Gesture Ontology for Informing Service-oriented Architecture

Catalin-Marian CHERA, Wei-Tek TSAI, Radu-Daniel VATAVU

Abstract— We present a system engineering approach for designing Service-oriented Architectures (SOA) for software applications that use gesture commands. The approach employs ontology for gesture-based interaction which was designed on three levels: user execution, system implementation, and gesture reflection. The ontology borrows concepts from several research communities interested in gestures such as human-computer interaction, pattern recognition, and cognitive psychology. We show how the ontology can be used in order to inform the design of Service-oriented Architectures for engineering new systems and applications and describe a software architecture design for controlling smart homes with gesture commands.

I. INTRODUCTION

Current trends in interacting with information systems employ gestures and body movements as natural means for users to enter commands [2, 15, 21]. As advances in technology empowered developers to implement gestural interfaces for new gadgets and devices, it is mandatory for the community to possess a clear understanding of what a gesture command is, what are its attributes, and what relationships can be drawn between the various concepts involved in gesture execution, recognition, and interaction. This type of knowledge is needed not only in computer science and engineering but also in other communities that show interest in gestures such as psychology and social sciences. For example, Cuffari [5] explored different views on gestures by comparing ideas from cognitive psychology in opposition to treatments of gesture from existentialphenomenology.

One way to understand the concepts, attributes, and relationships of a study domain is to use ontologies [8]. Ontology modeling can benefit researchers and practitioners by defining common vocabulary, meanings, attributes, and relationships together with hierarchical structures of the employed concepts. This representation technique is heavily used for developing the Semantic Web while practitioners of Service-oriented Computing (SOC) use ontologies as key technology. As gesture-based interfaces will become more present on the web (e.g., mouse gestures are already used by web browsers [17]), SOC applications that employ dedicated

*http://www.gestureontology.fcint.ro is the companion web site for this paper that contains visual illustrations and representations of the gesture ontology.

Catalin-Marian CHERA is with the Computer Science Department, University Politehnica of Bucharest, 060042, Romania (e-mail: catalin.chera@cs.pub.ro).

Wei-Tek TSAI, is with the School of Computing, Informatics and Decision Systems Engineering, Arizona State University, AZ85287, USA. He is also with the Department of Computer Science and Technology, Tsinghua University, Beijing, China (e-mail: wei-tek.tsai@asu.edu).

Radu-Daniel Vatavu is with the Computer Science Department, University Stefan cel Mare of Suceava, Romania (e-mail: vatavu@eed.usv.ro).

web services for gesture recognition and processing are likely to be developed. In this context, a common understanding of the concepts, hierarchies, and relationships relating to gestures will be critical for their integration into the semantic web.

Although gesture ontologies have been designed and used before, they were mostly developed to suit the needs of a particular application. It is therefore necessary to propose more comprehensive gesture ontology so that researchers from different communities can benefit of an understanding of gesture concepts and relationships. To achieve this goal, this paper seeks a new ontology for gestures that builds on the experience and results of multiple communities. Specifically, the ontology includes three dimensions: gesture execution, gesture implementation, and gesture reflection that provide practitioners with a reasonably wide view on gestures and their applications.

In the process of designing the gesture ontology, three key communities were identified: (1) human-computer interaction practitioners interested in developing user interfaces that employ gestures; (2) the pattern recognition community that develops and utilizes techniques for representing, recognizing, and interpreting gestures; and (3) the psychology community that studies the roles played by gestures and how gestures relate to the human discourse. Furthermore, this paper also presents a service-oriented architecture as informed by the ontology for controlling devices of a smart home via gesture commands. We provide full details the ontology design http://gestureontology.fcint.ro*.

II. RELATED WORK

Gesture ontologies have been proposed and used before in the literature of interface design. For example, Garcia-Rojas et al. [6] proposed an ontology for animating emotions for virtual humans. The ontology allows retrieval of animations that have been previously annotated with emotional descriptors. Wang et al. [25] described an ontology for multi-touch by considering atomic gestures and ways to combine them with temporal, spatial, and logical relationships. Sonntag et al. [22] defined ontological representations for pointing gestures in the SmartWeb project. Gesture strokes performed using the stylus are described in terms of time and x/y position together with references to a tagged object in the pointed image. Seghbroeck et al. [21] introduced WS-Gesture which represents a framework for controlling devices based on DPWS (Devices Profile for Web Services). DPWS relies on WS-* protocols initially designed for enterprise computing and then ported to the device networks that may have hardware constraints not present in an enterprise system.

