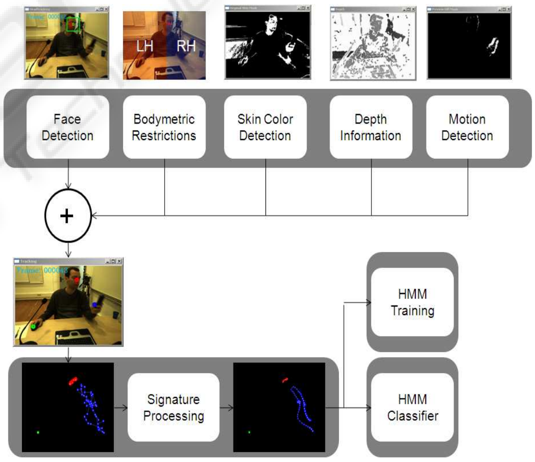

Figure 1: System overview.

The second step involves the processing of the tracked points acquired, so that the biometric, activity-related features of each user are revealed. Kalman Filtering (Welch and Bishop, 1995), followed by spatial interpolation is mobilized so as to generate the final biometric features. In the last step the use of continuous Hidden Markov Models (Rabiner, 1989) for the evaluation of the extracted biometric signatures is introduced. The proposed biometric system consists of two modes: a)The enrollment mode, whereby a user is registered by training an HMM from a certain number of his biometric signatures for each activity. b)The authentication mode, where the HMMs evaluate the claimed ID request by the user, as valid or void.

The effectiveness of the proposed system is demonstrated in Section 6 by two experiments. Experiment A examines the potential of a persons authentication using the biometric signature of just one activity as the identification metric, while Experiment B shows the authentication potential by combining the biometric signatures from two separate activities.

3 TRACKING UPPER-BODY

The detection of the head and hand points throughout the performed activity will be achieved with the successive filtering out of non-important regions of the captured scene. The methods and the algorithms mobilized for that are extensively discussed below.

3.1 Face

Tracking the face tracker implemented in the current approach is a combination of a face detection algorithm, a skin color classifier and a face tracking algorithm.

First, face detection is the main method that confirms the existence of one or more persons in the region under surveillance. The face detection algorithm, which has been implemented in our framework, is based on the use of haar-like feature types and a cascade-architecture boosting algorithm for classification, (i.e. AdaBoost) as described in (Viola and Jones, 2001) and in (Freund and Schapire, 1999). In case of multiple face detection, the front-most face (blob) is retained, utilizing the depth information provided by the stereo camera, while all others are discarded. Detected facial features allow the extraction of the location and the size of human face in arbitrary images, while anything else is ignored.

As an enhancement to the algorithm described above, the output of the latter is evaluated according to a skin classifier 3.2. In case the skin color restrictions are met, the detected face is passed over to an object tracker.

The face tracker, integrated in our work relies on the idea of kernel-based tracking (i.e. the use of the Epanechnikov kernel, as a weighting function), which has been proposed in (Ramesh and Meer, 2000) and later exploited and developed forward by various researchers.

3.2 Skin Color Classifier

A first approach towards the locating of the palms is achieved by the detection of all skin colored pixels on each image. Theoretically, the only skin colored pixels that should remain in a regular image in the end are the ones of the face and the ones of the hands of all people in the image (Figure 2). Each skin color modeling method defines the metric for the discrimination between skin pixels and non-skin pixels. This metric measures the distance (in general sense) between each pixel's color and a pre-defined skin tone.

The method incorporated in our work has been based on (Gomez and Morales, 2002). The decision rules followed, realize skin cluster boundaries of two color-spaces, namely the RGB((Skarbek and Koschan, 1994)) and the HSV((Poynton, 1997)), which render a very rapid classifier with high recognition rates.

Figure 2: a.Original image. - b.Skin colored filtered image.

3.3 Using Depth to Enhance Upper-body Signature Extraction

An optimized L1-norm approach is utilized in order to estimate the disparity images captured by the stereo *Bumblebee* camera of Point Gray Inc.(Scharstein and Szeliski, 2002). The depth images acquired (Figure 3) are gray-scale images, whereas the furthest objects are marked with darker colors while the closest ones with brighter colors.

