HAND GESTURES FOR REMOTE CONTROL

Sigal Berman
Department of Industrial Engineering and Management
Ben-Gurion University of the Negev
Beer-Sheva 84105, Israel
E-mail:sigalbe@bgu.ac.il

KEYWORDS
Gesture recognition, Remote control, Human computer interaction

ABSTRACT

Gestures can be used as an alternative to traditional human computer interaction devices such as the keyboard and mouse. Gesture interaction can be flexible, natural, and intuitive. They can increase user immersion in virtual environments and reduce the need for learning systems specific interaction methods. In the talk, the general structure and main building blocks of gesture recognition systems will be discussed. The concepts discussed will be demonstrated as part of two applications employing hand gestures for remote control: control of the TV menu and control of a robotic manipulator.

INTRODUCTION

People use gestures as part of their daily interaction for communication of directives, thoughts, and feelings, enhancing or replacing spoken words. Use of gestures for human-computer interaction (HCI) can replace traditional interface devices such as mouse, keyboard or hand-held remote control. Gestures offer a natural, intuitive, and flexible HCI modality.

For computer-based simulation training, gestures can offer several advantages over traditional interfaces. Gestures can improve user immersion in virtual environment, they can reduce or even eliminated the need for learning simulation specific interaction methods, and gestures can enhance training to include actual interaction required in the physical scenario.

Previously gesture-based HCI was limited by the requirement for expensive and/or cumbersome sensing equipment such as magnetic trackers, sensor gloves, high-end image sensors, and touch screens. In recent years a plethora of low cost sensors suitable for gesture-based HCI has entered the market. For unencumbered gestures sensors such as the Kinect™ vision sensor suitable for remote hand and full body gesturing (0.5-5 meter) or the newly introduced Leep Motion™ controller for near hand and finger gestures (up to 1 meter) are now available at very low costs. For encumbered gestures various technologies are readily available and touch screens are now fully integrated within Windows 7™ operating system.

In order to utilize the benefits of gesture-based HCI it is important to gain understanding of gesture recognition technology. Gesture recognition systems (GRS) integrate both hardware and software components which should suit the task, environment, and user capabilities. In the current talk GRS architecture and main building blocks will be presented along with two examples of GRS for remote control: a GRS for television menu control and a GRS for tele-operation of a robotic manipulator.

GESTURE RECOGNITION SYSTEMS

The GRS operation cycle (Figure 1) includes the gesture executed by the user, which is sensed using a capture device and recognized using various algorithms (segmentation, tracking, feature extraction, classification). The recognized gesture is reported to the controlled application, which changes its state accordingly, and displays the outcome to the user. Due to the high interest in GRS, there are many reviews of their operation from various perspectives (e.g., (Aggarwal and Park 2004; Berman and Stern 2012; Garg and Aggarwal 2009; Mitra and Acharya 2007; Rautaray and Agrawal 2012; Wachs et al. 2011; Yilmaz et al. 2006).

There are many different types of gestures. Gestures may be performed using different body parts such as hands, head, legs, full body etc. Currently most GRS are based on hand gestures. Manipulative gestures are gestures where the executed motion directly influences the trajectory of the controlled object, similar to the relationship between the motion of the computer mouse and the curser on the screen. Semantic gestures encode commands by gesture features similar to the operation of keyboard keys. In dynamic gestures meaning is encoded by motion and in static gestures the meaning is encoded by the static hand or arm pose. Hybrid gestures encode meaning by both pose and motion. A gesture vocabulary is the set of gestures used within the application context.

The required building blocks of GRS are greatly influenced by the context of use, that is, by the gesture vocabulary and by the environment in which the system operates. For instance when manipulation gestures are

Proceedings 27th European Conference on Modelling and Simulation ©ECMS Webjørn Rekdalsbakken, Robin T. Bye, Houxiang Zhang (Editors)
ISBN: 978-0-9564944-6-7 / ISBN: 978-0-9564944-7-4 (CD)

used, the system does not require a classification module, or when the environment has very poor illumination RGB cameras cannot be used for gesture capture.

