Multimed Tools Appl
DOI 10.1007/s11042-013-1635-1

User identification approach based on simple gestures

**Jože Guna · Emilija Stojmenova · Artur Lugmayr ·
Iztok Humar · Matevž Pogačnik**

Abstract We present an intuitive, implicit, gesture based identification system suited for applications such as the user login to home multimedia services, with less strict security requirements. The term "implicit gesture" in this work refers to a natural physical hand manipulation of the control device performed by the user, who picks it up from its neutral motionless position or shakes it. For reference with other related systems, explicit and well defined identification gestures were used. Gestures were acquired by an accelerometer sensor equipped device in a form of the Nintendo WiiMote remote controller. A dynamic time warping method is used at the core of our gesture based identification system. To significantly increase the computational efficiency and temporal stability, the "super-gesture" concept was introduced, where acceleration features of multiple gestures are combined in only one super-gesture template per each user. User evaluation spanning over a period of 10 days and including 10 participants was conducted. User evaluation study results show that our algorithm ensures nearly 100 % recognition accuracy when using explicit identification signature gestures and between 88 % and 77 % recognition accuracy when the system needs to distinguish between 5 and 10 users, using the implicit "pick-up" gesture. Performance of the proposed system is comparable to the results of other related works when using explicit identification gestures, while showing that implicit gesture based identification is also possible and viable.

Keywords Accelerometer · Gesture · Human-computer interaction · Non-invasive · User identification

1 Introduction and motivation

The personalization of applications and services has in recent years evolved from a desired feature to an almost mandatory feature. Additionally, the users are facing an increasing

J. Guna (✉) · E. Stojmenova · I. Humar · M. Pogačnik
Faculty of Electrical Engineering, University of Ljubljana, Tržaška 25, 1000 Ljubljana, Slovenia
e-mail: joze.guna@fe.uni-lj.si
URL: http://www.ltfe.org

A. Lugmayr
EMMi - Entertainment and Media Management Lab., Department of Business Information Management and Logistics, Tampere University of Technology (TUT),
Korkeakoulunkatu 8, 33101 Tampere, Finland

Published online: 16 August 2013

phenomenon of information overload and a user tailored selection of content is one of the logical solutions of the problem. Regardless of the approach used for personalized content recommendation, these systems need to identify the user before recommendations can be made. Consequently, user identification is used for logins in most of the multimedia systems including IPTV systems [15] offering personalized recommendations of TV shows and VoD content, converged IPTV and telephony services [20], as well as gaming consoles, where personalized settings, game progress, etc., are strongly bound to individual users. However, from the point of view of the technology, provision of these heterogeneous and multimedia rich applications is very demanding. Knowing the user's identity can alleviate this problem and thus be used to provide personalized delivery of multimedia rich content as well as a Quality of Experience (QoE) level adaptation and assurance [24].

User identification strategies are usually based on something the user owns (a card, a key, etc.) or on some secret information the user knows (a password of some kind). A lot of effort has been invested into prevention of accidental or malicious transfer of this secret information, thus many of the systems rely on biometrics for user identification, which identifies people using their physiological features [6]. Most advanced mechanisms include iris recognition [16], fingerprint recognition [13], hand geometry recognition [22], face recognition [1], and multi-modal biometrics [21]. As these approaches often require the integration of additional hardware and software components (e.g. special cameras, fingerprint readers) they are usually not widely integrated into mobile/home devices, such as mobile phones and remote controllers. In addition, these approaches require users to perform specific interventions. The readings of iris or fingerprints are thus considered as invasive activities in the identification process.

On the other hand, an increasing number of consumer electronic devices, including mobile phones and remote controllers, are equipped with accelerometers, enabling a device to "sense" how it is physically moved and manipulated by the user. This information can unobtrusively be used for detecting the users' identity. Of course, the greatest impact of the proposed method is seen when applied in traditionally multiuser environments, such as the living room equipped with a smart television or multimedia set-top-box/home theater PC devices. However, also when used on traditionally single user devices, such as smart mobile phones or tablet PCs, it can still be applied, but not primarily for personalization, but for unlocking of the device, for example. Also, tablet PCs (with Android 4.2) support multiple user profiles, where our proposed solution is again suitable.

Therefore, we have developed a new approach for user identification, named "Gesture Based user iDentification" method (GBiD). The term "gesture" in our work refers to a natural physical manipulation of the controlling device performed by the user. Standard control gestures (e.g. directional gestures, simple geometry shapes related gestures and word/letter related gestures), as presented in [8], can be classified as an example of explicit and well defined gestures. On the other hand, very loosely defined gestures, which are inherently embedded into other user actions, can be classified as implicit. An example of such implicit gesture is a simple "shake" of the control device (e.g. remote controller, mobile phone). An even more intuitive and implicit gesture is the "pick-up" gesture. It is performed when the user picks up the control device from its neutral motionless position and takes hold of it in a no particular predefined way. In the presented paper we will show, that implicit gesture based user identification using a single 3-axis accelerometer, based on loosely defined gestures, which can even be identical for all users, is possible and viable. A Dynamic Time Warping (DTW) method [9] is used in the proposed user identification system. The paper is a continuation of work presented at Pervasive/SAME conference [4] and includes a novel method of constructing recognition and processing wise more efficient dataset from raw data and corresponding new evaluation results.

The rest of the paper is organized as follows: related work is presented in Section 2; a description of the proposed implicit gesture based system is described in Section 3;

experiment setup and evaluation method in Section 4, results are presented in Section 5; while discussion, key conclusions and future work references are drawn in Sections 6 and 7 respectively.

2 Related work

The science of biometrics has received a lot of research attention because it provides a reliable and very adaptable method of confirming specific user's identity or distinguishing between identities of different users. A biometric system is essentially a pattern recognition system that operates by collecting, processing and matching various user biometric data. Depending on the specific application context, user verification or user identification modes are possible. In the verification mode (1:1 mode), a specific user's identity is either confirmed or denied by matching captured biometric data to templates stored in the system database. In the identification mode (1:*N* mode), an individual user is recognized among many (*N*) users by searching the templates of all users for a match. An introduction to biometric recognition is given in [5].