Figure 3: Depth disparity image.

Given that the face detection was successful, the depth value of the head can be acquired. Any object with a depth value greater than the one of the head can now be discarded and thus excluded from our observation area. On the other hand, all objects in the foreground, including the user's hands, remain active. A major contribution of this filtering is the exclusion of all other skin colored items in the background, including other persons or surfaces in skin tones (wooden floor, shades, etc) and noise.

3.4 Using Motion History Images to Enhance Upper-body Signature Extraction

Another characteristic, which can be exploited for isolating the hands and the head of a human, is that they are moving parts of the image, in opposition to the furniture, the walls, the seats and all other static objects in the observed area. Thus, a method for extracting the moving parts in the scenery from a sequence of images could be of interest. For this reason, the concept of Motion History Images (MHI) (Bobick and Davis, 2001) was mobilized.

The MHI is a static image template, where pixel intensity is a function of the recency of motion in a sequence (recently moving pixels are brighter). Recognition is accomplished in a feature-based statistical framework. Its basic capability is to represent how (as opposed to where) motion in the image behaves within the regions that the camera captures.

Its basic idea relies on the construction of a vector-image, which can be matched against stored representations of known movements. A MHI H_T is a multi-component image representation of movement that is used as a temporal template based upon the observed motion. MHIs can be produced by a simple replacement of successive images and a decay operator, as described below in Eq. 1:

$$H_T(x,y,t) = \begin{cases} \tau, & \text{if D(x,y,t)=1} \\ max(0, H_T(x,y,t-1)-1), & \text{otherwise} \end{cases} \tag{1}$$

The use of MHIs has been further motivated by the fact that an observer can easily and instantly recognize moves even in extremely low resolution imagery with no strong features or information about the three-dimensional structure of the scene. It is thus obvious, that significant motion properties of actions and the motion itself can be used for activity detection in an image. In our case, if we set τ=2, we restrict the motion history depicted on the image to one frame in the past, which only concerns the most recent motion recorded (Figure 4). In other words, a mask image is

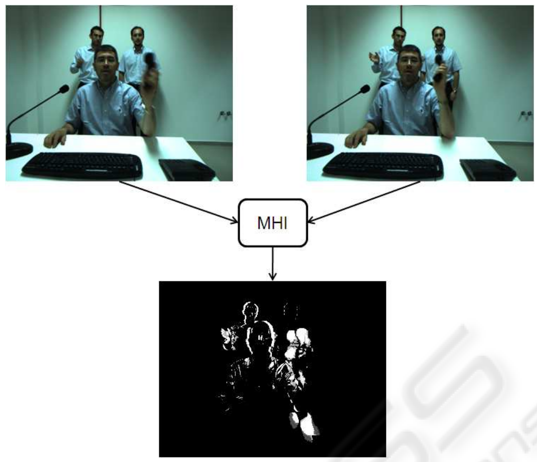

Figure 4: Motion history image generation.

created, where all pixels are black except for the ones, where movement last detected.

Summarizing, after having the face at least once successfully detected (red spot), the left/right hand location is marked with a blue/green spot on its location, when motion and skin color and the depth criteria are met (Figure 5). The tracking is only then complete, when the last image of the annotated sequence has been processed.

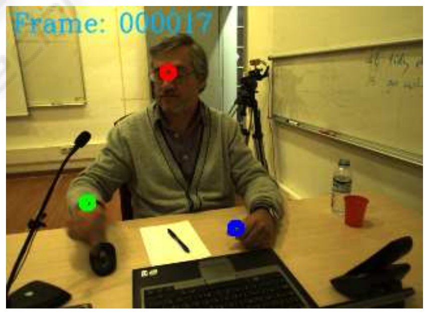

Figure 5: Tracked head & hands.

4 PREPROCESSING OF EXTRACTED SIGNATURES

Once the tracking process is over, the respective raw trajectory data is acquired. In this section the location points of all tracked body parts (head & hands) will be handled as a 3D signal. The block diagram of this refinement approach can be seen in Figure 6 and Figure 7. The refined, filtered raw data will result to an optimized signature (Wu and Li, 2009), which will carry all the biometric information needed for a user to be identified.

