Towards Sensing Behavior Using the Kinect

Wouter van Teijlingen[1,2], Egon L. van den Broek[1,3], Reinier Könemann[4], John G.M. Schavemaker[1]

[1] *Media and Network Services, Technical Sciences, TNO, Delft, The Netherlands*

[2] *Department of Computer Engineering, Faculty of Natural Sciences & Technology, Utrecht University of Applied Sciences, Utrecht, The Netherlands*

[3] *Human Media Interaction, Faculty of Electrical Engineering, Mathematics, and Computer Science, University of Twente, Enschede, The Netherlands*

[4] *Sustainable Productivity, Quality of Life, TNO, Hoofddorp, The Netherlands*

wouter@van-teijlingen.nl, vandenbroek@acm.org, reinier.konemann@tno.nl, john.schavemaker@tno.nl

Abstract

A method is proposed to validate Microsoft's Kinect as a device and, hence, to enable low fidelity, unobtrusive, robust sensing of behavior. The Xsens MVN suit is proposed as the measurements' ground truth. An overarching framework is introduced that facilitates a mapping of both devices upon each other. This framework includes a complete processing pipeline for both the Xsens and the Kinect data, recorded in parallel, which, at the end of both pipelines, are mapped upon reach other. Next, two strategies are presented that aim to interpret the data gathered. We close with a brief discussion on the pros and cons of the protocol proposed.

Introduction

In November 2010, Microsoft launched its Kinect, a motion sensing input device, as an alternative game controller for Xbox 360 video game consoles. Kinect competes with game controllers such as the Nintendo Wii's Remote Plus and Sony Computer Entertainment's PlayStation 3 Eye motion controllers. However, the Microsoft Kinect can also be used for various other applications than games. This has drawn the interest from science and engineering to this game controller. Also Kinect's low price (i.e., approx. €100,-), compared to that of other (traditional) scientific tools, has added to the interest it receives from science and engineering. Already within months from the introduction of the Kinect, various researches were presented and the first papers were published. The Kinect has already been used in various domains for several applications. For example, Gallo and colleagues [1] employed the Kinect to enable controller-free exploration of medical image data; Chang and colleagues [2] developed a Kinect-based system for physical rehabilitation; and Kamel Boulos and colleagues [3] introduced Kinoogle, a Kinect interface for interaction with Google Earth. This triplet illustrates the vast amount of research and development that has already been done using the Kinect, within two years since its introduction.

This article addresses yet another application domain for the Kinect: *sensing behavior*. This domain is not new [4,5]; however, the Kinect can make sensing behavior affordable for every day's (real world) practice. Next, in Section 2, we present our research rationale including its key elements. In Section 3, we discuss our quest towards low fidelity, robust, unobtrusive sensing of behavior. Its essence is a method to validate the Kinect for (real-time) behavior classification, using the Xsens suit (see also [6]). We close this article with Section 6, which provides a brief discussion and the conclusion.

On sensing of behavior

Traditionally (e.g., in experimental psychophysiology), "human motor activity is indexed by surface electromyography (EMG), which requires the placement of electrodes on the skin surface. This limits the mobility of the subject and reduces the possible contexts in which motor responses can be recorded and evaluated. In addition, EMG captures only the activity of specific muscle groups … ." [7]. With the rise of computing power and the steep progress of image processing and computer vision, behavior monitoring became

possible via the visual domain [4,5]. However, as with EMG-based monitoring, computer vision-based monitoring of human behavior requires a very high level expertise and rather expensive apparatus [7,8]. With the Kinect these two issues seem to resolve and, additionally, traditional computer vision is augmented with an infrared (IR) depth sensor [9]. Sensing behavior is easier said than done. It starts with agreeing upon a definition of behavior. Although all of us have an intuitive understanding of what behavior is, providing a stipulate definition remains challenge. In particular, in engineering human-related concepts, such as behavior is, are often ill defined; in other words, such research has poor construct validity. To tackle this traditional flaw, we define behavior as an observable, measurable movement of some part of the body through space and time. Moreover, we further specify behavior by providing concise definitions of motion, action, and activity, as given in Table 1.