Haptics can also be employed for human authentication as shown in [18]. The authors explore the use of haptics, which allows a user to touch, feel, manipulate, create, and/or alter simulated three-dimensional objects in a virtual environment. Most of the existing applications of haptics are dedicated to hone human physical skills such as sensitive hardware repair, medical procedures, handling hazardous substances, etc. These skills can be trained in a realistic virtual world, and describe human behavioral patterns in human–computer interaction environments. The measurement of such psychomotor patterns can be used to verify a person's identity by assessing unique-to-the-individual behavioral attributes.

User body and hand gestures can be used as an effective and very intuitive way of interaction in virtual environments and manipulation of 3D data, as shown in [23]. While traditional gesture recognition systems are visual input based, recent advances in electronics allow for accurate and cost-effective gesture data acquisition using accelerometer sensor equipped devices. A lot of research effort has been given to user body and hand gesture recognition systems based on accelerometer data. Most acceleration based gesture recognition systems use Bayesian networks [2] or Hidden Markov Models for recognition [7, 19, 26]. At the same time, gestures used are either very simplified, such as directional movements, circles or Arabic numbers, or very different [8, 19] from each other, which, to some extent, simplifies the recognition process. Applications based on gesture recognition have been presented in [12].

Relatively little research effort has been given to acceleration based gesture user identification, however. Among relevant works we find [25], where authors describe a gesture based user verification system. The authors introduce a concept of "muscle memory" where a given suite of motor skills (a handwritten signature for example) through repetition over time and the ability through brain activity to inculcate and instill can be memorized in such a way they become unique for each person. Gestures used are personal signatures performed in space—"3D signatures", which are specific and very different for each user. The importance of this work is the idea, that locomotive physiological features can be used for user identification.

The work by Okumura et al. [14, 17] presents a gesture based biometric identification method using arm swing motions. The authors explore several identification algorithms, including squared error of Euclidean distance, error of angle and Dynamic Programming-matching. Gestures used need to be again very well defined in a way they are performed by users.

Another gesture based user authentication study has been reported by Liu et al. [10, 11]. The authors focus on feasibility and usability aspects of gesture based authentication and present an evaluation study results based on their own "uWave" recognition system. The implementation is based upon the dynamic time warping approach. The study reports very good results with 98 % success rate. However, only groups with five concurrent users were tested and gestures used are very deliberately chosen to be specific and very different for each user. Furthermore, gestures used are quite distinctive and well defined or explicit.

In comparison to other gesture based user identification systems works [10, 11, 14, 17, 25], our approach does not require the user to perform some predefined user interaction gestures, such as circles or signature gestures. Furthermore, gestures used in our system can be identical for all users and implicit in nature, describing the way they are performed by the user.

3 Intuitive gesture based user identification system

Our research has shown that it is possible to use the non-invasive and implicit gestures, to differentiate between users with sufficient statistical relevance. In contrast to the most known gesture recognition systems, our method supports a non-invasive and user-friendly approach, namely the usage of very simple and loosely defined user gestures (e.g. remote control "pickup" or "shake" gesture) Finally, the identification can be based on similar or identical gestures for all users.

3.1 Identification gestures and the dataset

For identification, three different types of gestures were selected: "a signature", "a shake" and "a pick-up" gesture. For comparison with existing gesture based identification systems we selected a personal name signature gesture in a form of particular user's name. For example, a user named Emilija would write "Emilija" with suitable hand movements in a form of a free 3D gesture in space. These gestures, according to our definition, can be classified as "well defined and explicit" and are different for each participant. On the other hand, the "shake" and "pick-up" gestures are classified as "implicit" in our identification experiment. A "shake" gesture depicts a movement, which occurs when the user simply shakes the control device in a no predefined way (e.g. remote controller, mobile phone).

An even more intuitive and implicit gesture represents the "pick-up" gesture. It describes the action of simply picking up the control device from its neutral position, usually resting on a flat surface (e.g. a table). This gesture is implicitly inherent in any active user-to-system interaction and thus convenient and very intuitive to use. When compared to the personal name signature gesture, the "pick-up" gesture is significantly less defined from the end user's point of view.

For practical reasons a simple, wide-spread and cost-effective solution was selected as a gesture acquisition device—the Nintendo WiiMote remote controller. With a built-in three-axis accelerometer and a guaranteed dynamic range of ±3g, 8 bits per axis resolution and 100 Hz sample rate frequency in average this device turned out be a reasonable choice. The recorded gestures are acquired as a time series and are stored in the form of a four column matrix (N x M). The first column contains the timestamp data, while the other three columns represent acceleration data in all three dimensions (*accX, accY, accZ*) respectively. Typical duration of performed gestures is in the interval of 1–5 s, thus corresponding to 100–500 data samples.

A typical example of the "pick-up" gesture acceleration footprint is presented in Fig. 1. The x-axis units are data samples, where 100 samples stretch over a 1 second period. The y-axis units are fractions of g-force and thus contain acceleration data, for samples in x, y and z

directions. They are presented with dotted-dashed blue (x), dashed green (y) and solid red lines (z). As all users are identified through execution of the same gesture, the acceleration footprints are similar for all users. Thus, user identification presents a greater technological challenge.

3.2 Dynamic time warping

Dynamic Time Warping (DTW) is a method for calculating the similarity between two time series. Time series compared can have different sampling rates and be of different lengths. Compared to simple Euclidean (Fig. 2 left) distance measure the DTW (Fig. 2 right) method has the ability to warp time axis and thus allows for optimal alignment between the two time series. Because gestures acquired with accelerometer sensors equipped devices can be typically described as multidimensional time series, the proposed method is very well suited for our purposes, as is also evident from related work [10, 12, 14, 17, 25].