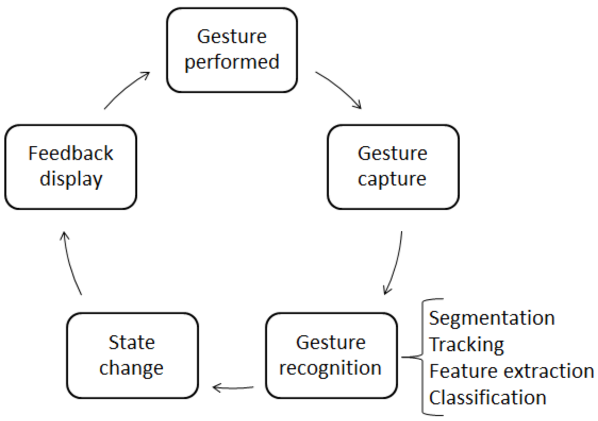

Figure 1: Gesture Recognition Cycle

Gesture capture devices can be divided by gesture stimuli into five groups (Berman and Stern, 2011): electric, optic, acoustic, magnetic, and mechanic. While optic based gesture capture facilitates unencumbered (free-hand) gestures, other stimuli require some assistive device either worn or in contact with the hand/body for gesture capture. For unencumbered GRS 2D vision sensors are still in use, but the low cost of currently available vision based distance sensors makes them the popular choice for GRS implementation.

Motion tracking is required when using manipulative or dynamic gestures. In vision-based GRS motion tracking requires hand segmentation from the background image. Segmentation difficulty is greatly alleviated when using distance information available from 3D vision sensors. Color cue is often used for hand tracking as people have a distinctive Hue range (Stern et al. 2010).

Classification is required when using symbolic gestures. With dynamic symbolic gestures, tracking provides a stream of segmented hand positions, and the classification algorithm must also determine gesture start and end. An additional challenge stems from motion variability within and between users. Hidden Markov Models (HMM), Dynamic Time Warping (DTW), and Longest Common Sub-sequence (LCS) are among the successfully applied classification methods (Frolova et al. 2012; Stern et al. 2013).

GRS FOR TELEVISION OPERATION

A vision-based GRS was developed for remote control of an Internet Protocol TV (IPTV) in a living-room environment (Stern et al, 2010). The living-room environment dictates a recognition range of up to a few meters and robustness to low and changing lighting conditions. Two additional constraints were imposed for the developed system: use of low-cost hardware components facilitating suitability to the consumer market, and adhering to the existing TV set-top box interface.

The gesture vocabulary was based on dynamic symbolic hand gestures. Dynamic unencumbered hand-gestures were used since their recognition is robust even with low quality sensing equipment, due to the distance requirements, and to facilitate "come-as-you-are" operation. Symbolic gestures were used for interfacing to the existing displays. Two system prototypes were developed based on 2D and 3D vision sensors. The vision sensors were located above or below the TV screen so several users sitting in the room could use the system together. For the 2D sensor a modified CAMShift algorithm was implemented based on motion and color cues with face detection and handling. For the 3D camera the OpenNI (http://www.openni.org/) tracking functionality was used. Two gesture classification algorithms were implemented: a tree-based classifier ensemble based on gesture trajectory features, and a modified Longest Common Subsequence (LCS) classifier. The system was implemented using a standard home theater streamer.

While system feasibility was clearly demonstrated the requirement of adhering to the existing interface was very limiting. The displays were originally developed for hand-held remote operation. Using the same displays for gesture-based control prevented full exploitation of the usability facilitated by gesture-based interfaces.

GRS FOR TELE-OPERATION

A sensor-glove based system was developed for tele-operation of a robotic manipulator (Berman et al. 2008). Maintaining a spatial resemblance between human and robotic motion is problematic when the structure of the robotic system is considerably different from that of the human. In order to harness the high level task understanding capabilities of the human while alleviate the burden of translating the required actions to those feasible for the robotic system an object-action abstraction was implemented. Such an abstraction is in-line with current finding on manipulation task representations in the central nerves system (CNS). Human actions in a virtual environment were recognized based on hand and arm motion and the object on which they were performed. A robotic grasp suitable for the object and recognized action was selected from a set of a-priori programed grasps. The object motion trajectory was formed based on the demonstrated trajectory.

