Creating a Database for Objective Comparison of Gesture Recognition Systems

Róbert VARGA[1], *Zoltán PREKOPCSÁK*[2]

[1, 2] Dept. of Telecommunications and Media Informatics, Budapest University of Technology and Economics, Magyar Tudósok körútja 2., 1117 Budapest, Hungary

[1] robertvarga87@gmail.com, [2] prekopcsak@tmit.bme.hu

Abstract. *In the past decade, accelerometer sensors became small enough to embed in mobile devices that people are using every day. Thus, arm movements became attractive to control mobile games and game consoles. However, there is no widely-used gesture-vocabulary for controlling devices and gesture recognition is not standardized. Furthermore, there is no widely-accepted benchmark for measuring performance of the developed gesture recognition methods, and it is impossible to decide objectively, which gesture recognition system is better than another. We present how to effectively build a benchmark database for gesture recognition systems, and we developed a data collection application which shortens the measurement process. We collect data for 23 gesture-words by Nintendo Wii Remote and HTC Touch Diamond devices. We recruited 30 people to participate in our data collection task and they produced the largest gesture recognition database to date.*

Keywords

Benchmark database, Accelerometer-based gesture recognition

1 Introduction

Gestures have recently become popular as natural interfaces for game consoles and mobile phones. These devices contain small accelerometer sensors, which has become an inexpensive tool in our days.

Arm movements are attractive to users to control their games or mobile applications so that is why some research aims to develop segmentation methods that can recognize gestures in real time, what makes the interface more natural for its users. The procedure can predict the beginning and the ending of the performed gesture from accelerometer data.

User-based recognition purposes to identify users from unique marks of the performed gesture. This procedure can be useful to multi-user environments for authentication tasks or the system can be associate a personalized gesture vocabulary for its user.

Gesture recognition tasks require a vocabulary, which contains valid gesture-words that we want to recognize. There is no standard vocabulary, so the users have to define the trajectory and the meaning of the gestures, which they want to use for the interface. Thus, the desired command can be executed as a result of the performed gesture. There is a need for a standard vocabulary to simplify the system development and to make the user interaction easier.

Developed gesture recognition methods require a data set that contains adequate number of gesture-samples from different persons. Until recently, the researchers have not published the collected gesture dataset that was used for training phase and performance evaluation for their research. In addition, gesture recognition lacks a large published database so the researchers have to create their own dataset. The data-collection procedure includes recruiting participants who are performing gestures, accelerometer data recording, storing labeled gestures, checking the correctness of the data samples. These tasks require precise work, which take a lot of time.

A large published database, which contains data-samples from different users helps to develop different type of methods for gesture recognition. This database can be used for benchmark, what allows to an objective performance comparison of the developed methods.

Our contribution is two fold. First, we provide a general gesture vocabulary which can be used for different tasks if we assign meaning for some of the gestures. Second, we provide a reference benchmark database, which can be used to objectively test different methods.

In this paper, we discuss how to effectively build a large benchmark gesture database that can be used for different accelerometer based tasks such as gesture recognition, user-based recognition and segmentation. In Section 2, we review the developed accelerometer based methods and their challenges. In Section 3, we present an application, which makes the data collection more effective, describe a gesture vocabulary and the data structure that was used for building the benchmark database. In Section 4, we discuss the circumstances of data collection and analyze the recorded data-samples. Finally, in Section 5, we present the collected dataset and we discuss our future works.

2 Related Work

Researchers published a lot of methods for accelerometer-based gesture recognition in the past few decades, such as Hidden Markov Model (HMM) that is widely-used in speech-recognition and gesture-recognition in accelerometer-based systems.

Researchers from VTT of Finland classified [1] the types of the gesture recognition systems from different aspects. The purpose of user-dependent system is to use the interface by one person. This condition is acceptable for mobile phones. User-independent interfaces apply in multi-user environments such as game consoles. In discrete gesture-recognition, the beginning and ending of gesture-words is defined by a button press. In continuous gesture-recognition, it is not necessary to sign the gesture, because the system automatically recognizes when it occurs.

Gesture-segmentation aims to recognize gesture-patterns from the accelerometer sensor data-stream so the system can continuously receive gesture-commands. Advantage of this method that is the user does not have to sign the beginning and ending of the arm movements manually. To solve this problem, Schlömmer applied threshold filters [2] to eliminate the acceleration values, which are not part of the occurred gesture. Prekopcsák analyzed [3] the collected data and determined rules to segment gestures automatically. These rules use the duration of the gestures and the changing of the acceleration values.