Given two time series $X=(x_1,x_2,\ldots,x_i,\ldots,x_{|X|})$ and $Y=(y_1,y_2,\ldots,y_j,\ldots,y_{|Y|})$ of lengths $|X|$ and $|Y|$ where x_i, $y_j\in\Re$, a cost matrix D with dimensions $|X|$ by $|Y|$ is constructed. Each matrix element (x_i, y_j) corresponds to distance between elements x_i and y_j. Different distance measures $d(x_i, y_j)$ can be used whereas Euclidean distance measure is most commonly used.

A warp path $W=w_1, w_2, \ldots, w_r, \ldots, w_R$ with elements w_r of length R is then constructed. Each element $w_r=(i, j)$ corresponds to a point $(i, j)_r$ of the cost matrix D and represents a mapping or warping of elements of time series x_i to elements of time series y_j. The warp path starts at the beginning of each time series at $w_1=(1, 1)$, finishes at the end at $w_R=(|X|, |Y|)$ and increases monotonically. A sum of all distances $d(x_i, y_j)$ between the elements of both time series X and Y that the warping path connects is calculated. Different warp paths with different distance sums can be constructed. From all possible warp paths a minimum distance warp path is calculated, as shown by equation in Eq. (1).

$$DistDTW(X,Y) = \min W\left[\sum_{r=1}^{R} d(w_r)\right], \text{where}\, d(w_r) = d\left(x_i, y_j\right) \tag{1}$$

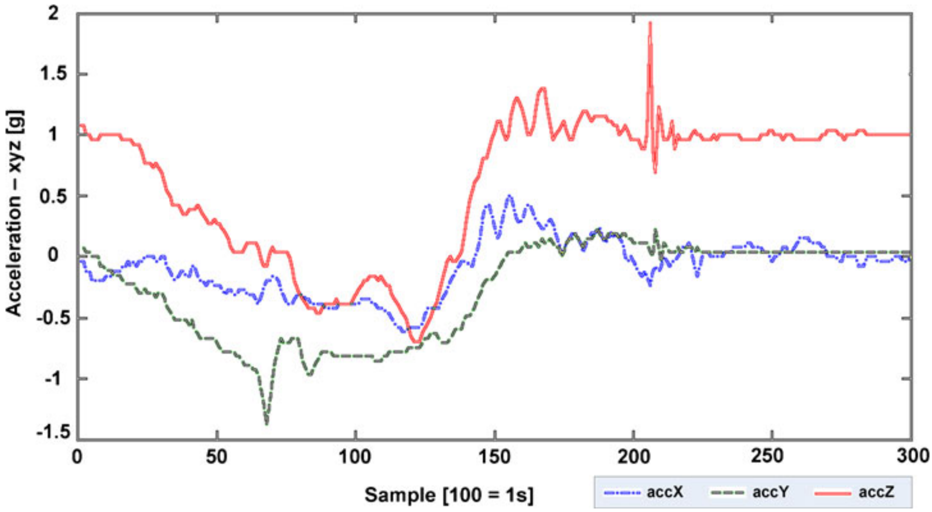


Fig. 1 Example of gesture "pick-up" acceleration footprint (processed in MatLab)

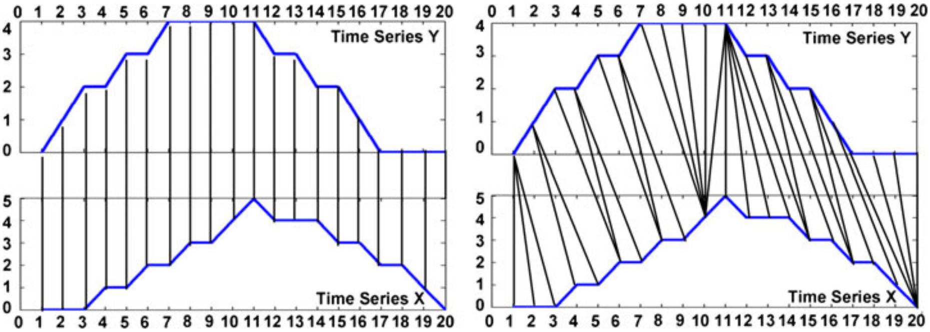


Fig. 2 Euclidean distance (*left*) vs. DTW time series similarity measure (*right*)

The minimum distance *DistDTW(*|X|,|Y|*)* is also used as a similarity measure between the two time series *X* and *Y*. If the two time series are identical, the similarity measure is zero. Otherwise it is greater than zero. An example of the DTW warp path is shown in Fig. 3.

3.3 Super-gesture templates

A "super-gesture" combines acceleration features of multiple original gestures in just one gesture. This is done with the purpose of maximizing the DTW similarity between the super-gesture and original gestures. Thus a dataset consisting only of super-gestures can be created by inherently incorporating acceleration features of multiple original learning gestures in a considerably fewer number of super-gestures, up to only one super-gesture template per each user.

Super-gestures are created by using information which the DTW method inherently uses to determine the similarity between the two time series (i.e. gestures) - the warp path itself. The warp path (W=w_1, w_2, …, w_r, …., w_R) records how elements of the

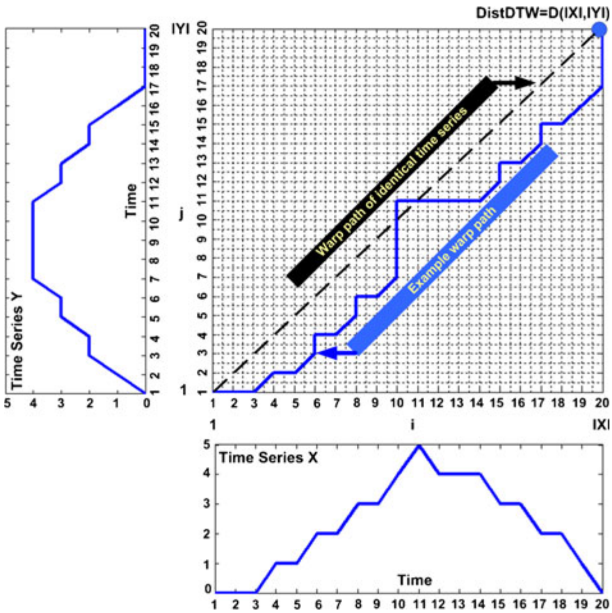


Fig. 3 Optimal minimum-distance DTW warp path

first time series $X=(x_1,x_2,\ldots,x_i,\ldots,x_{|X|})$, i.e. gesture X, are mapped $w_r=(i, j)$ to the elements of the second time series $Y=(y_1,y_2,\ldots,y_j,\ldots,y_{|Y|})$, i.e. gesture Y. The resulting super-gesture SG ($SG=sg_1, sg_2,\ldots,sg_r,\ldots,sg_R$) averages values of corresponding elements (x_i and y_j) of both time series belonging to the warp path w_R, as shown in Eq. (2):

$$SG = sg_1, sg_2, \ldots, sg_r, \ldots, sg_R; sg_r = \frac{x_i + y_j}{2}; w_r = (i, j) \tag{2}$$