The approach feasibility was demonstrated for control of a five degree of freedom serial manipulator. The tasks performed included building a block tower, pouring from a jar and screwing/unscrewing jar lids. Users unfamiliar with robotic operation succeeded in performing the missions, yet their performance improved over time as they gained understanding of system operation. That is the users did learn interface specific knowhow that aided them in attaining improved operation.

SUMMARY AND FUTURE RESEARCH

Gesture-based interfaces offer a highly natural modality for HCI. Available low-cost sensors today facilitate fast development of GRS for various applications. GRS have both hardware and software component with a multitude of different implementation options. Understating GRS operation and the different alternatives is important for implementing a high quality interface.

ACKNOWLEDGEMENTS

This work was supported in part by Deutsche Telekom AG and by the Paul Ivanier Center for Robotics Research.

REFERENCES

Aggarwal J.K. and Park, S. 2004. "Human motion: Modeling and recognition of actions and interactions". *in Proc. 2nd Int. Symp. 3D Data Process., Vis., Transmiss*., 640–647.

Berman, S. Friedman, J., Bakır, G. and Flash, T. 2008. Action Identification for Teleoperation Based on Object - Action Abstraction, *IEEE SMC International Conference on Distributed Human-Machine Systems (DHMS)*.

Berman, S. and Stern, H. 2012. "Sensors for Gesture Recognition Systems", *IEEE Transactions on Systems, Man, and Cybernetics-Part C* 42(3), 277-290.

Garg, P. Aggarwal, N. and Sofat, S. 2009. "Vision based hand gesture recognition*". in Proc. World Acad. Sci., Eng. Technol*. 49, 972–977.

Frolova, D., Stern, H. and Berman, S. 2012. "Most Probable Longest Common Subsequence for Recognition of Gesture Character Input". *IEEE Systems Man and Cybernetics, part B, September (in press)*.

Mitra S. and Acharya, T. 2007. "Gesture recognition: A survey". *IEEE Trans. Syst., Man, Cybern*. 37(3), 311–324.

Rautaray, S.S. and Agrawal, A. 2012. "Vision based hand gesture recognition for human computer interaction: a survey". *Artificial Intelligence Review*, 1-54.

Stern, H., Frolova, D. and Berman S., 2010. Hand Gesture Recognition for TV Remote Control using Tree-Based Ensemble and LCS Classifiers, *WORLDCOMP'10 The 2010 World Congress in Computer Science, Computer Engineering, and Applied Computing*.

Stern, H., Shmueli, M. and Berman, S. 2013. "Most Discriminating Segment - Longest Common Subsequence (MDSLCS) Algorithm for Dynamic Hand Gesture Classification". *Pattern Recognition Letters* (Special issue on 'Smart Approaches for Human Action Recognition'), (in press).

Wachs, J., Kolsch, M., Stern, H. and Edan, Y. 2011. "Vision-based hand gesture applications: Challenges and innovations". *Commun. ACM* 54(2), 60–71.

Yilmaz, A. Javed, O. and Shah, M. 2006. "Object Tracking: A Survey". *ACM Comput. Survey* 38(4), 1–45, 2006.

AUTHOR BIOGRAPHIES

SIGAL BERMAN is a lecturer and head of the intelligent system M.Sc. track in the Department of Industrial Engineering and Management at the Ben-Gurion University of the Negev, Beer-Sheva. She received a M.Sc. in Electrical and Computer Engineering, and a Ph.D. in Industrial Engineering, both from Ben-Gurion University of the Negev, and a B.Sc. in Electrical and Computer Engineering, from the Technion, Haifa; Dr. Berman was the primary investigator of the gesture recognition for remote control user interface project for Deutsche Telekom laboratories at BGU. Her research interests include: robotics and motor control, computer integrated manufacturing, and human machine interfaces. Her e-mail address: sigalbe@bgu.ac.il and her Web-page can be found at http://www.bgu.ac.il/~sigalbe.