4.1 Filtering Processes

During the performance of the activity, there are some cases whereby the user stays still. Absence of motion as well as some inaccuracies in tracking may result in the absence of hand tracking. Within the first pre-interpolation step, the missing spots in the tracked points are interpolated between the last -if any- and the next -if any- valid tracked position.

The second step involves the smoothing of the 3D signals as it is implemented by a cumulative Moving Average method (Eq. 2). In this step short-term fluctuations, just like small fluctuations or perturbations of the exact spot of the hands due to light flickering or rapid differentiations in the capturing frame rate of the camera, are discarded, while longer-term movements, which show the actual movement of the head or hands, are highlighted.

$$CA_i = \frac{p_1 + ... + p_i}{i}\text{, whereby } p_i \text{ is each tracked point} \quad (2)$$

Further, the relative distances between the head and the hands are checked and the signals are trimmed accordingly. More specifically, the distance between the user's head and his hands is not expected to be bigger than a certain (normalized) value in the 3D space, due to anthropometric restrictions. Thus, in the unusual case of a bad tracked point, this point is substituted by the correct one or is simply discarded.

Next, each signal x_k undergoes a Kalman filtering process (Welch and Bishop, 1995). Since there are two points each time in the 3D images we are observing (the head and one hand), a six-dimensional Kalman filter is needed. The Kalman filter is a very powerful recursive estimator, where its stages are described by the following set of equations:

Predict:
predicted state

$$\hat{x}_{k|k-1} = F_k \hat{x}_{k-1|k-1}$$

predicted estimate covariance

$$P_{k|k-1} = F_k P_{k-1|k-1} F_k^T + Q_{k-1}$$

Update:

updated state estimate $\quad \hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k \bar{y}_k$

updated estimate covariance $\quad P_{k|k} = (I - K_k H_k) P_{k|k-1}$

whereby $K_k = P_{k|k-1} H_k^T (H_k P_{k|k-1} H_k^T + R_k)^- 1$,
$\bar{y}_k = z_k - H_k \hat{x}_{k|k-1}$;
$z_k = H_k x_k + v_k$;
F_k the state-transition model; H_k the observation model; v_k is the observation noise; R_k its covariance; and Q_k the covariance of the process noise.

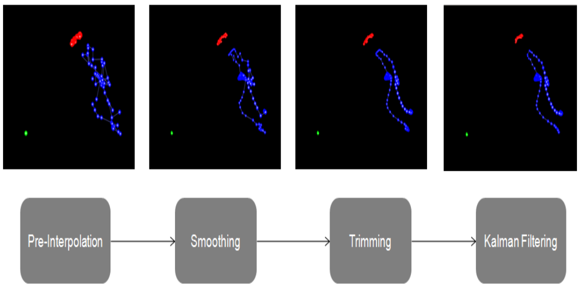

Figure 6: Signature signal processing.

Last but not least, a uniform resampling algorithm targetting interpolation is applied on the signals (Figure 7). The uniform interpolation ensures a uniform spatial distribution of the points in the final signature. The sampled points of the raw signature are rearranged, in such a way that a minimum and a maximum distance between two neighboring points is preserved. When necessary, virtual tracked points are added or removed from the signature. The result is an optimized, clean signature with a slightly different signature data set, without loss of the initial motion information (Wu and Li, 2009). With this, the proposed signature concerns more about the description for a continuous trajectory rather than a sequence of discretely sampled points.

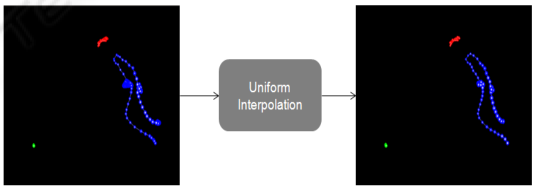

Figure 7: Uniform interpolation.

The proposed homogenized signature speeds up trajectory recognition, especially for large scale databases and image sequences.

4.2 Clustering and Classification using HMMs

At the enrollment and authentication stages, we use Hidden Markov Models (HMMs) to model each sign and classify according to the maximum likelihood criterion (Rabiner, 1989). HMMs are already used in many areas, such as speech recognition and bio-informatics, for modeling variable-length sequences and dealing with temporal variability within similar sequences (Rabiner, 1989). HMMs are therefore suggested in gesture and sign recognition to successfully handle changes in the speed of the performed sign or slight changes in the spatial domain.