This ensures maximum DTW similarity between original gestures and the resulting super-gesture. The length of the super-gesture equals the length of the warp path (R). Multiple gestures can be consecutively joined into one super-gesture.

The final result is a training dataset consisting of only 1 super-gesture template per each user instead of the multiple N original gestures per each user. Even though the resulting super-gestures are longer than the original gestures (approximately 20–30%, judging from experimental results) the training dataset gesture reduction (N-times) substantially increases the overall efficiency of the calculation during the gesture based identification process. An example of the resulting "pick-up" super-gesture acceleration footprint is shown in Fig. 4.

3.4 Gesture based user identification system

A Dynamic Time Warping method is used at the core of our implicit gesture based user identification method. Because of its inherent abilities to compare time series of different lengths and sample rates, and because it is simple to implement, it is well suited for this task.

An overview of the proposed gesture based user identification system (GBiD) is presented in Fig. 5. It consists of learning and identification phases. In the learning phase one or more training gestures TG_Y of user Y are recorded as input to the system. The system allows for recognition with as few as only one training gesture per user.

During the identification phase, a user X performs the identification gesture which is passed as input to the GBiD process as a validation gesture (VG_X). The validation gesture is compared to all gesture templates in the dataset belonging to all users using the DTW method and the best matching gesture (maximum gesture similarity which equals minimum DTW similarity distance) is selected (gesture Y). The gesture templates in the dataset can

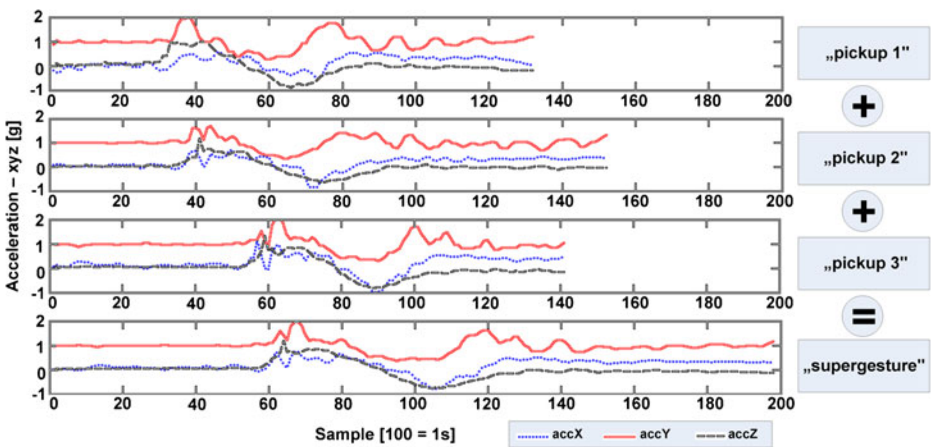


Fig. 4 Examples of original gestures and resulting super-gesture "pick-up" acceleration footprints

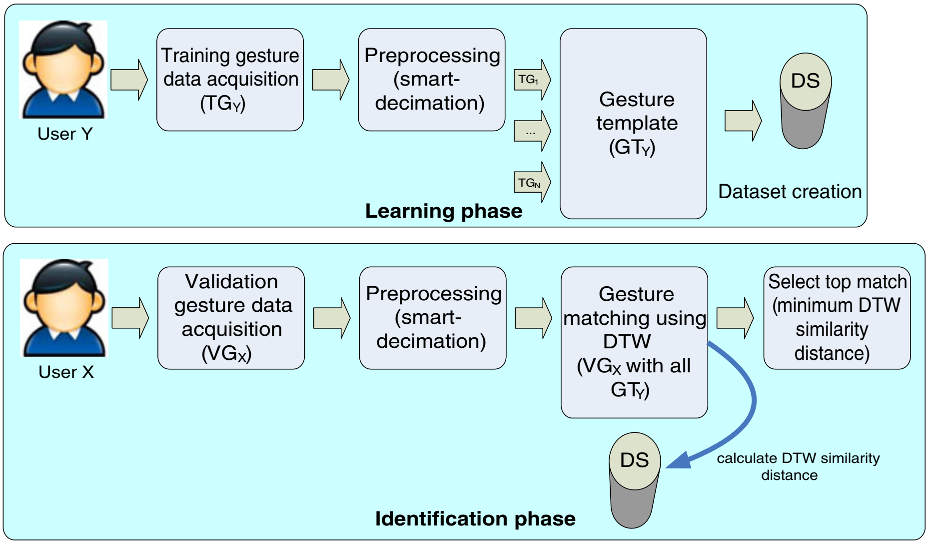


Fig. 5 Gesture based user identification system architecture

consist of either originally recorded gestures or of generated super-gestures. The identification is successful if the selected gesture Y indeed belongs to the user X.

3.4.1 Gesture pre-processing using the smart-decimation method

DTW performance is directly correlated to input time series length. Our "smart-decimation" method operates on acceleration data in all three axes simultaneously and removes only those vector gesture samples (consisting of timestamp information and x, y and z acceleration values) that do not change significantly over time and thus carry less relevant data. Gesture data represented as time series is first derived and then vector samples that do not change over the set gradient limit in all three dimensions simultaneously are removed. Thus all static gesture parts with limited acceleration changes, i.e. static user hand movements, are removed and the gesture length is significantly reduced.