Validation of the Kinect: An overarching framework

In parallel, motion data is captured with the Kinect camera and the Xsens MVN body suit. An overarching framework is designed that includes two processing pipelines that are connected at their start (i.e., the data acquisition) and at their end (i.e., their mapping); see also Figure 1. First, we will describe the Xsens processing pipeline. Second, we will describe the Kinect processing pipeline. Third and last, we will describe their mapping. The Xsens body suit captured motion via 17 inertial motion trackers (i.e., 3D gyroscopes, 3D accelerometers, and 3D magnetometers). Their data is sent to the affiliated MVN suit, which saves it as a Biovision Hierarchy file (BVH) file, an animation file type (see also Figure 1). The Kinect captures an infrared signal (IR) and a video (RGB) stream in parallel. Both are sent to the newly developed ONI (OpenNI) recorder, which was founded on the Nestk library [10]. The ONI recorder generates an ONI file, which is sent to the newly developed Kinalyze module. The Kinalyze module converts the ONI file to a BVH file. Hence, both processing pipelines have a BVH file as output, which can be mapped upon each other. Having two BVH files, one that originates from the Xsens body suit and one that originates from the Kinect, a mapping between them can be realized. This requires the alignment of both data files. Next, a digital human model (DHM) through time has to be generated using both BHV files. To realize this, angles between joints have to be determined. Subsequently, the movements of the DHM have to be interpreted, which depends on the context of use. This phase of interpretation will be discussed next.

Table 1. Classification of four level of behavior: motion, action, and activity, adapted from Chaaraoui et al. [8]

	Time window	Description
motion	(parts of) seconds	Image segmentation, movement detection, and gaze and head-pose estimation.
action	seconds – minutes	Identify interaction of users with objects. Recognize primitives such as sitting, standing, walking, and running.
activity	minutes – hours	Sequence of actions with a particular order. Subsequently, recognition of daily activities (e.g., cooking or showering).

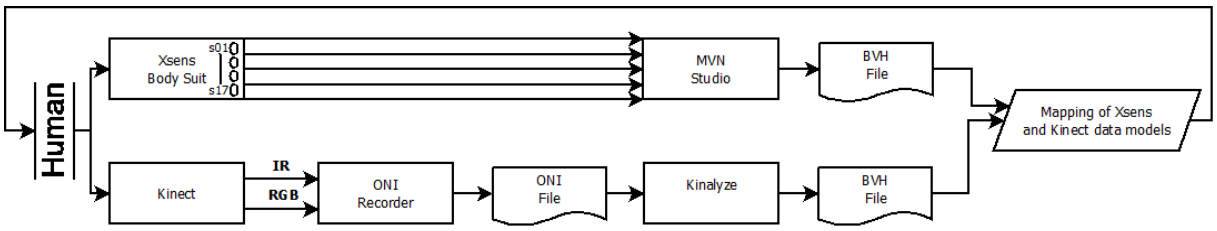

Figure 1. An overarching framework or processing pipeline for mapping the Xsens body suit data and the Kinect data upon each other.

Towards low fidelity, unobtrusive, robust sensing of behavior

Microsoft Kinect and the Xsens MVN exploit distinct modalities that can be used in parallel, hardly without any restrictions. This enables parallel data gathering for both devices. As such, an optimal set up can be achieved to validate the precision of the Kinect's measurements in practice. Although, the Kinect has been evaluated for kinematic measurement [9,11], to the authors' knowledge, it has not yet been compared directly with the Xsens MVN suit as a solid ground truth. When tracking of motion, action, and activities with a high temporal and spatial precision is needed, Xsens is the preferred device among this duo [6]. However, in practice, for various (real life) applications activity monitoring is at hand (see also Table 1), where precision in both the temporal and spatial domain is less important and ease of use is crucial. In this case, Kinect could provide an interesting low fidelity alternative, as it enables unobtrusive and robust sensing of activity and behavior [cf. 1,2,3,9,11].

One of the core differences between the output of Xsens and Kinect is the DOF in their measurement [6]. Xsens provides 45 DOF model (i.e., 23 segments and 22 joints) to characterize movements, where Kinect relies on 15 DOF (i.e., when using the ONI skeleton tracking algorithm). This is a crucial distinction when analyzing motion and action. The software packages available with both Xsens and Kinect support the use of skeleton models, which provide accurate modeling of motion and, subsequently, action. This enables knowledge-based dimension reduction. Depending on the task or context at hand, possibly only parts of the models are of interest. For example, when sitting behind a desk, the lower part of the body is of little interest. We will select the appropriate DOF of Xsens and, subsequently, select Kinect's DOF that map upon Xsens's DOF. Where no mapping is possible, we aim to determine mappings by way of interpolation.

An alternative for knowledge-based dimension reduction, as can be done using skeleton models, is data-driven dimension reduction. To exploit this, a set of standard motions, actions, and activities will be identified. Participants in experiments will be asked to execute these motions, actions, and activities, while wearing the Xsens MVN suit and while being observed by the Kinect. Offline signal processing and classification of the signals can be utilized to explore possible mapping between the signals generated by both devices. In a nutshell, of both devices the signals are captured, the signals are processed. A measurement space is defined, preprocessing (e.g., filtering and artifact removal), synchronization, and segmentation is applied. Next, features and, subsequently, their parameters are extracted from each of the signals. A pattern space is identified and appropriate parameters are selected, which results in a reduced pattern space. This pattern space is fed to a classifier (e.g., principal component analysis or a support vector machine, SVM). In case of supervised learning (i.e., the SVM), errors in the classification process are detected and the classification process is adapted. Note that the latter phase requires appropriate validation, which can be realized by separating training and test sets or by conducting cross validation.