In order to form a cluster of signatures to train an $HMM_{n,k}$ for the activity n of Subject k, we use three (3) biometric signatures, extracted from the same subject, when repeating the same activity. Each behavioral signature of activity-related feature vectors contains two 3D vectors (left/right hand and head times x,y and z position in space), corresponding to the biometric signature.

Specifically, each feature vector set $e_k(l_{head}, l_{hand})$ is used as the observation 3D vector for the training of the HMM of Subject k. The training of the HMM is performed using the Baum-Welch algorithm as proposed in (Rabiner, 1989). The classification module receives all the features calculated in the hand- and head-analysis modules as input. After the training, each cluster of behavioral biometric signatures of a subject is modeled by a five-state (5-state), left-to-right model, fully connected HMM, i.e. $HMM_{n,k}$.

In the authentication/evaluation process, given a biometric signature, we can claim that the inserted feature vectors form the observation vectors. Given that each HMM of the registered users is computed, the likelihood score of the observation vectors is calculated according to (Rabiner, 1989), where the transition probability and observation *pdf* are used. The biometric signature for an activity is only then recognized as Subject k, if the signature likelihood, returned by the HMM classifier, is bigger than an empirically set threshold.

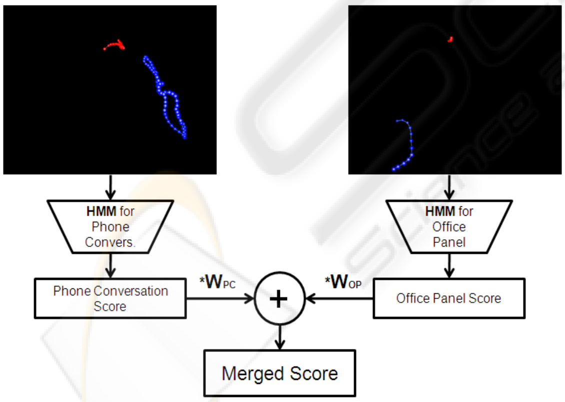

Figure 8: Combined scores.

In the proposed framework, the final decision is formed by the combination of the two activities' results. Specifically, the scores from each activity (i.e.*phone Conversation* & *Interaction with the panel*) contribute according to a weight-factor to the final score for a given subject Figure 8). Given that one activity usually demonstrates higher recognition capacity than the other one, it is logical that the final decision will mainly base on the outcome of the first activity, while the outcome of the latter will be supportive.

5 DATABASE DESCRIPTION

The proposed methods were evaluated on the proprietary ACTIBIO-dataset. This database was captured in an ambient intelligence indoor environment. Currently, there are two available sessions that were captured with a difference of almost six months. The first session consists of 35 subjects while the second includes 33 subjects, that have also participated in the first session. Briefly, the collection protocol had each person sat at the desk and act naturally, as if he is working. "NORMAL" activities, such as answering the phone, drinking water from glass, writing with a pen, interacting with a keyboard panel, etc. and "ALARMING" ones, such as raising the hands, etc. were captured. Five calibrated cameras have been constantly recording the scene:

- 2 usb cameras (Lateral, Zenithal) and 1 stereo camera (Bumblebee Point Grey Research) for office activity recognition)
- 1 camera (Grasshopper) for face dynamics/soft biometrics and 1 Pan-Tilt-Zoom Camera.

Further, a propriety annotator tool was developed so as to generate activity segments and to perform single and multiple searches for all annotated activities sequences, using any possible criterion.

6 EXPERIMENTAL RESULTS

The proposed algorithms have been tested on one large dataset and considerable potential in recognition performance has been seen in comparison. The proposed framework was evaluated in the context of three verification scenarios. Specifically, the potential of the verification of a user has been tested, based on his a)activity-related signature during a phone conversation, b) activity-related signature during the interaction with an office panel and c) the combination of the scores of the activity-related signatures from the two latter activities.