3.4.2 DTW gesture matching and decision process

The validation gesture is compared to all gestures of the training dataset by pairs using the DTW method. For performance issues it is important that the training dataset consists of as few gestures as possible. Therefore, the super-gesture concept was introduced (see Section 3.3).

The DTW method originally processes only one-dimensional signals. To increase performance and accuracy we have extended the original DTW method to process three-dimensional signals which represent acceleration in all three Cartesian axes. The Euclidean distance measure used is extended to three dimensions as shown in Eq. (3).

$$d(X(i),Y(j)) = \sqrt{(X(i).x{-}Y(j).x)^2 + (X(i).y{-}Y(j).y)^2 + (X(i).z{-}Y(j).z)^2} \qquad (3)$$

A DTW similarity distance $DistDTW_i$=DTW(VG_X, GT_Y) (see Section 3.2) is calculated and used as a similarity measure to determine the best match. A list of top K-best matches with lowest DTW distances (i.e. highest DTW similarity) from training instances (GT_Y) is created. The K factor signifies how many of the best matches are further used in the decision process. In our study only the best match (a gesture with the highest DTW similarity in the data set) was used in the decision process (K=1), although multiple matches could be used and thus even better results could be achieved. We however wished to show, that even using only one best match for identification corresponding to one super-gesture template per user, is enough for successful implicit gesture based identification.

The system was implemented partly in C# and partly in MatLab. For flexibility and performance reasons a special gesture data acquisition application was developed in the . Net C# environment. Data preprocessing, analysis and validation algorithms were implemented in MatLab.

4 Experiment setup and evaluation method

Ten participants took part in the user evaluation study. Seven of them were male and three were female. The youngest participant was 25 years old, whilst the oldest participant was 45 years old. The average participant's age was 32 years (6.21 standard deviation). Three participants declared themselves as expert users of multimedia systems and four as regular users. Three participants said they were occasional users and only had little experience with multimedia systems. Previous experience with multimedia systems was assessed on a five point Likert scale, where 1 indicated little or no experience and 5 an expert user. An expert user has a lot of experience with multimedia systems and control devices on different devices (set-top-box, smart phone, tablet PC, etc.) on a daily basis. Among the participants one user was left handed while the other were right handed.

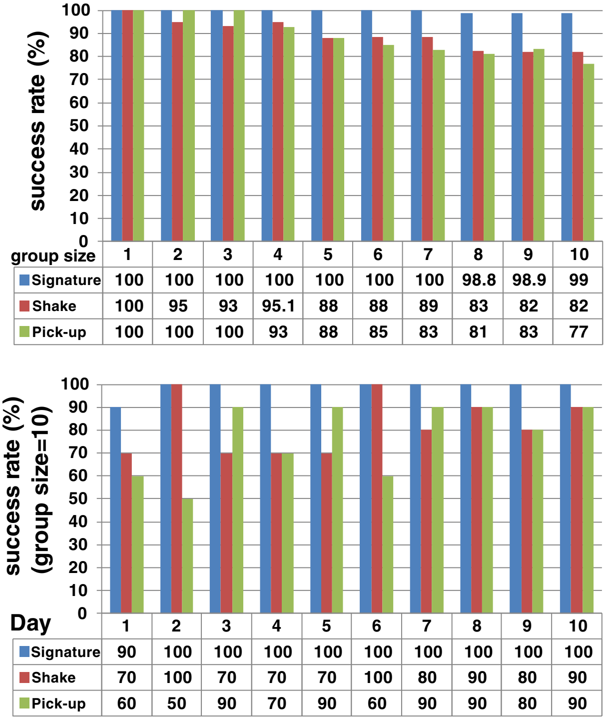
The main experimental objective was to determine the recognition success rate of the proposed method dependent on the size of target group population and also in comparison to the traditional gesture based user recognition using well defined explicit (e.g. signature gestures) and different gestures per each user.

The experiment itself was conducted over a period of 11 days in order to learn about temporal stability and recognition success rate. On the first day (day 0) three training gestures for each selected gesture type (personal signature gesture, "shake" and "pick-up" gestures) per each user were collected. Then on each following day (day 1 to day 10) only one additional validation gesture for each selected gesture type per each user was collected. A 10-fold temporal cross validation procedure was used for evaluation of the proposed system. Appropriate dataset consisting of all collected gestures was dynamically constructed and validation gestures were compared to the dataset (e.g. dataset for day 5 consisted of original three training gestures recorded on day 0 as well as from five validation gestures recorded from days 1 to 5 and was used to validate gestures recorded on day 6). Two datasets were constructed, one using original gestures and another using super-gestures.

5 Results

User recognition success rate is expressed as a number of successful user recognitions vs. all recognition attempts. Results obtained from processing original gesture datasets and datasets using super-gestures are presented in graphs in Figs. 6 and 7, respectively.

The presented graphs in Fig. 6 (left) and Fig. 7 (left) illustrate the various gesture type detection methods in relationship to successful user recognitions as a function of various user group sizes (from 1 to 10 users). The results are cummulative, thus the percentage of success rates is given over a period of time of the full 10 test days for



group size	1	2	3	4	5	6	7	8	9	10
Signature	100	100	100	100	100	100	100	98.8	98.9	99
Shake	100	95	93	95.1	88	88	89	83	82	82
Pick-up	100	100	100	93	88	85	83	81	83	77

Day	1	2	3	4	5	6	7	8	9	10
Signature	90	100	100	100	100	100	100	100	100	100
Shake	70	100	70	70	70	100	80	90	80	90
Pick-up	60	50	90	70	90	60	90	90	80	90

Fig. 6 User recognition success rate (in %) for various user target group sizes (*above*) and a detailed daily results for target group size of ten users (*below*) for the original dataset

a particular user group size. For example, for target group size of 10 users, a total of 100 recognition attempts were performed per each recognition gesture type—10 users X 10 days.

Figure 6 (right) and Fig. 7 (right) show the detailed changes of the results of various detection methods over a time period of 10 days with a constant group size of 10 users.