There is no widely-accepted evaluating procedure for segmentation methods so far. To solve this problem, auto-segmentation methods require a different data-sample, which contains the whole measurement from the performed gestures. This dataset also contains information about the beginning and ending of the gesture-words, so the user has to sign these events. This information can be used for the testing phase.

Most researchers have not published the developed methods and the collected gesture dataset, which they used for training phase and performance evaluation for their research. However, they have tried to compare the prediction results [4], [5] with other published results [1] by the described number of gesture-samples and performance results. This comparison is not objective, because the researchers used different data samples and accelerometer sensors.

Researchers from the Rice University [4] have published their own gesture database. This is the largest published data set so far, which contains 4480 gesture sample from 8 participants. Researchers from the University of Augsburg [5] have created a gesture-learning environment for Wii remote. This application is available online on the internet, which includes their recorded gesture sample and the developed methods.

Data collection can be a difficult task. Therefore some researchers [1] tried to shorten the training phase with duplicated gesture data. To duplicate the recorded gesture data, the researchers added Gaussian noise to the samples. This approach is not good for a reference database, because the tested procedure could learn the distortion, so it can acquire higher performance than the real prediction capability.

Up to this day, the researchers of gesture-recognition have not determined the objective performance-criteria. Thus, the published or unpublished system implementation can not compare by performance values. This problem was mentioned by other researchers of activity recognition [6]. The Pervasive 2010 Workshop papers agreed that, there is a need for a benchmark [7]. The CVPR workshop aims at gathering researchers to share algorithms and techniques of camera-based gesture recognition. The workshop also aims to collect the researcher's data sample and share with the community of video-based gesture-recognition.

Our work focuses to build a public benchmark accelerometer-based gesture database, which contains a large gesture-vocabulary. To use this database as a benchmark, we collected data continuously, which can be used for segmentation methods and we recorded gestures from different users and different accelerometer sensors.

3 Data recording application

Creating a reference database requires a large gesture vocabulary and a lot of participants, who are performing the arm movements repeatedly. Each participant has to perform hundreds of gesture because of the large gesture-dictionary and the repeated gesture-words. Therefore, data collection task can be an exhausting procedure, so we wanted to automate its processes. Thus, we can record adequate number of gestures in shorter time.

Technical parameters of accelerometer sensors can be various for each sensor and the researchers used different type of sensors for their studies. These parameters such as sensitivity, number of axis, quantization and sample rate can be important for the gesture-recognition systems. Therefore, containing accelerometer data from different mobile devices is an expectation for a reference database.

To improve the gesture recording tasks for creating a benchmark database, we developed an application for the data-collection. Our application, called MeasureGesture, has three main functions. First, it receives accelerometer data and controls signals from different mobile devices via Bluetooth interface. The second function is to inform the participants about the measurement tasks. This information is a short video, which describes the arm movement and pictogram of the gestures. The third is to store the collected accelerometer data in labeled-gesture form in a database. The reference database was created in Microsoft SQL Server environment. With these functions, the application allows to record gestures automatically. Thus, we can effectively collect hundreds of gestures from participants within a short time.

3.1 Gesture vocabulary

There is a lack of a standard gesture dictionary, so we gathered together the published vocabularies into a large gesture dictionary [1]. The created gesture vocabulary of 23 gesture words as is shown in Fig. 1. . The defined gestures are limited to only one plane. The author of [1] has already determined meanings of the gesture-words to control a VCR. Their gesture-vocabulary consists of 8 gestures (see Fig. 1. 1-8). We adopted 9 gesture-words from [8] (see Fig. 1. 9-17) and 6 gestures from [9] (see Fig. 1. 18-23).

