A Virtual Trainer by Natural User Interface for Cognitive Rehabilitation in Dementia

Alessandro Leone, Andrea Caroppo, and Pietro Siciliano

CNR-IMM, via Monteroni presso Campus Universitario – Lecce, Italy
{alessandro.leone,andrea.caroppo,
pietro.siciliano}@le.imm.cnr.it

Abstract. The aim of this work is the design and the development of an ICT platform integrating advanced Natural User Interface technologies for multi-domain Cognitive Rehabilitation without the direct physician involvement to the rehabilitation session. The platform is made up of a set-top-box connected to a TV monitor, a Microsoft Kinect RGB-D sensor and a (optional) WWS Smartex e-shirt for clinical signs monitoring.

Customized algorithms for calibration, people segmentation, body skeletonization and hands tracking through the RGB-D sensor have been implemented in order to infer knowledge about the reaction of the end-user to the Graphical User Interface designed for specific cognitive domains. For proper interaction, gestures of Alzheimer Disease's patients are acquired by Microsoft Kinect in the nominal functioning range, allowing 100% hands detection rate, useful for an error free human-machine interaction.

Keywords: Natural User Interface, Active Vision System, Cognitive Rehabilitation.

1 Introduction

In the recent years, the phenomenon of ageing population is receiving increasing attention mainly for healthcare and social impacts, so a great effort has been addressed by the scientific community in order to provide specific enabling solutions. Alzheimer's Disease (AD) is a chronic neuro-degenerative disease (dementia) in which the first symptom is a slowly increasing memory loss. As the disease progresses, the brain deteriorates more rapidly with apparent cognitive limits.

Several scientific publications [1-5] highlight the usefulness and importance of Cognitive Rehabilitation (CR) in the treatment of patients with dementia. In order to increase the chances of an appropriate care, the development of a cost-effective home-care service with CR functionalities could be very useful. Moreover, although education and ICT skills level among the elderly is often low, various pilot studies on small samples show that ICT tools are accepted improving quality of life and increasing the permanence at home. In the field of healthcare, technologies such as virtual reality,

C. Stephanidis and M. Antona (Eds.): UAHCI/HCII 2014, Part III, LNCS 8515, pp. 300–309, 2014.

augmented reality and serious games have being applied for long time, including cognitive training and rehabilitation [6-8].

Virtual reality offers training environments in which human cognitive and functional performance can be accurately assessed and rehabilitated [9-10]. On the other hand, augmented reality provides safer and more intuitive interaction techniques allowing interaction with 3D objects in real world [11-12]. In this scenario social communication channels (natural speech, para-language, etc.) are not blocked, breaking down mental barriers applying such a technology to specific problems or disabilities. New solutions for cognitive assistance based on serious games have been implemented: in the field of CR commercial products (Nintendo's Brain Age, Big Brain Academy, etc.) have been tuned as educational tools helping to slow the decline of AD [13-16].

More recently, the large diffusion of interaction devices enabling body movements to control systems have been investigated, with specific focus on ICT technologies for natural interaction. Microsoft Kinect is the state-of-the-art [17] as device for body movements acquisition and gesture recognition; the effects of this kind of technology for rehabilitation purposes is widely investigated [18,19].

In this paper, a Natural User Interace (NUI) platform for remote CR has been designed with the aim to support AD patients during the rehabilitation pratice without the presence of any caregiver. A new hands tracking filter has been implemented with the aim to overcome the well-known limitations of the royalties-free NUI middleware architecture used in the platform.

The paper is organized as follows. Section 2 presents the overall implemented platform with specific focus on the rehabilitation practice in a cognitive multi-domain scenario. Section 3 presents the HCI methodology in which the customized hand tracking filter is described. Section 4 presents the evaluation of the proposed system on real data in different scenarios.

2 Platform Overview

The developed ICT platform provides a system for Cognitive Home Rehabilitation (called AL.TR.U.I.S.M.) through a customized Virtual Personal Trainer (VPT) allowing the patients to perform the rehabilitation practice at home. The platform gives the opportunity to perform the cognitive therapy without the presence of a caregiver. Details about the platform are in the following.

2.1 AL.TR.U.I.S.M. Platform Architecture

The CR platform is made up of a set-top-box connected to a TV monitor with Internet connection, a Microsoft Kinect RGB-D sensor for human body tracking and gesture recognition and a (optional) WWS Smartex e-shirt [20] with textile electrodes for clinical signs monitoring (Fig. 1).