6 Discussion

The results of our research clearly show that motion gesture based user identification is possible. For comparison with relevant related works, our proposed gesture based identification system was tested on the explicit identification gestures, such as user signatures, as well as on the implicit "shake" and "pick-up" gestures. Two sets of results obtained from the same original data are presented—dataset based on original data and dataset based on optimized "super-gestures" data.

When discussing the results based on the original gesture dataset, the experimental results show that nearly 100 % recognition success rate (lowest score was 98.8 % on group size of 8 users) was achieved when users used their name signatures as identification gestures. This suggests that the Dynamic Time Warping method performs very well when comparing distinct and different input time series. Consequently we can conclude that our system's performance is comparable to results of other respectable related works.

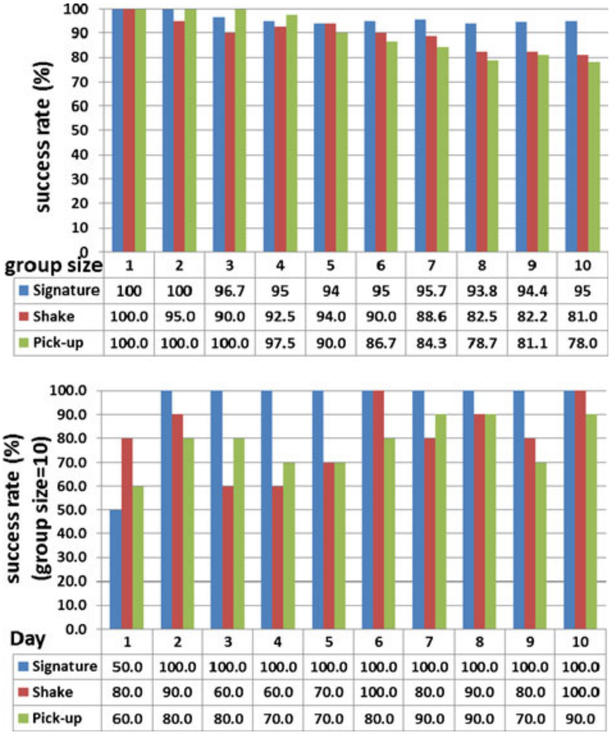


Fig. 7 User recognition success rate (in %) for various user target group sizes (*above*) and a detailed daily results for target group size of ten users (*below*) for the super-gestures dataset

To estimate how user's gesture execution may change over time, the experiment was conducted over a period of 10 days. The achieved average recognition success rate tested on 10 users is 82 % for the "shake" gesture and 77 % for the "pick-up" gesture as shown in Fig. 6 right. More specifically, the results range from 70 % on day 1 to 90 % on day 10 for the "shake" gesture and from 60 % on day 1 to 90 % on day 10 for the "pick-up" gesture. This suggests good overall temporal stability of the proposed system, however recognition success rate variations on days 2–3 and 6–7 for "shake" and "pick-up" gesture suggest different users' execution of identification gestures. The gesture based user identification system learns from users as they use it and thus provide more training gesture data. Users can perform their identification gestures in any way they wish and do not have to adjust to the system.

When comparing the results of various target user group sizes (Fig. 6 left), we achieved 95 % ("shake") and 93 % ("pick-up") recognition success rate for the group size of 4 users. The results then drop to approx. 88 % ("shake") and 85 % ("pick-up") recognition success rates for the group sizes of 5–7 users.

When discussing the results based on the "super-gestures" dataset, the results obtained are comparable or better than results from using the original (single) gesture dataset. For target group size of 5 users the results are clearly better – 94 % vs. 88 % for the "shake" gesture and 90 % vs. 88 % for the "pick-up" gesture. For a target group size of 10 users the results are comparable – 81 % vs. 82 % for the "shake" gesture and 78 % vs. 77 % for the "pick-up" gesture. There is an interesting discrepancy when comparing results using the name signature gesture, however. The results are still excellent and clearly higher recognition rates were achieved using the name signature gesture compared to implicit "shake" and"pick-up" gestures. But, interestingly enough, lower results were obtained when compared to the results from the original gesture datasets. At the closer examination of daily results for different group sizes for the name signature gesture we find a specific discrepancy for group sizes of 3 to 10 and only for the recognition results on the first day (day 1). It turns out that the users in question were named "Matevž" and "Mateja". Since the "super-gestures" concept was based on the idea of combining different gestures into one gesture thus maximizing the similarities, the two users were misidentified because of their similar names and consequently similar personal name gestures. The effect is also present in the results obtained from the original gesture dataset, but less evident. This can be seen on detailed daily results for group size of 10 users (Figs. 6 and 7 right) for both datasets. The recognition success rate on day 1 for the original dataset was 90 % vs. 50 % for the "super-gesture" dataset.

Beside better overall results the "super-gesture" concept also allows for a much more optimized processing wise solution. The reduction of dataset from N original gestures to a significantly less super-gestures (or even only one) correspondingly reduces the DTW processing by the same ratio. In our case at the group size of 10 users the reduction rate is 130 (original gesture dataset size) : 10 (super-gesture dataset size).

According to [3] the average household size in Europe is less than three users per family. This suggests that the proposed system could be used for non-critical and intuitive user identification in a home family environment for a number of multimedia applications and services. The proposed identification system could also be used on mobile devices for simple functions, such as unlocking the device with a simple shake or twist. However, this is not the primary application of the proposed system and additional single post-processing (e.g. motion compensation) will be required due to the possible external environmental movement distractions (e.g. walking movement).

Overall, the results for the "shake" and "pick-up" gestures are understandably lower from the ones obtained for the "signature" gesture, because the gestures used were the same for all users. Furthermore, slightly lower results for the "pick-up" gesture indicate that this gesture is less distinct and thus more implicit. The trade-off is of course usage of a simple and intuitive gesture, which was one of the main goals of our work.