In order to test the robustness of the proposed algorithm, additive white Gaussian noise (AWGN) with a distribution of $\sigma = 0.02$ and zero mean value $\mu = 0$, has been added to all signatures of each user, while the trained HMMs remained the same, and the evaluation process was carried out again. The evaluation of the proposed approach in an authentication scenario mobilizes false acceptance/rejection rates (FAR, FRR) and equal error rates (EER) diagrams. The lower the EER, the higher the accuracy of the system is considered to be.

In Figure 9, the authentication potential of using just one signature, namely the one of the activity *Phone Conversation*, is demonstrated. It is obvious in Diagram 9, that the EER score lies at 15%, while in the "noisy" case (Diagram 10) the EER score is found at about 18%.

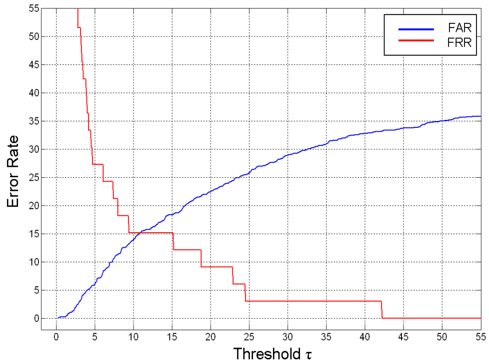

Figure 9: EER: activity phone conversation.

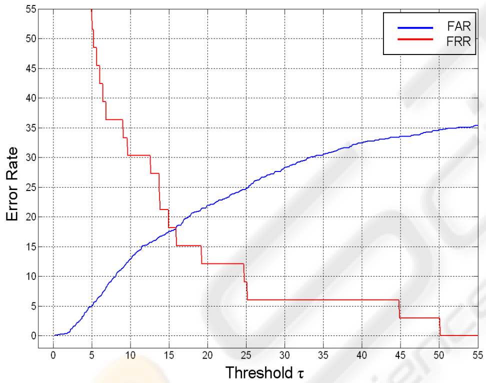

Figure 10: Activity phone conversation with AWGN ($\sigma = 0.02$).

The authentication capacity of the second activity which is the *Interacting with an Office Panel* is quite smaller as it can be seen in Diagram 11, whereby the EER is 9.8%. The insertion of white noise, however, takes it at an EER slight bigger than 12% (Diagram 12).

The overall performance of our system can be further improved, if we try to combine the recognition potential of each of the former activities. Since the second activity *Interaction with an Office* exhibits a higher recognition capacity than the *Phone Conversation* activity, it is logical to give a bigger weight factor to the first one. So, an EER score of less than 7.4%

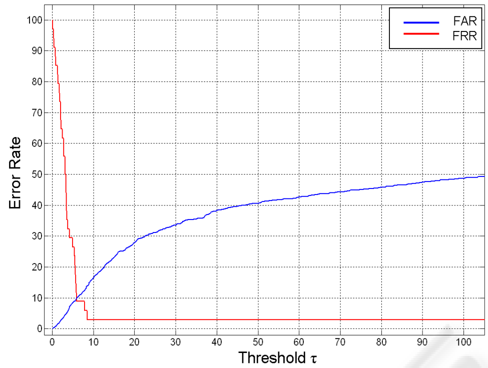

Figure 11: EER: activity office panel.

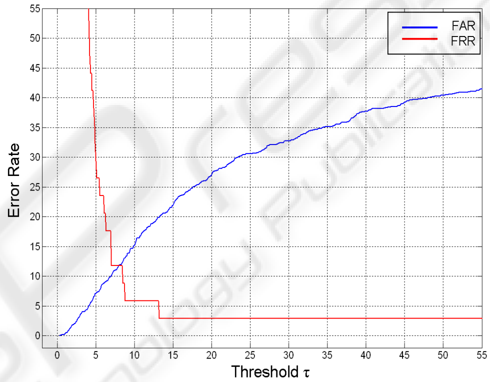

Figure 12: Activity office panel with AWGN ($\sigma = 0.02$).

(Diagram 13) is only then achieved, when the final score for each subject is formed by a higher contribution from the *Office Panel* activity and a lower contribution from the *Phone Conversation* one, as shown in Figure 8. It is obvious that in this case the insertion of noise has the least effects, since the EER score does not exceed the value of 9% in Diagram 14. Thus, the multi-activity scores approach is highly recommended for user verification purposes, since it is extremely difficult, both activities to be spoofed simultaneously.