Gesture ontologies were also discussed in other communities such as psychology and philosophy. For example, the goal in [5] is to provide reconciliation on the understanding of embodied meaning in the context in which psychology sees gestures and speech different while in philosophy speech is already gesture and co-occurrence of expression and thought leads to gesture.

explicitly developing ontologies, Although not researchers have classified gestures by their common properties. For example, Wobbrock et al. [28] introduced a taxonomy of surface gestures by taking into consideration multiple criteria such as gesture form, nature, binding, and flow. The nature dimension classifies gestures in accordance to their abstract or physical meanings with four classes: symbolic, physical, metaphorical, and abstract. The binding dimension classifies gestures into object-centric, worlddependent, world-independent, and mixed dependencies. Finally, flow specifies whether the action occurs after the gesture has been executed (discrete) or during execution (continuous). Ruiz et al. [20] introduced a taxonomy of motion gestures for mobile phones by considering gesture mappings (nature, context, temporal) and physical characteristics (kinematic impulse, dimension, complexity). The nature, context and temporal dimensions are similar to Wobbrock et al.'s nature, binding, and flow criteria [28]. According to kinematic impulse, gestures can have low, medium, and high jerk. When classifying by dimension, gestures can be executed on one axis, 3-axis, or 6-axis. Also, judging by complexity, gestures can either be simple or compound. Vatavu and Pentiuc [24] proposed a simple classification of hand gestures into four types: static simple, static generalized, dynamic simple, and dynamic generalized by considering the amount of hand posture and motion present in the commands.

III. GESTURE ONTOLOGY

We designed the gesture ontology on three dimensions that group aspects of a gesture-based control application: execution, implementation, and reflection (see Fig. 1).

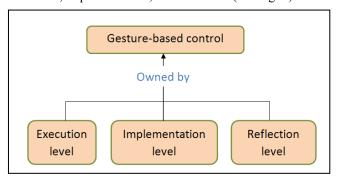


Figure 1. The three dimensions of the gesture ontology.

The Execution level refers to aspects involved during the process of producing a gesture: users, body movements, and acquisition devices. Implementation includes gesture representation and recognition. The Reflection level models the way gestures are used as commands or companions to speech as well as their usage in the linguistics and

psychology communities. The proposed ontology aims to unify concepts and relationships from multiple communities interested in the various aspects of gesture which makes it stand as the most comprehensive gesture ontology available today.

A. Gesture Execution Level¹

Execution level considers gesture production aspects and includes the following concepts:

- Users: Who is performing the gesture? What are the relevant users' attributes for gesture execution? How do users transition from novices to experts?
- **Body Movements:** Which body parts are involved in producing the gesture?
- **Acquisition Devices**: How are gestures being captured? In which environments?
- **Execution Enhancers:** What type of feedbacks do users receive while executing gestures?

All these concepts relate to the process of gesture execution including production, acquisition, and the feedback users receive while executing gestures. Users represent the main concept of the Gesture Execution level. From the interface perspective, users can be either novices or experienced. This distinction is critical for gesture-based interfaces due to the nature of gesture as command: gestures act as shortcuts that are supposed to be fast comparing to the traditional use of menus, mouse, and keyboard. However, in order for users to become experts, a certain amount of practice is required and researchers have proposed several techniques for assisting this process [2]. A natural relationship between the two concepts is the one that describes the transition from novices to experienced: Novices become Experienced with training. This transition can be achieved with practice or by using one of the available assisting techniques. Users perform Body Movements that can be further classified by the body part involved during the motion such as arms and legs [15, 18], hands [13], fingers [27], and head [4]. The simplified ontology view from Figure 2 illustrates instances of such gestures. For example, sit down, jump or lean represent whole body gestures while pinch and open are gestures performed with the hand.

The process in which *Users* perform *Body Movements* is referred as *Acquisition*. Movements are captured using acquisition *Devices* inside specific *Environments*. There are many devices available today for acquiring gesture movements that differ in terms of resolution, data reporting frequency, data format, and device cost. Also, there are works that attempt a classification of such devices using criteria such as the type of events they generate or the main type of feedback they provide [23]. For simplicity, *Devices* can be either *Wearable* or *Environmental*. *Wearable* devices need to be held or worn by *Users* such as game controllers [29], body suits [30], data gloves [31], and mobile devices

¹ See a visual illustration of the Execution level at http://www.gestureontology.fcint.ro/images/pdf/gestureexecutionlevel.pdf (not inserted in the text due to limited space)

[1]. Environmental Devices are installed in Environments and watch or sense Users' actions. They include video cameras [15, 21] or interactive displays in the form of tabletops and vertical surfaces [16]. Also, motion sensing can be employed to trigger an action when the simple presence of the user is detected.

