

J. Guna, E. Stojmenova, A. Lugmayr, I. Humar, and M. Pogacnik, "User identification approach based on simple gestures," in Proceedings of the 5th International Workshop on Semantic Ambient Media Experience (SAME) - in conjunction with Pervasive 2012, A. Lugmayr, T. Risse, B. Stockleben, J. Kaario, B. Pogorelc, and E. S. Asensio, Eds. Newcastle, UK: Tampere University of Technology (TUT), June 2012.

User identification approach based on simple gestures

¹Jože Guna, ²Emilija Stojmenova, ³Artur Lugmayr, ¹Iztok Humar and ¹Matevž Pogačnik

¹Faculty of Electrical Engineering, University of Ljubljana, Slovenia

²IskrateL, Kranj, Slovenia

³Tampere University of Technology, Tampere, Finland

joze.guna@fe.uni-lj.si, stojmenova@iskratel.si, artur.lugmayr@tut.fi, iztok.humar@fe.uni-lj.si,
matevz.pogacnik@fe.uni-lj.si

Abstract. This paper presents 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. 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 phenomenon of information overload and 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 [1] offering personalized recommendations of TV shows and VoD content, converged IPTV and telephony services [2], 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 [3].

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 [4]. Most advanced mechanisms include iris recognition [5], fingerprint recognition [6], hand geometry recognition [7], face recognition [8], and multimodal biometrics [9]. 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. 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 [10], 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 [11] is used in the proposed user identification system.

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 [12].

Haptics can also be employed for human authentication as shown in [13]. 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 [14]. 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 [15] or Hidden Markov Models for recognition [16-18]. At the same time, gestures used are either very simplified, such as directional movements, circles or Arabic numbers, or very different [18-19] from each other, which, to some extent, simplifies the recognition process. Applications based on gesture recognition have been presented in [20].

Relatively little research effort has been given to acceleration based gesture user identification, however. Among relevant works we find [21], 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. [22-23] 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 are 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. [24-25]. 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 [21-25], 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, are 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 $\pm 3g$, 8 bits per axis resolution and 100Hz 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 ($N \times M$). 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 seconds, 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.

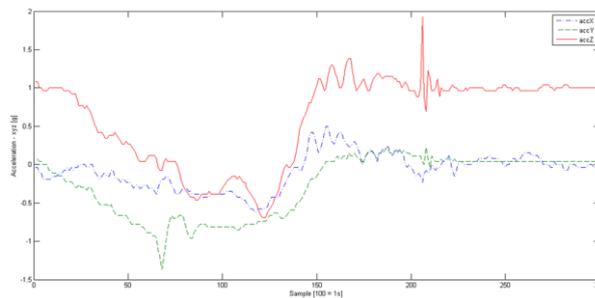


Fig. 1. Example of gesture “pick-up” acceleration footprint (processed in MatLab).

3.2 Gesture Based User Identification System

The procedure used in our gesture based user identification system (GBiD) is presented in Fig. 2. 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. The identification is successful if the selected gesture indeed belongs to the dataset of user X .

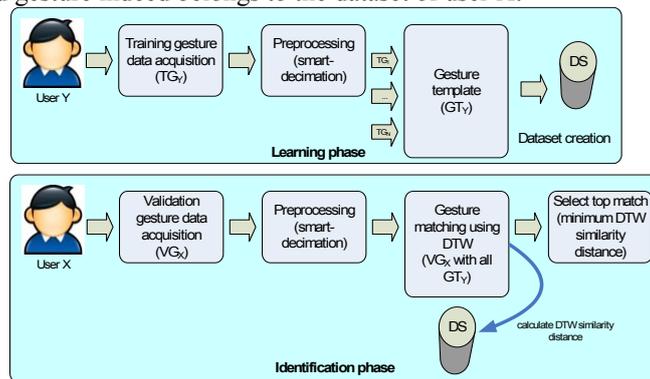


Fig. 2. Gesture based user identification system architecture.