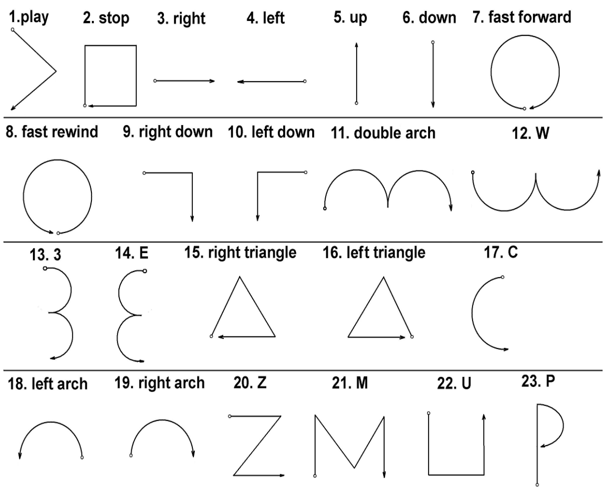

Fig. 1. 23 gesture-words of the gathered gesture vocabulary

3.2 Data structure

The designed schema allows storing data from 3-dimensional acceleration vectors, participants, mobile devices and gesture labels. The MeasureGesture stores the collected accelerometer values in raw form, so the researchers can execute arbitrary data transformation methods on the benchmark. The application also stores a logical timestamp for the sensor data. This information can be useful for data preparation methods such as checking sample rate, checking duration of the recorded gestures and the researchers can also use for their studies. The E-R diagram is shown (see Fig. 2).

Segment entity represents the recorded data and shows which gesture was performed by which participant and by which mobile device. Each segment determines a data-vector series, which are the measured accelerometer values during the performed arm movement. These data-vectors are ordered by time.

Data structure also allows to query the gestures without segmentation information, which can be useful for developing or testing segmentation methods. Measure entity is similar to Segment, but each Measure contains data from the whole measurement of 10 same gesture-words by one participant without segmentation information.

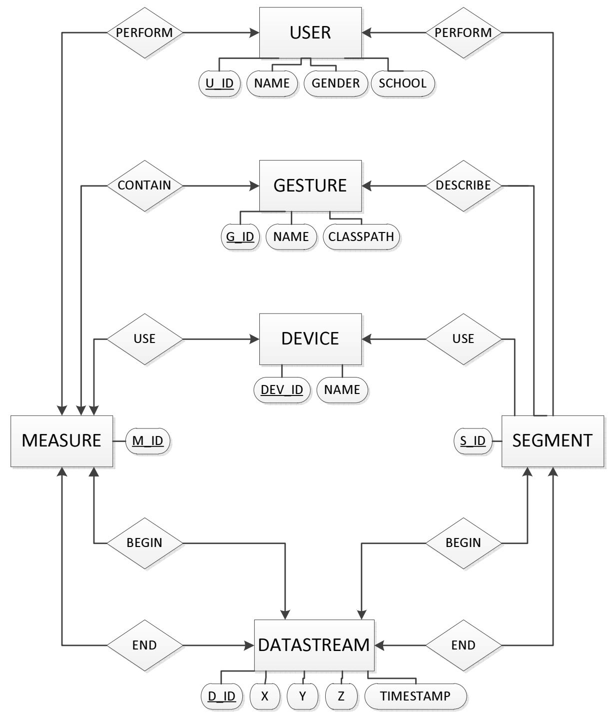

Fig. 2. E-R diagram of the benchmark database

4 Data Collection

Data recording requires a person, who supervises the measurements of the gestures. The supervisor's job is to show for the participants how to hold the mobile device in hand, to prepare the MeasureGesture for data recording and to control the developed application, when the participant makes mistake during the arm movements.

Process of the measurement (see Fig. 3) consists of iterative tasks such as to record each repeated gesture or to describe and record all gesture-words. These tasks are occurring frequently during the measurement so it is not practical to schedule manually.

To make the data acquisition more effective, MeasureGesture schedules these processes (on Fig 3. Change device, Describe gesture, Record gesture) automatically so it simplifies the supervisor's job. Thus, the application allows to finish the measurement in shorter time.

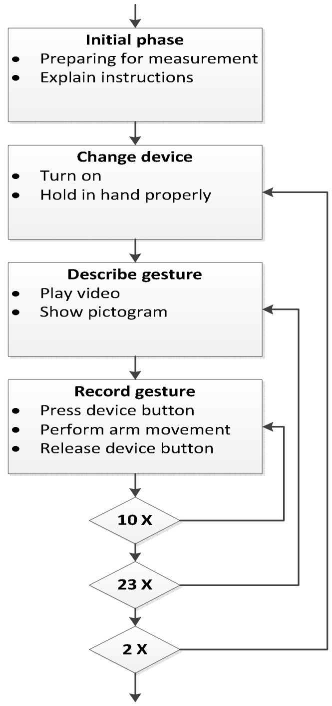

Fig. 3. Block diagram of the measurement. The described measurement consists of 10 repetitions of 23 gesture words by 2 mobile devices.