Several relationships can be drawn between our concepts so far. For example, *Users* wear *Wearable Devices* while *Users* are being watched by *Environmental Devices*. Also, *Environmental Devices* are installed inside *Environments*. The *Environments* class sets the location in which gestures are being executed. Subclasses are represented by *Home*, *Office* or *Public* settings. A relationship that can be drawn here is that *Environments* influence *Users* in their decision to embark interaction with a gesture-based interface.

The process of executing gestures can be enhanced for several reasons. One is to provide feedback that gesture acquisition is running or that gestures are being captured correctly. Such feedback can be Visual, Audio, or Haptic, concepts located in the ontology under the general term Feedback Providers. A good example including all three feedback types is the tennis game of the Nintendo Wii Console Sports Pack [33]: when the user hits the ball by performing a swing gesture with the Wii Remote controller, a ball hitting sound is being played while the remote rumbles to simulate contact with the ball. At the same time, the display shows the ball changing direction after being hit. Another type of enhancer is represented by a Training Tool that assists users during the process of learning gesture commands. Feedback Providers together with Training Tools make up the Execution Enhancers class. At relationships level, Execution Enhancers accompany Body Movements; Feedback Providers provide feedback to Users; and Training Tools assist Novice Users while learning gestures helping them become Experienced.

B. Gesture Implementation Level²

The Implementation level contains concepts employed in the process of developing gesture recognition systems such as:

- Representations: How are gestures represented for recognition? How are they represented for humans?
- **Recognition**: What are the recognition algorithms for classifying gesture motions? What are the main tools and techniques? How is performance measured?

Gestures may use different *Representations* at different levels. This paper uses a two-level view of *Representations* by considering machines and users. *Machine Readable Representations* can be *Features* as they are heavily used in statistical pattern recognition [26], *Primitives* as used in structural pattern recognition such as curves, lines and corners [7], and *Encodings* that have been proposed for storing gestures. Examples of such encodings are the

UNIPEN format [9] initially designed to store pen strokes for pen-gesture interfaces and the InkML format [10]. The InkML (Ink Markup Language) is being developed by W3C and serves as data format for ink entered with electronic pens or styluses with specific XML tags. Representations include Human Readable Representations of gestures such as Videos, and Descriptions. These serve to communicate gestures to users in forms of manuals and training sessions. For example, a manual accompanying a gesture-based interface game would contain a series of images describing how to execute gestures. The Nintendo Wii Sports Pack includes video tutorials showing users how to perform gestures such as swinging a golf club or throwing a bowling ball. Descriptions serve to accompany other representations by providing additional details, such as "hold the device firmly during execution", "use moderate speed", etc.

To build the concept hierarchies for the recognition part of this level, this paper relies heavily on the results of the pattern recognition community [11, 26] for which the main concepts are Approaches, Tools, Performance, and Training Samples. The main approaches in pattern recognition [11] are Statistical, Syntactic, Template Matching, and Neural *Networks*. Each approach uses a specific tool. For example, the Statistical approach uses Discriminant Functions, the Syntactic approach uses Grammars, while Template matching uses Metrics. A couple of metrics have been found popular for gesture recognition and they include the Euclidean Distance, or the Dynamic Time Warping. Neural Networks have also been used for gesture recognition [27]. The *Performance* of gesture recognizers is being assessed in terms of their Complexity and Error Rate. The complexity refers to both *Processing* and *Memory* and it usually relates to how many samples are available in the Training Set. Analysis of complexity uses notions from algorithmic complexity. For example, the Nearest Neighbour rule has a linear complexity function of the number of samples in the training set. With regards to the Error Rate, a couple of measures have been proposed in the literature [11] such as classification error, acceptance error, or mean square error.

C. Gesture Reflection Level³

The Reflection level models gestures as commands or part of speech and employs results from speech and discourse [3, 12, 14]. The main concepts are:

- Class: the identifier of a gesture as it is being delivered by a recognizer and further employed by the system.
- Commands: What commands are associated to gestures? How do they relate to real world actions?
- **Attributes:** What do we see when we look at gestures?
- **Function-oriented Gestures**: How are gestures classified by their function?