Conclusion

This article merely proposed a method to validate Microsoft's Kinect as a device to enable low fidelity, unobtrusive, robust sensing of behavior. The Xsens MVN suit is presented as means to establish the measurements' ground truth (cf. [9,11]). To facilitate this, an overarching framework has been presented that facilitates the mapping of both devices' data streams. Moreover, the required alignment of both data streams is discussed to achieve this mapping.

The Kinect has tremendous advantages compared to most other devices. However, similar as most computer vision based approaches, the Kinect suffers from the physiognomies of body and faces. These vary considerably among individuals due to age, ethnicity, gender, facial hair, cosmetic products, and occluding objects (e.g., glasses and hair). Furthermore, both body and face can appear to be distinct from itself due to pose and/or lighting changes, or other forms of environmental noise. When the Kinect indeed would prove itself as a low fidelity, unobtrusive, and robust device for sensing of behavior, this could yield significant progress on measuring behavior in general [9,11] (cf. [1,2,3]). Its broad usage has the significant advantage that software packages are constantly both improved and extended (e.g., [10]). High level programming languages enable easy

access, processing, and interpretation of Kinect data. Consequently, we expect that the measuring behavior community will embrace the Kinect and include it in its standard measurement setups.

Acknowledgments

This publication was supported by the Dutch national program COMMIT (project P7 SWELL). The authors also thank Joop Kaldeway (Utrecht University of Applied Sciences) and Tim Bosch (TNO) for their comments. Further, the authors gratefully acknowledge the two anonymous reviewers for their to the point and constructive remarks.

References

1. Gallo, L., Placitelli, A.P., and Ciampi, M. (2011). Controller-free exploration of medical image data: Experiencing the Kinect. In M. Olive and T. Solomonides (Eds.), *Proceedings of the 24th International Symposium on Computer-Based Medical Systems (CBMS)* (Bristol, UK, 27-30 June 2011), 1-6. Piscataway, NJ, USA: IEEE.
2. Chang, Y.-J., Chen, S.-F., Huang, J.-D. (2011). A Kinect-based system for physical rehabilitation: A pilot study for young adults with motor disabilities. *Research in Developmental Disabilities* **32**(6), 2566-2570.
3. Boulos, M.N.K., Blanchard, B.J., Walker, C., Montero, J., Tripathy, A. and Gutierrez-Osuna, R. (2011). Web GIS in practice X: a Microsoft Kinect natural user interface for Google Earth navigation. *International Journal of Health Geographics* **10**(1), 45.
4. Klein Breteler, M.D., Meulenbroek, R.G.J., Gielen, S.C.A.M. (1998). Geometric features of workspace and joint-space paths of 3D reaching movements. *Acta Psychologica* **100**(1-2), 37-53.
5. Noldus , L.P.J.J., Spink, A.J., Tegelenbosch, R.A.J. (2001). EthoVision: A versatile video tracking system for automation of behavioral experiments. *Behavior Research Methods, Instruments, & Computers* **33**(3), 398-414.
6. Roetenberg, D., Luinge, H, and Slycke, P. (2008). 6 DOF motion analysis using inertial sensors. In *Proceedings of Measuring Behavior 2008: 6th International Conference on Methods and Techniques in Behavioral Research* (Maastricht, The Netherlands, 26-29 August 2008), 14-15. Wageningen, The Netherlands: Noldus Information Technology.
7. Vousdoukas et al. (*in press*). SVMT: A MATLAB toolbox for stereo-vision motion tracking of motion reactivity. *Computer Methods and Programs in Biomedicine*.

   DOI: http://dx.doi.org/10.1016/j.cmpb.2012.01.006
8. Chaaraoui, A.A., Climent-Pérez, P., and Flórez-Revuelta, F. (2012). A review on vision techniques applied to human behaviour analysis for ambient-assisted living. *Expert Systems with Applications* **39**(12), 10873-10888.
9. Dutta, T. (2012). Evaluation of the Kinect sensor for 3-D kinematic measurement in the workplace. *Applied Ergonomics* **43**(4), 645-649.
10. Burrus N. (2012). Developing your own software based on the nestk library. <http://labs.manctl.com/rgbdemo/index.php/documentation/nestk>. Accessed 1 July 2012.
11. Ray, S.J., Teizer, J. (2012). Real-time construction worker posture analysis for ergonomics training. *Advanced Engineering Informatics* **26**(2), 439-455.