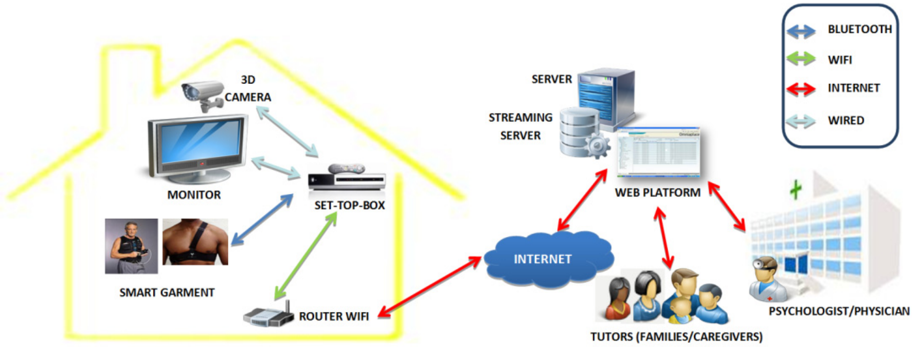

Fig. 1. AL.TR.U.I.S.M. platform overview

The left hand side part of the Fig. 1 shows the hardware equipments in which the set-top-box (a commercial embedded pc) is the gateway able to automatically downloads sequences of exercises from a remote server through the internet connection. The system provides a web-based platform that allows the physician to customize directly the therapy: this process is a highly innovative compared to existing systems [21-22] as the caregiver/physician defines a specific sequence of exercises (the therapeutic session) according to the residual abilities of the patient.

In order to infer more knowledge during the rehabilitation activities, the main clinical parameters (heart-rate, breathe-rate, electrocardiogram, etc.) are monitored by wearing the WWS Smartex e-shirt that integrates several sensing devices for biomedical applications. Through a Bluetooth radio link with the platform, each useful clinical parameter provided by the WWS device is stored on the set-top-box and, then sent to the physician with a multi-modal paradigm for clinical evaluations. Moreover, the platform integrates streaming functionalities allowing visual/audio recording for post-verification or online feedback to the physician which is able to follows the execution of the exercises from a remote architecture. From this perspective, the physician/psychologist of the reference center could communicate to the patients through a remote connection and then monitor the progress or trouble in the execution of the different required tasks. At the end of the rehabilitation session, the central platform collects different kinds of data locally stored on the set-top-box. An ad-hoc multimodal messaging procedure (e-mail, SMS, App, …) is performed and relevant data are sent to the physician allowing instant verification of the performance through an easy-to-use Graphical User Interface (GUI).

2.2 Multi-domain Cognitive Rehabilitation Practice

On the basis of the severity of cognitive impairment and residual skills of the target, the CR program provides specific categories of exercises, in order to assess specific domains of deficits. In order to make the system reliable, flexible and compliant with the international evaluation scales (Mini Mental State Examination [23]), few input

Fig. 2. Rehabilitation approach by gesture as Natural User Interface

parameters need to be defined a-priori (execution time, maximum numbers of allowed errors, movement sensitivity). The platform provides sixteen different exercises belonging to the following cognitive domains:

- Orientation (temporal, personnel and spatial)
- Memory (topographic, verbal and visual)
- Attention (hearing and visual)
- Categorization
- Verbal fluency
- Logic

In order to allow an appropriate display of every exercise and become independent from the specific output device (digital monitor, HD TV, …) a software module for the best video rendering is implemented. The definition of graphics objects displayed on the GUI has been designed according to the principles of ergonomics, usability and acceptability as referred in ISO/IEC 2001a [24].

3 Contact-Less Natural User Interface for Cognitive Rehabilitation

Each exercise is performed via a multi-modal contact-less NUI (Fig. 2) available through the use of both the Microsoft Kinect device for gesture recognition and Microsoft Text-To-Speech engine (TTS) [25] for human voice synthesis (italian language). The platform allows the user to interact with the GUI in a natural way without the use of a mouse or any kind of controller. The interaction is achieved by hand gestures performed by the patient according to the ad-hoc designed GUI, compliant the specific CR exercise. From the functioning principle point of view, the Kinect device is a RGB-D camera integrating both a high resolution RGB camera (640×480 at 30 fps) and an infrared depth sensor (640×480 at 30 fps), providing a metric reconstruction of a scene. The RGB-D device allows to capture mid-resolution depth and

