Online Segmentation of Continuous Gesturing in Interfaces

F.W. Fikkert, P.E. van der Vet, and A. Nijholt

Human Media Interaction, Department of Computer Science,
University of Twente, Enschede, The Netherlands,
{f.w.fikkert|p.e.vandervet|a.nijholt}@ewi.utwente.nl

Segmentation of Continuous Gesturing

An increasing demand exists for more intuitive ways to interact with ever larger displays. Natural interfaces fulfill this demand by directly analyzing, reacting to and reasoning about observed human behavior. We focus on gesturing which is a key modality in a more complex natural interface. Most gesture recognition research today focuses on computer vision techniques. Although promising, such techniques are not yet mature enough for markerless, robust and in-the-field deployment. Instead, we deploy a more down-to-earth solution by using existing motion capture systems. Our goal is to develop methods for automatic online segmentation and interpretation of continuous gesturing. We propose a two-stage approach. First, we use commodity hardware equipped with accelerometers and buttons that explicitly mark gesture boundaries. Experience gained there feeds our second stage in which we investigate motion trajectories and their meaning. In the second stage we use a combination of a full body motion capture suit and two data gloves.

Motion segmentation – extracting distinct units from continuous motion – is hard because the boundaries are both subjective and highly sequence-dependent. More so, even predefined gestures are made with variation. Individual users are not fully consistent and two users do not make exactly the same gestures [1]. Applications of automatic gesture segmentation deal with this variety by basing start and end cues on local minima or turn points in the trajectory [2]. This approach is applied often, with varying success, for more coarse, whole body movements such as dancing [3] and conducting music [4]. For finer motion such as sign language [5], extensive machine classifier solutions are mostly used. They are fed with the whole motion sequence. These solutions require extensive training and their computational load prevents real-time motion analysis.

Strong, familiar metaphors are the key of a natural interface [6]. We have devised a series of user experiments in which we hope to discover a gesture repertoire consisting of such metaphors [7]. We distinguish two semantically separate classes of metaphors in gesture-based interfaces. The first class involves manipulations of virtual 3D objects as if they were tangible. This form of interaction is directly based on the way humans interact with their everyday surroundings. Examples are picking up, rotating and zooming of a 3D mesh. The second class involves more abstract metaphors that are ideally inspired by everyday activities [8]. Gesturing to throw an item away means to delete it, for example. This class contains the six communicative gesture classes defined by [9]. Clearly, these two classes are strongly linked. Consider a case in which a person selects two visualizations – by picking them up – and then analytically relates them to each other – by moving them together. The key is to both find these metaphors and to recognize the intended meaning of the gestures.

Approach and results

To discover gesture boundaries and gesture meaning we propose a two-stage approach. In the first stage, we have collected trajectories and explicit gesture boundaries using commodity hardware: the WiiRemote controller from Nintendo's Wii game console. Gesture trajectories and boundaries were gathered through a user study. Users interacted with both a 3D mesh and a collection of 2D images. Trajectories were explicitly bounded using the buttons on the WiiRemote. We will show that these explicitly marked gesture boundaries will feed our segmentation method. This is shown by comparing multiple machine classifier techniques that are trained on these data to recognize the gesture boundaries. For our research, the WiiRemote functions as a stepping stone for high-resolution full body motion capture that we will deploy in stage two. We work towards this complexity because human movement is an effort of the whole body. This system will combine a full body motion tracking suit with a pair of CyberGloveII data gloves.

Acknowledgements

References

1. J. Epps, S. Lichman, and M. Wu. (2006) A study of hand shape use in tabletop gesture interaction, presented at CHI '06: *CHI '06 extended abstracts on Human factors in computing systems*, Montreal, Quebec, Canada

2. J. Barbič, A. Safonova, J.-Y. Pan, C. Faloutsos, J. Hodgins, and N. Pollard. (2004) Segmenting motion capture data into distinct behaviors. presented at GI '04: *Proceedings of Graphics Interface 2004*, London, Ontario, Canada

3. K. Kahol, P. Tripathi, S. Panchanathan, and T. Rikakis. (2003) Gesture segmentation in complex motion sequences. presented at *ICIP 2003 - International Conference on Image Processing.*

4. T.-S. Wang, H.-Y. Shum, Y.-Q. Xu, and N.-N. Zheng. (2001) Unsupervised Analysis of Human Gestures. in Advances in Multimedia Information Processing — PCM 2001, vol. 2195/2001, *Lecture Notes in Computer Science: Springer Berlin / Heidelberg* pp. 174-181.

5. S. Ong and S. Ranganath, Automatic Sign Language Analysis: A Survey and the Future beyond Lexical Meaning, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **27**, pp. 873-891, 2005.

6. T. Ni, G. Schmidt, O. Staadt, M. Livingston, R. Ball, and R. May. (2006) A Survey of Large High-Resolution Display Technologies, Techniques, and Applications, presented at *Proceedings of IEEE Virtual Reality 2006*, Alexandria, VA

7. W. Fikkert and P. van der Vet. (2007) Towards a Gesture Repertoire for Cooperative Interaction with Large Displays, presented at *7th International Workshop on Gesture in Human-Computer Interaction and Simulation* (GW 2007), Lisbon, Portugal

8. T. Grossman, R. Balakrishnan, G. Kurtenbach, G. Fitzmaurice, A. Khan, and B. Buxton. (2002) Creating principal 3D curves with digital tape drawing, presented at CHI '02: Proceedings of the SIGCHI conference on Human factors in computing systems, Minneapolis, Minnesota, USA

9. A. Kendon. (2004) Gesture: Visible Action as Utterance