² See a visual illustration of the Implementation level at http://www.gestureontology.fcint.ro/images/pdf/gestureimplementationlevel.pdf

³ See a visual illustration of the Reflection level at http://www.gestureontology.fcint.ro/images/pdf/gesturereflectionlevel.pdf

• **Linguistic Gestures**: How are gestures classified by their relationship to human discourse?

In accordance to task type, commands can be *Natural* or *Abstract*. Natural commands replicate the same actions from the real world such as pointing, rotating and manipulating objects, answering yes and no with head nods. On the other hand, no natural gestures exist for cut & paste or undo tasks (abstract commands).

Gestures posses *Attributes* which can be *Objective* or *Subjective*. Objective attributes refer to the way people perceive the geometry or the execution of a gesture movement. For example, gesture trajectories can be round-shaped, large or small, are performed clockwise or counter clockwise, etc. Subjective attributes may vary from one user to the next as they describe the ease of execution, familiarity, learnability, or the intuitiveness of a gesture command. Such attributes are critical when designing gesture sets for a given application and have been investigated by researchers before [19, 28].

A first classification of gestures by their function was provided by Cadoz [3] that identified 3 main classes: Ergotic, Semiotic, and Epistemic gestures. Semiotic gestures have meaning and are used to convey information. McNeill [14] further classifies semiotic gestures into *Deictic*, *Iconic*, Metaphoric, and Beat-like. Deictic gestures are used to point and they have been naturally reused as commands for gestural interfaces. A natural relationship can be drawn between Deictic gestures and Natural commands in the ontology. Iconics are gestures that describe a physical object or a real situation, usually accompanying speech. For example, iconic gestures describing size appear naturally during speech in sentences such as "I caught a fish this big" or "I was this tall at that time", etc. Metaphoric gestures are similar to Iconics except they describe an abstract concept. An example would be pointing ahead to suggest something in the future. Iconics gestures can be associated to Natural commands but also to Abstract ones while Metaphorics can only be associated to Abstract commands. Beat-like gestures are simple and rapid movements used to augment the discourse. Other cognitive psychologists have classified gestures in direct relationship with speech. For example, Kendon [12] describes a continuum ranging from Gesticulation up to Sign Languages. Middle levels include Language-like gestures which are similar to the iconic that naturally appear in speech and Emblems which are cultural specific such as the ok sign or various Italian hand gestures.

D. Inter-level relationships

New relationships can be derived by looking at the Execution and Implementation levels. The execution level shows how *Users* perform *Body Movements* which are captured by *Acquisition Devices* inside *Environments*. *Acquisition Devices* use *Machine Readable Representations* to provide the output data. *Devices* are used to collect *Training Samples*. Device characteristics (e.g., resolution) impact system performance so there is a relationship between *Performance* and *Devices*. Users are presented with *Human Readable Representations* and *Novice Users* especially make use of them during training and practicing. *Execution*

Enhancers use *Tools* to provide real-time feedback and assist users learn gestures.

When analyzing the Execution and Reflection levels new relationships can also be derived. For example, *Users* can rate the *Subjective* attributes of gestures and describe gestures using their *Objective* attributes such as shape or scale and rate *Commands* with respect to their naturalness. *Deictic* gestures are performed using *Hands*, *Arms*, and *Fingers* as are most of the *Semiotic* gestures. *Ergotic* gestures that manipulate the environment and *Epistemic* movements that sense the environment are performed using *Hands* and *Fingers*. Both *Function-oriented* and *Linguistic Gestures* use *Representations*.

IV. AN EXAMPLE OF SOFTWARE SERVICE IDENTIFICATION FROM ONTOLOGY

We discuss in the following how the proposed ontology can be employed in order to inform the design of services for applications implementing Service-oriented Architectures. A practical example is presented in the context of a smart home scenario in which users control equipments and devices with gestures acquired with the Wii Remote controller.

At the most simple level, a gestural interface needs a recognizer working with a data set of training samples. The following basic services were therefore identified in relation to the concepts of the ontology (see Fig. 2):

- WII-GESTURE-ACQUISITION-SERVICE: implements
 details which are specific for the acquisition device
 such as communication protocols. The output is
 represented by a motion gesture that can be either
 stored as part of a training set or fed into a
 recognizer for classification. This service is part of
 the Gesture Execution Level and implements an
 instance of the Controller concept.
- GESTURE-REPRESENTATION-SERVICE implements the encoding of a motion gesture using the XML format. The service implements an instance of the *Machine Readable Representation* concept located under the Gesture Implementation Level.
- GESTURE-SET-MANAGEMENT-SERVICE implements gesture storage operations such as uploading and retrieving gesture samples to/from a database server. It implements an instance of the *Training Samples* concept under the Gesture Implementation Level. This service uses the *XML Gesture Representation Service* to store and retrieve gesture sets.
- NEAREST-NEIGHBOR-RECOGNITION-SERVICE implements a *Statistical* classifier which represents a pattern recognition *Approach*. This is the algorithm that classifies a candidate gesture resulted from the acquisition process.
- DYNAMIC-TIME-WARPING-SERVICE implements a distance for computing the dissimilarity between two motion gestures. It is used by the NEAREST-NEIGHBOR-RECOGNITION-SERVICE at the Implementation level.