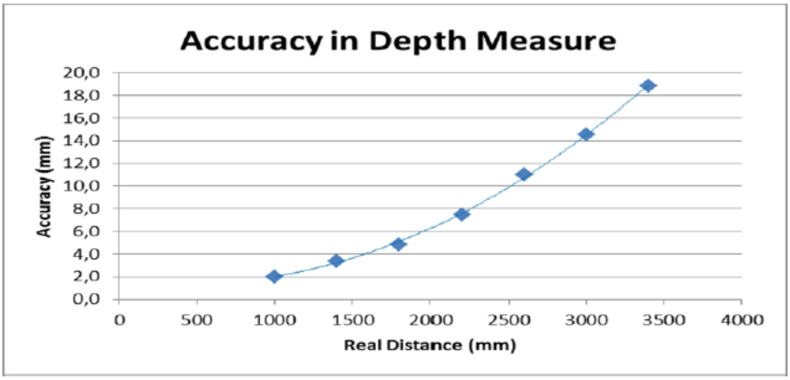

Fig. 3. Accuracy of Microsoft Kinect by varing the distance between target and the optical centre of the device

appearance information at high data rates (up to 30fps) for real-time functioning. The space resolution along the x and y axis is 3 mm at a depth of 2 meters, whereas the resolution of z-depth is 1 cm at the same depth. While increasing distance from the sensor, the accuracy decreases remaining within an acceptable range for people and hands tracking (Fig. 3).

3.1 Gesture Recognition by Hand Tracking for Human Computer Interaction

For the considered CR scenario, the upper part of the human body needs to be completely visible, avoiding situations in which large occlusions occur. As AD patient may have troubles moving the own hands, the procedure for hand tracking and gesture recognition provided by the Microsoft Kinect SDK [26] may be affected by critical issues, for example when hands and body torso are overlapped. In particular, AD patient may not be able to move hands in a spatially extensive environment, so that the interaction with objects belonging to the GUI could be hard and the CR practice could be dramatically affected. In this context graphical objects are codified as “hidden buttons” able to discover the hovering time for specific end-user choices. The NUI procedure is depicted in Fig. 4.

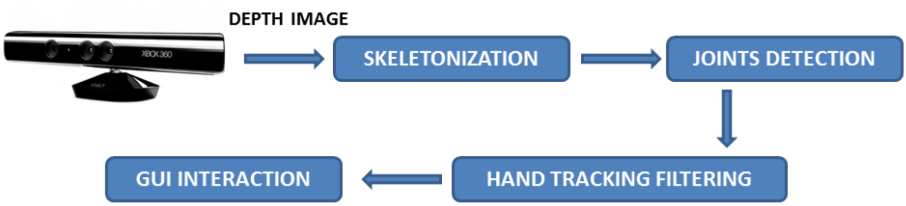

Fig. 4. Hand tracking procedure for contact-less natural user interaction

The skeletonization procedure provided by Microsoft Kinect SDK can be affected due to noise in the acquisition process so that the joint positions estimations can fail and the interaction with the GUI could be sometimes hard to handle. In order to overcome this kind of issue, a noise reduction filter has been designed removing as much as possible noise from raw data. The suggested filter operates as a smoothing filter that overcome the performaces of the Holt Double Exponential Smoothing Filter (HDESF) [27] built in Microsoft Kinect SDK. The HDESF procedure reduces the jitters from skeletal joint data providing a smoothing effect with lower latency than other smoothing filter algorithms. The main issue with HDESF application is that Kinect sensor does not have sufficient resolution to ensure consistent accuracy of the tracked joints over time. Observing real data, the problem is apparent when different joints are overlapped.

An improved smoothing algorithm has been designed, overcoming the performance of HDESF. From the analysis of the state of the art, the Exponential Weighted Moving Average Filter (EWMAF) [28] appears as the best trade-off between smoothing effect and jitter control. The EWMAF is given by the following formula:

$$S_t = \alpha X_{t-1} + (1-\alpha)\, S_{t-1} \qquad t > 1 \tag{1}$$

where X_t is the 2D coordinate sample at a time period t, S_t is the smoothed statistic as simple weighted average of the previous observation X_{t-1} and the previous smoothed statistic S_{t-1}. Generally, the setting of the initial smoothed value S_1 is performed by averaging the first six samples, limiting the delay of the interaction. The coefficient α is constant smoothing factor between 0 and 1 and it represents the decreasing weighting degree. Values of α close to 1 represents a lower smoothing effect, whereas values of α closer to 0 make a greater smoothing effect. Details about the best value of α are reported in Section 4. The last step of the hand tracking procedure involves the use of a function that allows to scale a joint's position to the maximum width and height specified (in our case the maximum screen resolution). Moreover, the setting of specific scaling factors simplifies the interaction of the end-user with the GUI, according the residual movement hands abilities of the end-user. For this purpose, three different combinations of scaling values (corresponding to different movement hands abilities) are set as reported in Tab. 1. High values of scaling require a wider real movement of the hand to cover the entire spatial resolution of the GUI, while low values of scaling allow the natural interaction of the end-user that may not be able to move the hand in a spatially extensive.