The proposed identification system, as any other biometric system, requires the initial user gesture templates in order to function. The initial data required is small, typically a few gestures of the same type per user (e.g. 3), although even one initial gesture may be enough in some cases. However, this requires the active participation of the user at the beginning. To make the process easier or simply more "fun" for the user, the initial gesture templates could be obtained using a simple game—the gamification approach. Users' actions could be visualized (using colours, changes of shapes and sizes) and used for control of the avatar, performing simple tasks during the data gathering. The whole process is not time consuming, and should mostly take a couple of minutes. Consequently, we don't see this as a problem, which needs significant attention.

7 Conclusions

The experimental results have shown that implicit gesture based identification of users is possible with statistical relevance. Even though this level of accuracy may not yet be suitable for employment in services with very strict security demands like online banking, etc., the results are satisfactory for typical home multimedia services and applications. Further experiments will be conducted using a number of different gestures such as twist, swing, etc., combined with additional gyroscopic data. In addition, combinations of gestures will be evaluated such as "pick-up and shake", as this approach may further increase the identification accuracy. Different gestures or their combinations may prove to provide a more reliable identification and the system itself could propose the most suitable gestures for certain types of services.

Furthermore, we will test our approach on the young and the elderly in order to assess the success rate in a typical home family environment. We expect that greater physiological user diversity should even improve the performance.

During these experiments, more than one gesture will be used in the decision process (e.g. K=3) and other soft decision making mechanisms will be evaluated like fuzzy logic, for example. This may increase the recognition accuracy and more importantly, provide a better measure of certainty of the recognition decision.

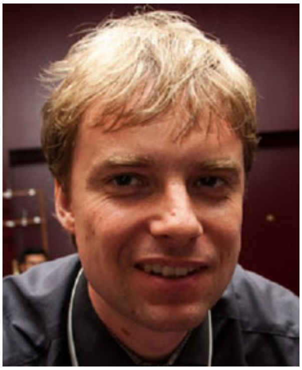
With the increasing amount of available Web and multimedia content, the need for personalized experience is of paramount importance. Personalization requires identification of users, which is the main motivation behind the presented research effort. Simple, intuitive and non-invasive identification approaches are the ones which will dominate future use of home, business, mobile and other multimedia devices and services. We see a strong potential for the system to succeed, due to it's simplicity and it's easy usage.

Acknowledgments The operation that led to this paper is partially financed by the European Union, European Social Fund and the Slovenian research Agency, grant No. P2-0246. The authors also thank the participants who took part in the evaluation study.

References

1. Abate AF, Nappi M, Riccio D, Sabatino G (2007) 2D and 3D face recognition: a survey. Pattern Recognit Lett 28:1885–1906
2. Cho SJ, Oh JK, Bang WC, Chang W, Choi E, Jing Y, Cho J, Kim DY (2004) Magic wand: A Hand-Drawn Gesture Input Device in 3-D Space with Inertial Sensors. Proceedings of the 9th Int'l Workshop on Frontiers in Handwriting Recognition
3. Eurostat Average household size, http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=ilc_lvph01&lang=en
4. Guna J, Stojmenova E, Lugmayr A, Humar I, Pogačnik M (2012) User identification approach based on simple gestures, Proceedings of the 5th International Workshop on Semantic Ambient Media Experience (SAME) - in conjunction with Pervasive 2012 [Elektronski vir] : Newcastle, UK, 18th June 2012/[editors] Artur Lugmayr … [et al.]
5. Jain AK, Flynn P, Ross AA (2008) Handbook of Biometrics. Springer http://www.springer.com/computer/image+processing/book/978-0-387-71040-2
6. Jain AK, Ross A, Prabhakar S (2004) An introduction to biometric recognition. IEEE Trans Circ Syst Video Technol 14(1):4–20. doi:10.1109/TCSVT.2003.818349
7. Jun-qi K, Hui W, Guang-quan, Z (2009) Gesture recognition model based on 3D Accelerations. Proceedings of 2009 4th InternationalConference on Computer Science & Education
8. Kela J, Korpipää P, Mäntyjärvi J, Kallio S, Savino G, Jozzo L, Marca D (2006) Accelerometer-based gesture control for a design environment. Pers Ubiquit Comput 10:285–299
9. Keogh E, Pazzani M (2001) Derivative dynamic time warping. Proc. of the First Intl. SIAM Intl. Conf. on Data Mining, Chicago, Illinois
10. Liu J, Wang Z, Zhong L, Wickramasuriya J, Vasudevan V (2009) "uWave: Accelerometer-based Personalized Gesture Recognition and Its Applications". Proc. IEEE Int. Conf. Pervasive Computing and Communication (PerCom)

11. Liu J, Zhong L, Wickramasuriya J, Vasudevan V (2009) User evaluation of lightweight user authentication with a single tri-axis accelerometer. Proc. ACM Int. Conf. Human Computer Interaction with Mobile Devices and Services (MobileHCI), September
12. Lugmayr A, Dorsch T, Humanes PR, Vallverdu J, ... (eds) (2009) Handbook of research on synthetic emotions and sociable robotics: new applications in affective computing and artificial intelligence, p 443-459. Information Science Reference, IGI Global, Hershey, New York
13. Maltoni D, Maio D, Jain A, Prabhakar S (2009) Handbook of fingerprint recognition. Springer professional computing, 2nd edn, XVI, p 496. 205 illus.. with DVD
14. Matsuo K, Okumura F, Hashimoto M, Sakazawa S, Hatori Y (2007) Arm swing identification method with template update for long term stability. Proc. Int. Biometrics
15. Montpetit MJ, Mirlacher T, Ketcham M (2010) IPTV: An end to end perspective. Journal Commun 5(5):358–373
16. Nabti M, Bouridane A (2008) An effective and fast iris recognition system based on a combined multiscale feature extraction technique. Pattern Recognit 41:868–879
17. Okumura F, Kubota A, Hatori Y, Matsuo K, Hashimoto M, Koike A (2006) A Study on biometric authentication based on arm sweep action with acceleration sensor. Proc. Int. Symp. Intelligent Signal Processing and Communications
18. Orozco M, Graydon M, Shirmohammadi S, El Saddik A (2008) Experiments in haptic-based authentication of humans. Multimed Tools Appl 37:73–92
19. Schlömer T, Poppinga B, Henze N, Boll S (2008) Gesture recognition with a Wii Controller. Proceedings of TEI, 11–14
20. Sedlar U, Zebec L, Bester J, Kos A (2008) Bringing click-to-dial functionality to IPTV users. IEEE Commun Mag 46:118–125
21. Toledano DT, Pozo RF, Trapote AH, Gómez LH (2006) Usability evaluation of multi-modal biometric verification systems. Interacti Comput 18:1101–1122
22. Varchol P, Levický D (2007) Using of hand geometry in biometric security systems. Radioengineering:16 (4):82–87
23. Varona J, Capó AJ, Gonzàlez J, Perales FJ (2009) Toward natural interaction through visual recognition of body gestures in real-time. Interact Comput 21:3–10
24. Volk M, Sterle J, Sedlar U, Kos A (2010) An approach to modeling and control of QoE in next generation networks. IEEE Commun Mag 48:126–135
25. Zaharis A, Martini A, Kikiras P, Stamoulis G (2010) User authentication method in implementation using a three-axis acceleorometer. Mobile Lightweight Wireless Systems, Second International ICST Conference, MOBILIGHT, Springer
26. Zappi P, Milosevic B, Farella E, Benini L (2009) Hidden Markov Model based gesture recognition on low-cost, low-power tangible user interfaces. Entertain Comput 1:75–84