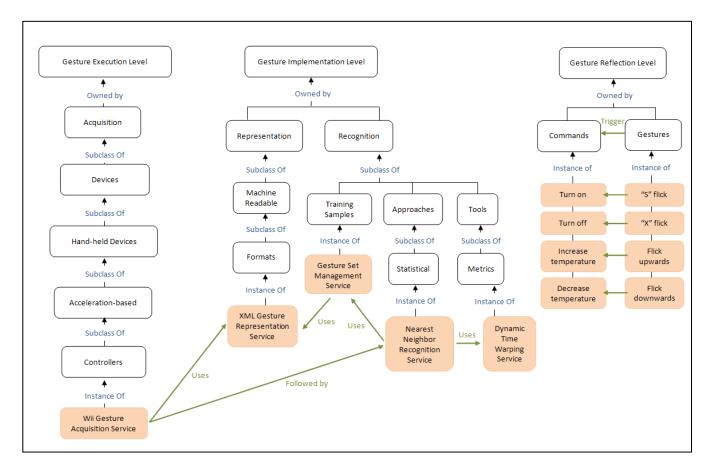


Figure 2. Gesture processing services and commands derived from the gesture ontology for a smart home application. Only a simplified view of the ontology is shown here. For a detailed view see http://www.gestureontology.fcint.ro/

The Gesture Reflection Level serves to define gestures and associate gestures to commands. For example, the following commands were available for controlling an air conditioner device: turn on, turn off, and increase/decrease temperature by 1 degree. Designed gestures included an S flick to start or turn on the air conditioner, X to turn it off and quick upward and downward motions for controlling the temperature.

V. CONCLUSION

We introduced in this paper gesture ontology for informing the design of gesture controlled applications. Unlike previous attempts to model gestural interfaces using ontologies, we built on the knowledge and research results from three independent communities: human-computer interaction, pattern recognition, and cognitive psychology. The result is a comprehensive gesture ontology covering aspects from user execution, system implementation, and gesture reflection. We demonstrated the ontology informing the design of Service-oriented Architectures for software applications employing gesture commands.

ACKNOWLEDGMENT

This work is supported partially by European Regional

Development Fund and the Government of Romania under the grant no. 181 of 18.06.2010, U.S. National Science Foundation project DUE 0942453 and the European project: Empowering Romanian Research on Intelligent Information Technologies (ERRIC) FP7-REGPOT-2010-1/264207.

REFERENCES

- [1] Apple iPhone, http://www.apple.com/iphone/
- [2] Bragdon, A., Zeleznik, R., Williamson, B., Miller, T., La Viola, J.J., Jr. (2009) GestureBar: improving the approachability of gesture-based interfaces. In Proceedings of the 27th international conference on Human factors in computing systems (CHI '09). ACM, New York, NY, USA, 2269-2278. DOI=10.1145/1518701.1519050
- [3] Cadoz, C. (1994) Les realites virtuelles. Dominos, Flammarion
- [4] Choi, H.-I., Rhee, P.-K, Head gesture recognition using HMMs, Expert Systems with Applications, 17 (3), 213-221.
- [5] Cuffari, E. (2008), Towards a Like-Minded Ontology: Gesture and the Embodied Nature of Thought in David Mcneill and Maurice Merleau-Ponty, 9th Conf. on Conceptual Structure, Discourse & Language.
- [6] Garcia-Rojas, A., Vexo, F., Thalmann, D., Raouzaiou, A., Karpouzis, K., Kollias, S. (2006) Emotional Body Expression Parameters In Virtual Human Ontology, 1st Int. Workshop on Shapes and Semantics, 63-70.
- [7] Gonzalez, R.C., Thomason, M.G. (1978) Syntactic Pattern Recognition: An Introduction, Addison-Wesley.
- [8] Gruber, T.R. (1993) A Translation Approach to Portable Ontology Specifications. Knowl. Acquis. 5, 2 (June 1993), 199-220. DOI=10.1006/knac.1993.1008.
- [9] Int. Unipen Foundation, http://www.unipen.org/