Table 1. Scaling parameters along the x axis and y axis of the screen

Level of Accuracy	*Scale X*	*Scale Y*
Low	0.6	0.4
Medium	0.3	0.2
High	0.15	0.1

4 Experimental Results and Discussion

The experimental evaluation of the proposed platform was based on real-world sequences obtained during the execution of specific movements acquired by Microsoft Kinect operating in the range 1.5 m - 4.0 m. The platform is hosted on an embedded PC equipped with an Intel® Core i5 CPU.

Since ground truth data for real-world image sequences is hard to obtain, according to the experimental section proposed in [29], the evaluation of hand tracking procedure was made analyzing the hand trajectory on the “Visual Attention” exercise GUI. The hand trajectory is reported in Fig. 5 in which a loop gesture is represented: the blu line is referred to the hand movement by using HDESF, while the red line is referred to the trajectory achieved by applying the proposed EWMAF. By comparing the two trajectories, it is apparent that EWMAF reduces the jittering of the signal, removing spikes presented in the hand tracking returned by HDESF. The analysis of the trajectories highlights the problems in some critical situations (yellow circle in Fig. 5): HDESF returns false paths that invalidate the process of graphic element selection, while the application of EWMAF allows the correct item selection by the end-user. The aforementioned issue may be negligible in gaming context, whereas in the considered CR scenario the consequences are apparent affecting the CR practice. A quantitative measurement of the performances can be carried out analyzing the difference between values calculated by a model (e.g. an ideal trajectory over the numbers placed on the board) and the tracked values recorded in the two previous cases.

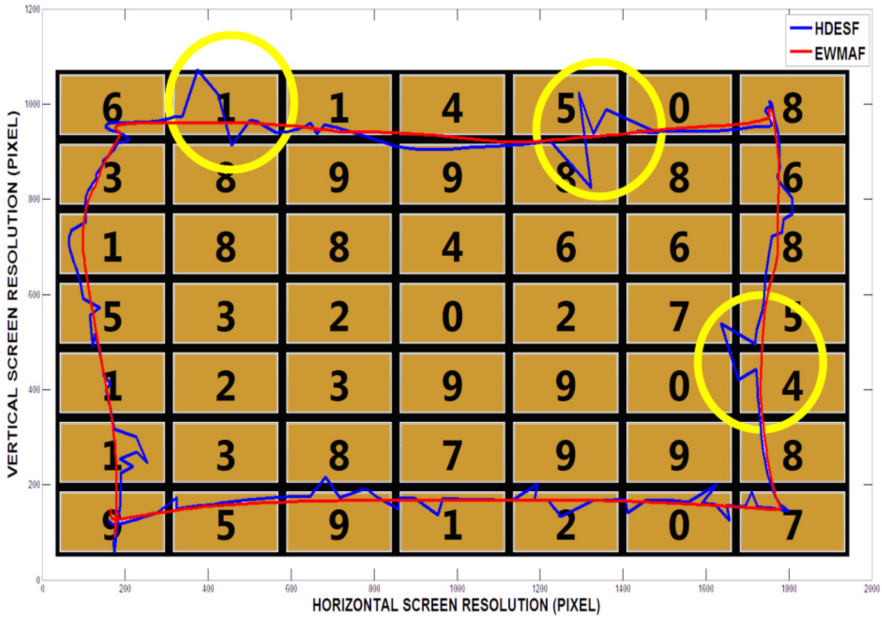

Fig. 5. Hand Trajectory obtained with HDESF (blu line) and EWMAF (red line) applications in “Visual Attention” exercise

For this purpose, a frequently used measure is the Root-Mean-Square Deviation (RMSD), that represents the sample standard deviation of the difference between predicted and observed values.