4.1 Circumstances of the data collection

The data recording employs Nintendo Wii Remote and HTC Touch Diamond mobile phone as 3 dimensional accelerometer sensors for collecting gesture data. The Wii remote and HTC sensor worked with 100Hz and 25Hz sample rate and 3g and 4g sensitivity.

The participants were standing during the measurement and each participant held the device in the same way. Participants had to hold a button on the device while performing a gesture. There were no constraints on the amplitudes and dynamics of the arm movements so the participants performed the gestures comfortably. Each participant repeated 10 times the 23 gesture-words with 2 mobile devices. The participants understood easily the task of gesture performing by the short videos.

4.2 Participants

We collected gesture data from 30 people to create the reference database. They are 15 females and 15 males and aged 12 to 57 (see Fig. 4). Most participants are undergraduate students.

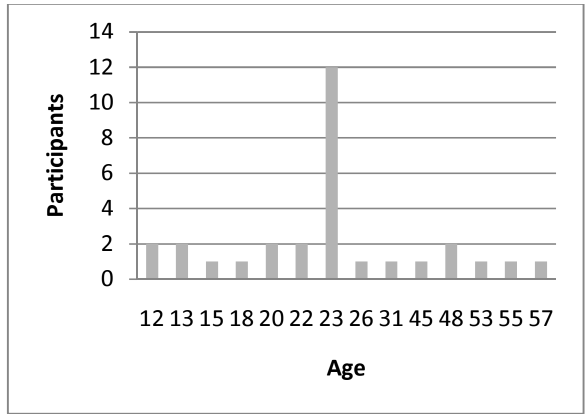

Fig. 4. Age of the participated people

4.3 Statistics of collected data

The collected benchmark contains 13800 gesture samples, 460 from each participant. Each gesture-word is represented with 600 samples in the benchmark.

We analyzed the duration of the measured gestures. It took 28 sec (σ = 9.789) on average (see Fig. 5.) to measure a gesture-word with 10 repetition.

The recorded gesture-words of the vocabulary (see Fig. 6.) took 1.3 sec (σ = 0.513) on average. The measurement of 460 gestures took 25 minutes on average with 2 devices. The author of [2] recorded 75 gestures in 15 minutes on average so our developed application really shortened the data collection task.

To compare our collected database with other works, we summarized the parameters of the collected data sets (see Tab. 1.) , which was used the researchers for their studies. It can be seen on Tab. 1. that we applied the largest gesture-vocabulary and we collected the largest gesture sample by 30 participants.

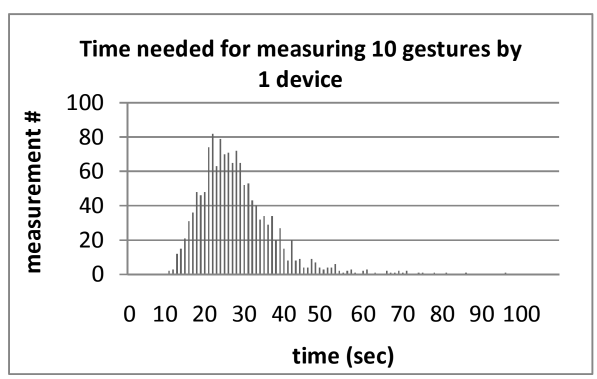

Fig. 5. Histogram of the duration of 10 repeated arm movements

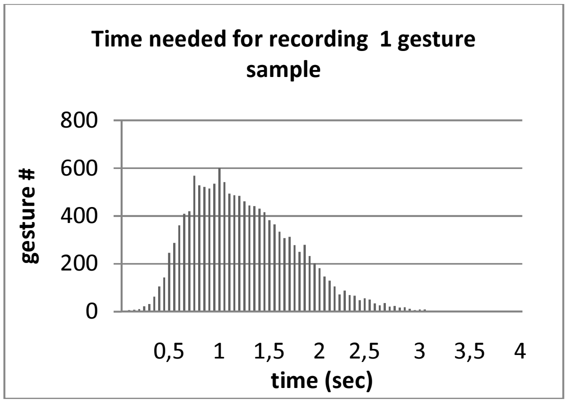

Fig. 6. Histogram of the duration of the gesture-words.