The last two diagrams (Diagram 15 & Diagram 16) exhibit the degradation of the EER score in both test-cases, respectively, (noise-free and with the insertion of AWGN to the users' signatures respectively) in a clearer, summarizing way. The intersection of the graph with the diagonal line drawn (x=y), indicates the Equal Error Rate score of each type of signature/experiment.

To the authors' knowledge, this is the first attempt of using signatures extracted from everyday activities for biometric authentication purposes.

The great authentication potential exhibited in the current approach lies in the fact that the extracted features encode not only behavioral information, which

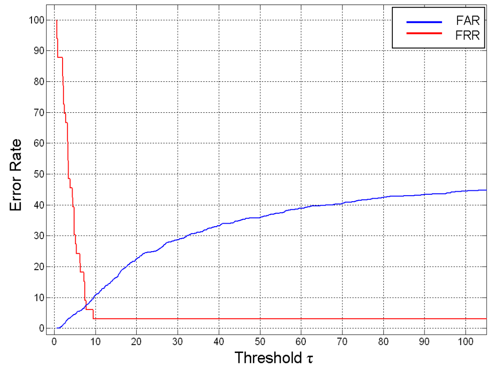

Figure 13: EER: merged activities.

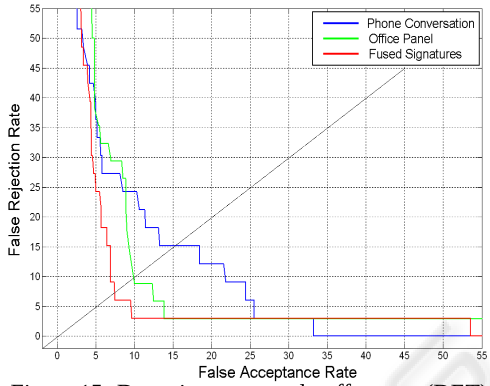

Figure 15: Detection error trade-off curves (DET).

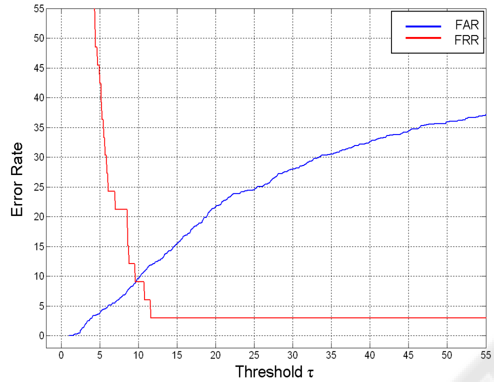

Figure 14: Merged activities with AWGN ($\sigma = 0.02$).

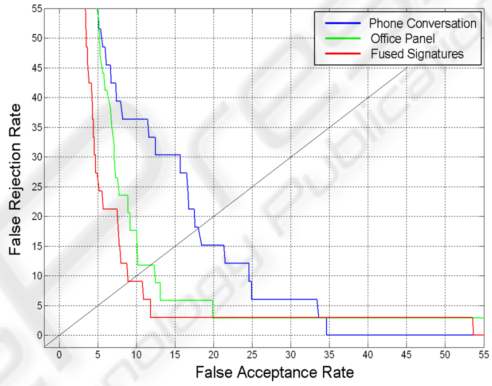

Figure 16: Detection error trade-off curves (DET) with AWGN ($\sigma = 0.02$).

relates to the way the user acts, but also to the user's anthropometric information, which describes the dimensions of his/her body and the relative positions of the body parts (head & hands).

7 CONCLUSIONS

In this paper, a novel method for human identification based on activity-related features is presented. The tracking of the user was handled by a robust visual-based upper-body tracker, while the activity-related signatures were processed by a series of successive filtering methods. Two office activities were combined and investigated for giving a robust identification result. The proposed system utilizes Hidden Markov Models to perform classification.

It has been experimentally demonstrated that the proposed approach on activity-related biometrics exhibits significant authentication potential, while being simultaneously totally unobtrusive for the users. A further advantage of the proposed system is that it allows the real-time, online recognition of the users, since it is very light in means of processing resources. The average frame rate recorded during the experiments had a value of 18fps.