Table 2. RMSD values at different distance for HDESF and EWMAF on 500 samples

Distance from Kinect	*RMSD (HDESF, in pixel)*	*RMSD (EWMAF, in pixel)*
1,5 m.	60,30	22,87
2,5 m.	61,06	23,36
3 m.	68,12	25,67
4 m.	87,98	28,32

Table 2 shows the results obtained at four different distances, selected in the operating range of Microsoft Kinect; the sample number used to estimate RMSD is approximately 500. RMSD values for EWMAF demonstrate the validity of the proposed solution: the average deviation never exceeds the value of 30 pixel, allowing greater precision in the selection of a graphic element. Instead, HDESF performs an average error always greater than 60 pixels, so that the gap in some cases affects the correct selection of a graphic element belonging to the GUI.

EWMAF presents the issue of introducing a lag relative to the input data. In real time application latency is a critical factor, therefore it is essential to tune up in the rigth way the parameters of the fiter, with the aim of obtaining the best performance with the lowest latency.

Table 3. RMSD values by evaluating EWMAF for different α smoothing factor and amount of observations

# of observations	$\alpha = 0.1$	$\alpha = 0.3$	$\alpha = 0.7$	$\alpha = 0.9$
5	34,3212	29,1258	25,3427	23,0045
10	33,5662	28,9083	24,9980	22,6754
15	32,0113	28,3397	24,6548	22,3359
20	32,9675	28,0989	24,1765	21,8675
30	31,3444	27,5674	23,8413	21,6754

As mentioned in Section 3.1, values of α close to 1 presents a lower smoothing effect, since greater weight to recent changes in data is considered. A choice of $\alpha = 0.9$ allows EWMAF to reduce significantly the RMSD compared to the result obtained with $\alpha = 0.1$ and $\alpha = 0.3$ (see Table 3). On the other hand, when the number of past observations grows in size results are better but a latency effect is introduced in the tracking procedure. For all considered scenarios, the best tradeoff is achieved for $\alpha = 0.9$ and the amount of past observations equal to 10 (as 300 ms).

4.1 Computational Load Evaluation

The computational workload, referred to the three main steps of Fig.4 (hand tracking methodology), is evaluated in terms of processing time. The skeletonization and Joint detection algorithms require exactly the same time on the target machine (2.5 ms). On the other hand, the time spent for EWMAF application is invariant to the amount of past observation used for the evaluation of the hand position, requiring an average processing time of 1.4 ms. These average values do not affect the frame rate of the application and consequently the overall computational load allows the execution of every exercise at 30fps.

5 Conclusions

The main idea of the proposed system is to promote multi-domain cognitive rehabilitation of Alzheimer Disease patient at home through an ICT platform integrating low-cost contact-less Natural User Interface devices.

As a patient with Alzheimer's disorder may have trouble moving the own hands, a new filter for hand tracking has been implemented, overcoming the performances of the built-in filter of Microsoft Kinect SDK. The improvement allows an easy and accurate interaction of the end-user and the platform, even in the presence of complex ad-hoc designed GUI. The proposed platform allows both the evalutation of the progress of the dementia (useful for the caregiver) and the cognitive stimulation of the end-user in several domains. Future works are addressed to deploy the platform to a wide class of Alzheimer's disease patients in order to have the feedback (validation) during the technological tool usage. This will allow to tune up each component of the platform (GUI, knowledge discovery logic, filter parameters) in order to make it highly compliant with the needs and the requirements of the patients, according to the recent User Centered Design paradigm.

Acknowledgements. This work has been carried out within AL.TR.U.I.S.M. project, funded by Apulia Region under the Health-care call for the promotion of regional partnerships for innovation.

References

1. Clare, L., et al.: Cognitive Rehabilitation in Dementia. In: Neuropsychologistical Rehabilitation, pp. 193–196. Psychology Press Ltd. (2001)
2. Neal, M., Briggs, M.: Validation therapy for dementia (Cochrane Review). The Cochrane Library, Chichester, UK, vol. (2) (2004)
3. Spector, A., Orrell, M., Davies, S., Woods, B.: Reality orientation for Dementia. Cochrane Database Syst. Rev. (4): CD001119. Review (2000)
4. Woods, R.T., Bruce, E., Edwards, R.T., Hounsome, B., Keady, J., Moniz-Cook, E.D., Orrell, M., Russell, I.T.: Reminiscence groups for people with dementia and their family carers: pragmatic eight-center randomised trial of joint reminiscence and maintenance versus usual treatment: a protocol. Trials 10, 64–74 (2009)
5. Clare, L., Woods, R.T., Moniz Cook, E.D., Orrel, M., Spector, A.: Cognitive Rehabilitation in Alzheimer's Disease. Aging Clin. Exp. Res. 18, 141–143 (2006)
6. Richard, E., et al.: Augmented Reality for Rehabilitation of Cognitive Disable Children: A preliminary Study. Virtual Rehabilitation, 102–108 (2007)