Jože Guna obtained his B.Sc. and M.Sc. degrees from the Faculty of Electrical Engineering at the University of Ljubljana, Slovenia in the field of telecommunications. He is an assistant, active researcher, and project manager with extensive knowledge and extpertise in the field of IP networking and interactive IP multimedia, with special focus on Ux and multimodal interaction approaches for the iTV environment. He is a member of IEEE, UxPA organizations and editor at IJATES scientific journal.

Emilija Stojmenova graduated in the field of electrical engineering at the Faculty of Electrical Engineering and Computer Science, University of Maribor, Slovenia in 2009. She joined Iskratel in 2009 as a young researcher from the industry and is responsible for the coordination of design and development following user-centered approaches. She has worked in the field of interactive TV and other telecommunication applications focusing on user-oriented development ranging from ethnographic studies to understand user needs to the development of new forms of user experience and usability evaluation. Her community engagement includes: WUD Slovenia 2010 Conference Chair and WUD Slovenia 2011 Conference Chair. Currently she is working on her PhD on user-centered design methodologies for applications for various devices used by specific user groups.

Prof. Dr. Artur Lugmayr describes himself as a creative thinker and his scientific work is situated between art and science. Starting from July 2009 he is full-professor for entertainment and media production management at the Department of Business Information Management and Logistics at the Tampere University of Technology (TUT): EMMi—Entertainment and Media Production Management (http://www.tut.fi/emmi/). His vision can be expressed as to create media experiences on future emerging media technology platforms. He is the head and founder of the New AMbient MUltimedia (NAMU) research group at the Tampere University of Technology (Finland) which is part of the Finnish Academy Centre of Excellence of Signal Processing from 2006 to 2011 (http://namu.cs.tut.fi). He is holding a Dr.-Techn. degree from the Tampere University of Technology (TUT, Finland), and is currently engaged in Dr.-Arts studies at the School of Motion Pictures, TV and Production Design (UIAH, Helsinki). He chaired the ISO/IEC ad-hoc group "MPEG-21 in broadcasting"; won the NOKIA Award of 2003 with the text book "Digital interactive TV and Metadata" published by Springer-Verlag in 2004; representative of the Swan Lake Moving Image & Music Award (http://www.swan-lake-award.org/); board member of MindTrek (http://www.mindtrek.org), EU project proposal reviewer; invited key-note speaker for conferences; organizer and reviewer of several conferences; and has contributed one book chapter and written over 25 scientific publications. His passion in private life is to be a notorious digital film-maker. He is founder of the production company LugYmedia Inc. (http://www.lugy-media.tv). More about him on http://www.tut.fi/emmi.

Iztok Humar received B.Sc., M.Sc. and PhD degrees in the field of Telecommunications at the Faculty of Electrical Engineering, University of Ljubljana, Slovenia, in 2000, 2003, and 2007 respectively. He also received PhD in the field of Information Management at the Faculty of Economics, University of Ljubljana, Slovenia, in 2009. In 2000 he joined Laboratory for Telecommunications as a researcher and teaching assistant. From 2004–2009 he was Senior Lecturer. Currently, he is an Assistant professor at the Faculty of Electrical Engineering, Slovenia, where he lecturers on The Fundamentals of Electromagnetics and Design, Management and Modeling of Telecommunication Networks. He is supervisor of many undergraduate and some postgraduate students. His main research topics include the design, planning and management of telecommunications networks and services, as well as measurement and modeling of network loads and traffic. In 2010, he was a 3-month Visiting Professor and Researcher at Huazhong University of Science and Technology, Wuhan, China, where he gave lectures and performed research on energy-efficient wireless networks. He is involved in many research and national industry projects as well in projects funded by EU FP and structural and Cohesion EU funds. Dr. Humar served as Chair of Vitel—national workshop from the field of telecommunications. He is a Senior member of IEEE Communication Society and a member of Electrical Association of Slovenia. He serves as IEEE Communication Society of Slovenia Chapter Chair.

Matevž Pogačnik graduated in 1997 and defended his Ph. D. thesis in 2004 in the field of telecommunication and informatics at the University of Ljubljana. He is presently employed as an assistant professor with the Faculty of Electrical Engineering in Ljubljana, where he heads the multimedia section of the Laboratory for Telecommunications. His research and scientific work is focused on development of interactive multimedia services for different devices including the development of systems for user-adapted content choice (personalization). Lately he’s been mostly involved in development of interactive services on all devices in the fields of e-learning and interactive TV systems using different interaction modalities. Matevž Pogačnik participated in numerous European projects in the field of interactive digital television, e-learning, e-tourism and sensor systems He is also a member of the international organizations IEEE and UxPA.