All gestures (training and validation) are first pre-processed using our smart-decimation method in order to increase the recognition accuracy and speed. Our “smart-decimation” method processes the acceleration data in all three axes simultaneously and removes only those vector gesture samples that do not change significantly over time and thus carry less relevant data. Consequently, the gesture length is significantly reduced, which increases the processing performance.

At the core of our implicit gesture based user identification method is the Dynamic Time Warping method. It has inherent abilities to compare time series of different lengths and sample rates and is therefore well suited for this task. The minimum DTW distance is used as a similarity measure between the time series. Gestures acquired with accelerometer sensors can be described as multidimensional time series, which makes the proposed method well suited for our purposes, as is also evident from related work [21-25].

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 acceleration data ($accX$, $accY$, $accZ$) is combined using the Euclidean distance measure and used in the DTW calculation.

The top K -best matches with the lowest calculated DTW similarity distances between training and validation gestures are selected. The K factor determines the number of best matches that are further used in the decision process. In our current study only the best match was used in the decision process ($K=1$), although multiple matches could be used and thus even better results could be achieved.

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 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).

5 Results

User recognition success rate is expressed as a number of successful user recognitions vs. all recognition attempts and is presented in graphs in Fig. 3.

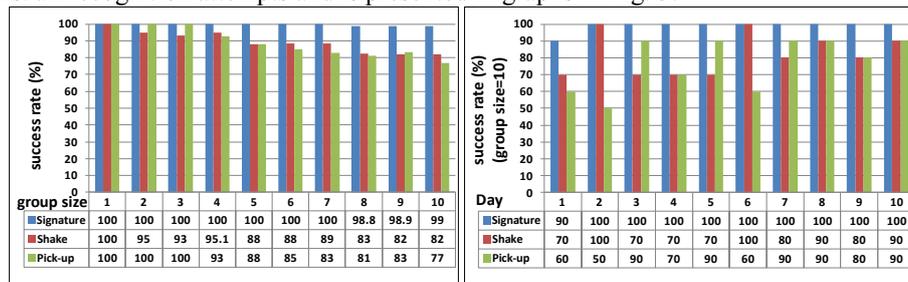


Fig. 3. User recognition success rate (in %) for various user target group sizes (left) and a detailed daily results for target group size of ten users (right).

The presented graph in Fig. 3 (left) illustrates the various gesture type detection methods in relationship to successful user recognitions as a function of various user group sizes (from 1 to 10). The results are cumulative, thus the percentage of success rates is given over a period of time of the full 10 test days for 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.

Fig. 3 (right) shows 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. 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.

To estimate how user's gesture execution changes 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. 3 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. 3 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. According to [26] 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).

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.

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.

On the other hand, the combination of gestures will undoubtedly increase the amount of required processing; therefore, additional data preprocessing procedures and especially efficient dataset combining methods are already being investigated.

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 its simplicity and its easy usage.

8 Acknowledgements

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. Montpetit, M.J., Mirlacher, T., Ketcham, M.: IPTV: An End to End Perspective. *Journal of communications*, vol. 5 no. 5 (2010)
2. Sedlar, U., Zebec, L., Bester, J., Kos, A.: Bringing click-to-dial functionality to IPTV users. *IEEE Communication Magazine*, 46 (2008), pp. 118–125
3. Volk, M., Sterle, J., Sedlar, U., Kos, A.: An approach to modeling and control of QoE in Next Generation Networks. *IEEE Communications Magazine*, 48 (2010), pp. 126-135
4. Jain, A.K.; Ross, A., Prabhakar, S.: An introduction to biometric recognition. *IEEE Transactions On Circuits And Systems For Video Technology*, 14 No. 1 (2004)
5. Nabti, M., Bouridane, A.: An effective and fast iris recognition system based on a combined multiscale feature extraction technique. *Pattern Recognition*, 41 (2008), pp. 868 – 879
6. Maltoni, D., Maio, D., Jain, A., Prabhakar, S.: *Handbook of Fingerprint Recognition*. Springer-Verlag, (2003)
7. Varchol P., Levický D.: Using of Hand Geometry in Biometric Security Systems. *Radioengineering*, Vol. 16, No. 4, December 2007
8. Abate, A.F., Nappi, M., Riccio, D., Sabatino, G.: 2D and 3D face recognition: A survey. *Pattern Recognition Letters*, 28 (2007), pp. 1885–1906