7. Standen, P., Brown, D.: Virtual reality and its role in removing the barriers that turn cognitive impairments into intellectual disability. Virtual Reality 10(3), 241–252 (2006)
8. Taylor, M.J.D., et al.: Activity-promoting gaming systems in exercise and rehabilitation. Journal of Rehabilitation Research 48(10), 1171–1186 (2011)
9. Pugnetti, L., Mendozzi, L., Barbieri, E., Motta, A.: VR experience with neurological patients: basic cost/benefit issues. Stud. Health Technol. Inform. 58, 243–248 (1998)
10. Rizzo, A.A., Buckwalter, J.G.: The status of virtual reality for the cognitive rehabilitation of persons with neurological disorders and acquired brain injury. Stud. Health Technol. Inform. 39, 22–33 (1997)
11. Azuma, R., Baillot, Y., Behringer, R., Feiner, S., Julier, S., MacIntyre, B.: Recent advances in augmented reality. IEEE Comput. Graph. 21(6), 34–47 (2001)
12. Azuma, R.T.: A survey of augmented reality. Presence: Teleoperators and Virtual Environments 6(4), 355–385 (1997)
13. Nacke, L.E., et al.: Brain training for silver aged gamers: effects of age and game form on effectiveness, self-assessment, and gameplay. Cyberpsychology & Behavior 12(5), 493–499 (2009)
14. Imbeault, F., et al.: Serious Games in Cognitive Training for Alzheimer's Patients. In: IEEE Int. Conference on Serious Games and Applications for Health, pp. 122–129 (2011)
15. Jiang, C.F., Chen, D.K., Li, Y.S., Kuo, J.L.: Development of a computer-aided tool for evaluation and training in 3d spatial cognitive function. In: 19th IEEE Symposium on Computer-Based Medical Systems, pp. 241–244 (2006)
16. Tremblay, J.B., Bouchard, A., Bouzouane, A.: Adaptive game mechanics for learning purposes: making serious games playable and fun. In: Proc. Int. Conf. on Computer Supported Education, vol. 2, pp. 465–470 (April 2010)
17. http://www.microsoft.com/en-us/kinectforwindows
18. Da Gama, A., et al.: Improving Motor Rehabilitation Process through a Natural Interaction Based System Using Kinect Sensor. In: Proceedings of IEEE Symposium on 3D User Interfaces, pp. 145–146 (2012)
19. Lange, B., et al.: Development and evaluation of low cost game based balance rehabilitation tool using the Microsoft Kinect Sensor. In: 33rd IEEE International Conference on Engineering in Medicine and Biology Society (2011)
20. http://www.smartex.it
21. Soong, Y.L.W., Man Way, K.D.: A Tele-cognitive Rehabilitation Platform for Persons with Brain Injuries. In: TENCON IEEE Region 10 Conference (2006)
22. Solana, J., et al.: PREVIRNEC, A new platform for cognitive tele-rehabilitation. In: The Third Int. Conference on Advanced Cognitive Technologies and Applications (2011)
23. Folstein, M., Folstein, S., McHugh, P.: Mini-mental state: a practical method of grading the cognitive state of patients for the clinician. J. Psychiat. Res. 12, 189–198 (1975)
24. ISO/IEC. 2001. International Standard ISO/IEC 9126 – 1. Software engineering - Product quality - Part 1: Quality model. International Organization for Standardization/ International Electrotechnical Commision, Geneva (2001)
25. http://msdn.microsoft.com/en-us/library/hh361572%28v=office.14%29.aspx
26. http://www.microsoft.com/en-us/kinectforwindowsdev/start.aspx
27. Holt, C.C.: Forecasting seasonals and trends by exponentially weighted moving averages, ONR Memorandum, vol. 52. Carnegie Institute of Technology, Pittsburgh (1957), available from the Engineering Library, University of Texas at Austin
28. Roberts, S.W.: Control Chart Tests Based on Geometric Moving Averages. Technometrics (1959)
29. Sangheon, P., Sunjin, Y., Joongrock, K., Sungjin, K., Sangyoun, L.: 3D hand tracking using Kalman Filter in depth space. EURASIP Journal on Advances in Signal Processing 2012, 36 (2012)