Researchers	Gesture #	Participants	repetition	sample
[10]	8	1	20	160
[1]	8	1	30	240
[3]	10	4	10	400
[9]	10	4	10	400
[2]	5	6	15	450
[5]	8 10 3	1 7 7	30 10 60	240 700 1260
[11]	7	5	30	1050
[12]	10	7	20	1400
[8]	18	7	30	3780
[4]	8	8	70	4480
In this paper	**23**	**30**	**20**	**13800**

Tab. 1. Parameters of data samples in gesture recognition studies.

5 Conclusion and future work

In this paper, we have presented the creation of a gesture recognition dataset. We have shortened the data collection with the MeasureGesture application, and recorded 13800 gesture samples from 30 people.

The gesture-vocabulary contains the 8 fundamental gesture-words [1], which the researchers widely-used to compare performance with other studies and 15 other gestures. This gesture-vocabulary is large enough for the researchers to use for their studies as a comprehensive gesture-dictionary.

We have recorded the gesture data with different accelerometer sensors, so the researchers can examine the device-independence of their developed system.

The created reference database can be used for gesture recognition, user recognition and segmentation benchmarking systems.

We will share the database with the community soon and we will determine performance-criteria for benchmarking gesture recognition systems.

Acknowledgements

We would like to thank all of 30 people, who participated in the data collection for their help and their useful advices. We would like to thank Márton Tóth for useful advices for the application development.

References

[1] KELA J., KORPIPÄÄ P., MÄNTYJÄRVI J., KALLIO S., SAVINO G., JOZZO L., MARCA D., Accelerometer-based gesture control for a design environment, *Personal Ubiquitous Computing* , 2006, vol. 10, no. 5, pp. 285 – 299.

[2] SCHLÖMER T., POPPINGA B., HENZE N., BOLL S., Gesture recognition with a wii controller, *in TEI '08: Proceedings of the 2nd international conference on Tangible and embedded interaction*, New York, USA, 2008, pp. 11 – 14.

[3] PREKOPCSÁK Z., Accelerometer based real-time gesture recognition, *in Proceedings of the 12th International Student Conference on Electrical Engineering*, 2008.

[4] LIU J., ZHONG L., WICKRAMASURIYA J., VASUDEVAN V., uwave: Accelerometer based personalized gesture recognition and its applications, *in Pervasive and Mobile Computing*, 2009, vol. 5, no. 6, pp. 657 – 675.

[5] REHM M., BEE N., ANDRÉ E., Wave like an Egyptian: accelerometer based gesture recognition for culture specific interactions, *in BCS-HCI '08: Proceedings of the 22nd British HCI Group Annual Conference on People and Computers*, Swinton, UK, 2008, pp. 13 – 22.

[6] AMFT O., On the need for quality standards in activity recognition using ubiquitous sensors, *in How To Do Good Research In Activity Recognition Workshop in conjunction with Pervasive 2010*, 2010.

[7] BRUSH A.J., SCOTT J., KRUMM J., Activity Recognition Research: The Good, the Bad, and the Future, *in Pervasive 2010 Workshop*, 2010.

[8] AKL A., VALAEE S., Accelerometer-based gesture recognition via dynamic time warping, a-ffinity propagation, & compressive sensing, *in Acoustics Speech and Signal Processing (ICASSP). IEEE International Conferenc*e, 2010, pp. 2270 – 2273.

[9] JOSELLI M., CLUA E., gRmobile: A Framework for Touch and Accelerometer Gesture Recognition for Mobile Games, *in 2009 VIII Brazilian Symposium on Games and Digital Entertainment*, 2009, pp.141 – 150.

[10] BAILADOR G., ROGGEN D., TRÖSTER G., TRIVNO G., Real time gesture recognition using continuous time recurrent neural networks, *in BodyNets '07: Proceedings of the ICST 2nd international conference on Body area networks*, Brussels, Belgium, 2007, pp. 1 – 8.

[11] VARCHOLIK P. D., MERLO J. L., Gestural communication with accelerometer-based input devices and tactile displays, *in Proceedings of the 26th Army Science Conference*, 2008, Orlando, FL, USA.

[12] PYLVÄNÄINEN T., Accelerometer based gesture recognition using continuous hmms, *Lecture Notes in Computer Science*, 2005, vol. 3522/2005, pp. 639 – 646.

About Authors...

Róbert VARGA was born in Budapest Hungary. He is a computer engineer BSc student at Budapest University of Technology and Economics. He is interested in data mining and accelerometer-based application development.

Zoltán PREKOPCSÁK is a PhD student at Budapest University of Technology and Economics. His research topic is pattern classification in time series with applications in human-computer interaction and ubiquitous computing.