- [10] Ink Markup Language, http://www.w3.org/TR/InkML/.
- [11] Jain, A.K., Duin, R.P.W., Mao, J. (2000) Statistical pattern recognition: a review. IEEE Trans. Pattern Anal. Mach. Intell. 22, 1 (January 2000), 4-37. DOI=10.1109/34.824819.
- [12] Kendon, A. (1986) Current issues in the study of gesture. In The Biological Foundation of Gestures: Motor and semiotic Aspects, Lawrence Erlbaum Assoc., 23-47.
- [13] Malik, S., Laszlo, J. (2004) Visual touchpad: a two-handed gestural input device. Proceedings of the 6th international conference on Multimodal interfaces (ICMI '04). ACM, New York, NY, USA, 289-296. DOI=10.1145/1027933.1027980.
- [14] McNeill, D. (1992) Hand and mind: What gestures reveal about thought, University of Chicago Press.
- [15] Microsoft Kinect, http://www.xbox.com/en-GB/kinect.
- [16] Microsoft Surface, http://www.microsoft.com/surface/
- [17] Mouse gestures in Opera, http://www.opera.com/browser/ tutorials/gestures/
- [18] Nintendo WiiFit, http://wiifit.com/
- [19] Nielsen, M., Störring, M., Moeslund, T.B. and Granum, E. (2004) A procedure for developing intuitive and ergonomic gesture interfaces for HCI. LNCS 2915, Springer, 409-420
- [20] Ruiz, J., Li, Y., Lank, E. (2011) User-defined motion gestures for mobile interaction. In Proceedings of the 2011 annual conference on Human factors in computing systems (CHI '11). ACM, New York, NY, USA, 197-206. DOI=10.1145/1978942.1978971
- [21] Seghbroeck, G. van, Verstichel, S., Turck, F. de, Dhoedt, B. (2010) WS-Gesture, a gesture-based state-aware control framework. Proceedings of the IEEE International Conference on Service-Oriented Computing and Applications (SOCA'10), 1-8. DOI= 10.1109/SOCA.2010. 5707162
- [22] Sonntag, D., Engel, R., Herzog, G., Pfalzgraf, A., Pfleger, N., Romanelli, M., Reithinger, N. (2007) SmartWeb handheld: multimodal interaction with ontological knowledge bases and semantic web services. In Proceedings of the ICMI 2006 and IJCAI 2007 international conference on Artifical intelligence for human computing (ICMI'06/IJCAI'07), Thomas S. Huang, Anton Nijholt, Maja Pantic, and Alex Pentland (Eds.). Springer-Verlag, Berlin, Heidelberg, 272-295.
- [23] Vatavu, R.D., Pentiuc, S.G., Chaillou, C. (2005) On Natural Gestures for Interacting in Virtual Environments. Advances in Electrical and Computer Engineering, 5(12), 72-79
- [24] Vatavu, R.D., Pentiuc, S.G. (2008) Multi-Level Representation of Gesture as Command for Human-Computer Interaction. Comp. and Inf. 27(6), 837-851
- [25] Wang, De-xin, Xiong, Zhi-hui, Zhang, Mao-jun (2011) An application oriented and shape feature based multi-touch gesture description and recognition method. Multimedia Tools and Applications, Springer
- [26] Webb, A. (2003) Statistical Pattern Recognition. Wiley
- [27] Weissmann, J., Salomon, R. (1999) Gesture recognition for virtual reality applications using data gloves and neural networks, Neural Networks, pp. 2043–2046
- [28] Wobbrock, J.O., Morris, M.R., Wilson, A.D. (2009) User-defined gestures for surface computing. In Proceedings of the 27th international conference on Human factors in computing systems (CHI '09). ACM, New York, NY, USA, 1083-1092. DOI=10.1145/1518701.1518866
- [29] Nintendo Wii controllers, http://www.nintendo.com/wii/ console/controllers
- [30] Xsens, 3D Motion Tracking, http://www.xsens.com/
- [31] 5DT Data Gloves, http://5dt.com/
- [32] PlaysStation2 EyeToy camera http://us.playstation.com/ps2/ accessories/eyetoy-usb-camera-ps2.html
- [33] Wii Sports, http://www.nintendo.com/games/detail/1OTt006 SP7M52gi5m8pD6CnahbW8CzxE.