The demonstrated results ignite further research in the the field of unobtrusive biometrics that could include the experimentation on new activities and signatures. Similarly, future work could also include the extraction of physiological characteristics - application of skeleton models in the tracking process - from the user and their insertion as further means in the recognition process. Given the existing technologies, one could not claim that the current biometric recognition module could be used as a stand-alone system. However, if combined with other modalities, such as face recognition or other soft biometrics, it is bound to exhibit great recognition results, by being in parallel totally unobtrusive.

ACKNOWLEDGEMENTS

This work was supported by the European Commission's funded project ACTIBIO.

REFERENCES

Bobick, A. and Davis, J. (2001). The recognition of human movement using temporal templates. *IEEE Trans. Pattern Anal. Mach. Intell.*, 23(3):257–267.

Bobick, A. F. and Johnson, A. Y. (2001). Gait recognition using static, activity-specific parameters. In *IEEE Proc. Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages 423–430.

Boulgouris, N. and Chi, Z. (2007). Gait recognition using radon transform and linear discriminant analysis. *IEEE Trans. Image Process*, 16(3):731–740.

Chen, L. F., Liao, H. Y. M., and Lin, J. C. (2001). Person Identification using facial motion. *IEEE Proc. International Conference of Image Proceeding*, 2:677–680.

Colmenarez, A., Frey, B., and Huang, T. S. (1999). A probabilistic framework for embedded face and facial expression recognition. *IEEE Proc. Computer Society Conference on Computer Vision and Pattern Recognition*, 1:592–597.

Freund, Y. and Schapire, R. E. (1999). A short introduction to boosting. *Proc. of the Sixteenth International Joint Conference on Artificial Intelligence (JSAI)*, pages 771–780.

Gomez, G. and Morales, E. F. (2002). Automatic feature construction and a simple rule induction algorithm for skin detection. In *Proc. of the ICML Workshop on Machine Learning in Computer Vision (MLCV)*, pages 31–38.

Ioannidis, D., Tzovaras, D., Damousis, I. G., Argyropoulos, S., and Moustakas, K. (2007). Gait Recognition Using Compact Feature Extraction Transforms and Depth Information. *IEEE Trans.Inf. Forensics Security.*, 2(3):623–630.

Jain, A. K., Ross, A., and Prabhakar, S. (2004). An Introduction to Biometric Recognition. *IEEE Trans. Circuits Syst. Video Technol.*, 14(1):4–20.

Junker, H., Ward, J., Lukowicz, P., and Tröster, G. (2004). User Activity Related Data Sets for Context Recognition. In *Proc. Workshop on 'Benchmarks and a Database for Context Recognition'*.

Kale, A., Cuntoor, N., and Chellappa, R. (2002). A framework for activity-specific human identification. In *IEEE Proc. International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 4, pages 3660–3663.

Li, B., Chellappa, R., Zheng, Q., and Der, S. (2001). Model-based temporal object verification using video. *IEEE Trans. Image Process*, 10:897–908.

Liu, X. and Chen, T. (2003). Video-based face recognition using adaptive hidden markov models. *IEEE Proc. Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 1:I340–I345.

Poynton, C. (1997). Frequently Asked Questions about Color.

Rabiner, L. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 53(3):257286.

Ramesh, D. and Meer, P. (2000). Real-Time Tracking of Non-Rigid Objects Using Mean Shift. In *IEEE Proc. Computer Vision and Pattern Recognition 2007 (CVPR)*, volume 2, pages 142–149.

Scharstein, D. and Szeliski, R. (2002). A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms. *International Journal of Computer Vision*, 47(1):7–42.

Skarbek, W. and Koschan, A. (1994). Colour image segmentation-a survey.

Viola, P. and Jones, M. (2001). Rapid Object Detection using a Boosted Cascade of Simple. In *IEEE Proc. Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages I511–I518.

Welch, G. and Bishop, G. (1995). An introduction to the Kalman filter.

Wu, S. and Li, Y. (2009). Flexible signature descriptions for adaptive motion trajectory representation, perception and recognition. *Pattern Recognition*, 42(1):194–214.