9. Toledano, D.T., Pozo, R.F., Trapote, A.H., Gómez, L.H.: Usability evaluation of multi-modal biometric verification systems. *Interacting with Computers*, 18 (2006), pp. 1101–1122
10. Kela, J., Korpipää, P., Mäntyjärvi, J., Kallio, S., Savino, G., Jozzo, L., D. Marca, D.: Accelerometer-based gesture control for a design environment. *Personal Ubiquitous Computing*, 10 (2006), pp. 285–299
11. Keogh, E., Pazzani, M.: Derivative Dynamic Time Warping. *Proc. of the First Intl. SIAM Intl. Conf. on Data Mining*, Chicago, Illinois, (2001)
12. Jain, A.K., Flynn, P., Ross, A.A.: *Handbook of Biometrics*. Springer, (2008).
13. Orozco, M., Graydon, M., Shirmohammadi, S., El Saddik, A.: Experiments in haptic-based authentication of humans. *Multimed Tools Appl* (2008) 37:73–92
14. Varona, J., Capó, A.J., González, J., Perales, F.J.: Toward natural interaction through visual recognition of body gestures in real-time. *Interacting with Computers*, 21 (2009), pp. 3–10
15. Cho, S.J., Oh, J.K., Bang, W.C., Chang, W., Choi, E., Jing, Y., Cho, J., Kim, D.Y.: 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*, (2004)
16. Jun-qi, K., Hui, W., Guang-quan, Z.: Gesture Recognition Model Based on 3D Accelerations. *Proceedings of 2009 4th International Conference on Computer Science & Education*, (2009)
17. Zappi, P., Milosevic, B., Farella, E., Benini, L., “Hidden Markov Model based gesture recognition on low-cost, low-power Tangible User Interfaces,” *Entertainment Computing*, 1 (2009), pp. 75–84
18. Schlömer, T., Poppinga, B., Henze, N., Boll, S.: Gesture Recognition with a Wii Controller. *Proceedings of TEI*, (2008), pp. 11–14
19. Kela, J., Korpipää, P., Mäntyjärvi, J., Kallio, S., Savino, G., Jozzo, L., Marca S.: Accelerometer-based gesture control for a design environment. *Pers Ubiquit Comput*, 10 (2006), pp. 285–299
20. Lugmayr, A., Dorsch, T., Humanes, P.R.: Emotional Ambient Media. In *Handbook of Research on Synthetic Emotions and Sociable Robotics: New Applications in Affective Computing and Artificial Intelligence*, IGI Global, (2009)
21. Zaharis, A., Martini, A., Kikiras, P., Stamoulis, G.: User authentication method in implementation using a three-axis accelerometer. *Mobile Lightweight Wireless Systems, Second International ICST Conference, MOBILIGHT*, Springer, (2010)
22. Okumura, F., Kubota, A., Hatori, Y., Matsuo, K., Hashimoto, M., Koike, A.: A Study on Biometric Authentication Based on Arm Sweep Action with Acceleration Sensor. *Proc. Int. Symp. Intelligent Signal Processing and Communications*, (2006)
23. Matsuo, K., Okumura, F., Hashimoto, M., Sakazawa, S., Hatori, Y.: Arm Swing Identification Method with Template update for Long Term Stability. *Proc. Int. Biometrics*, (2007)
24. Liu, J., Wang, Z., Zhong, L., Wickramasuriya, J., Vasudevan, V.: “uWave: Accelerometer-based Personalized Gesture Recognition and Its Applications”. *Proc. IEEE Int. Conf. Pervasive Computing and Communication (PerCom)*, (2009)
25. Liu, J., Zhong, L., Wickramasuriya, J., Vasudevan, V.: 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 (2009)
26. Eurostat Average household size,
http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=ilc_lvph01&lang=en