Human-Computer Intelligent Interaction: A Survey

Michael Lew[1], Erwin M. Bakker[1], Nicu Sebe[2], and Thomas S. Huang[3]

[1] LIACS Media Lab, Leiden University, The Netherlands
[2] ISIS Group, University of Amsterdam, The Netherlands
[3] Beckman Institute, University of Illinois at Urbana-Champaign, USA

Abstract. Human-computer interaction (HCI) is one of the foremost challenges of our society. New paradigms for interacting with computers are being developed which will define the 21st century and enable the world to communicate and interact effortlessly and intuitively. In this short survey, we explain the major research clusters comprising the state of the art, and indicate promising future research directions.

1 Introduction

Historians will refer to our time as the Age of Information. While information is indeed important, the machine interface, the interactive process between humans and computers will define the 21st century. In this short survey, we are primarily interested in human-computer *intelligent* interaction as opposed to *simple* human-computer interaction. When a user types a document at a word processor, there is a form of simple human-computer interaction - the human types and the computer shows the human the formatted keystrokes composing the document. However, this would not be considered *intelligent* interaction because the computer is not performing any intelligent processing on the keystrokes; there is no meaning extracted from the user's actions. The computer is simply mirroring the user's actions.

From a research perspective, we are interested in *intelligent* interaction where the computer understands the *meaning* of the message of the user and also the *context* of the message. As an example, interaction between humans is typically performed using speech and body gestures. The speech carries the meaning of the message but often not the context. Is the person happy or sad? Is the person serious or joking? To understand the context, it is also necessary to grasp the facial and body gestures. If we are to have truly intuitive communication, computers will need to have their own sense of vision and speech to naturally fit into the world of humans.

How can we achieve synergism between humans and computers? The term "Human-Centered Computing" is used to emphasize the fact that although all existing information systems were designed with human users in mind, many of them are far from being user friendly. Research in face detection and expression recognition [1, 2, 19] will someday improve the human machine synergism. Information systems are ubiquitous in all human endeavors including scientific, medical, military, transportation, and consumer. Individual users use them for learning, searching for

M. Lew et al. (Eds.): HCI 2007, LNCS 4796, pp. 1–5, 2007.

information (including data mining), doing research (including visual computing), and authoring. Multiple users (groups of users, and groups of groups of users) use them for communication and collaboration. And either single or multiple users use them for entertainment.

2 Current Research

In Human-Computer Intelligent Interaction, the three dominant clusters of research are currently in the areas of face analysis, body analysis, and complete systems which integrate multiple cues or modalities to give intuitive interaction capabilities to computers. Face and human body analysis includes modeling the face or body and tracking the motion and location. They are considered fundamental enabling technologies towards HCI.

Understanding the dynamics of the human face is certainly one of the major research challenges. The potential of the following work is toward detecting and recognizing emotional states and improving computer to human communication. Valstar and Pantic[3] propose a hybrid technique where a Support Vector Machine (SVM) is used to segment the facial movements into temporal units. Then a Hidden Markov Model is used for classifying temporal actions. Chang et al. [4] propose a framework for using multiple cameras in pose and gaze estimation using a particle system approach.

Another major research challenge is human body analysis. The following research has the potential to endow computers to see where humans are, what their gestures and motions are, and track them over time. Oerlemans, et al.[5] propose a tracking system which uses the multidimensional maximum likelihood approach toward detecting and identifying moving objects in video. Their system has the additional novel aspect of allowing the user to interactively give feedback as to whether segmented objects are part of the background or foreground. Cooper and Bowden[6] address the problem of large lexicon sign language understanding. The method detects patterns of movement in two stages involving viseme classifiers and Markov chains. Angelopoulou et al.[8] are able to model and track nonrigid objects using a growing neural gas network. Jung, et al.[9] use a multi-cue method to recognize human gestures using AdaBoost and Fisher discriminant analysis. Chu and Nevatia[10] are able to track a 3D model of the human body from silhouettes using a particle filtering approach and infra-red cameras.

In addition to novel face and human body analysis, current research is creating systems which are delving into the complete human-computer interaction, thereby endowing computers with new important functionality. State[11] had created a system where a virtual human can maintain exact eye contact with the human user thereby significantly improving the perception of immersion. Rajko and Qian[12] propose an automatic kinematic model building method for optical motion capture using a Markov random field framework. Vural, et al.[13] created a system for detecting

driver drowsiness based on facial action units and classified using AdaBoost and multinomial ridge regression. Thomee, et al.[14] used an artificial imagination approach to allow the computer to create example images to improve relevance feedback in image retrieval. Barreto, et al.[15] use physiological responses and pupil dilation to recognize stress using an SVM.

In the area of interfaces, researchers are actively developing new systems. Hua, et al.[16] presented a system for mobile devices for a visual motion perceptual interface. A tabletop interface using camera detected finger interactions was proposed by Song, et. al.[17]. Interaction with a wall sized display was presented by Stødle, et al. [18], whereby the user could interact with the system using arm and hand movements.

3 Future Research Directions

As the current research indicates, we are making rapid progress in face analysis, human body analysis and creating the early generation of complete human-computer interactive systems. Several fundamental directions for the future include (1) User Understanding - learn more about the user [23] and create systems which will adapt appropriately to the user's needs; (2) Authentic emotion recognition - be able to reliably recognize the authentic emotional states of humans [19]; (3) Education and knowledge - develop new paradigms which improve education, learning, and the search for knowledge [20]; (4) New features and similarity measures [7] - for example, development of new texture models [21,22] focusing on HCI tasks such as face and body tracking integration with color and temporal spaces; and (5) Benchmarking in HCI - in many areas such as body tracking there are negligible ground truth sets which would be considered scientifically definitive. We need to decide on the most important HCI tasks and collectively create credible ground truth sets for evaluating and improving our systems.

Acknowledgments

We would like to thank Leiden University, University of Amsterdam, University of Illinois at Urbana-Champaign, the Dutch National Science Foundation (NWO), and the BSIK/BRICKS research funding programs for their support of our work.

References

1. Cohen, I., Sebe, N., Garg, A., Lew, M.S., Huang, T.S.: Facial expression recognition from video sequences. In: ICME. Proceedings of the IEEE International Conference Multimedia and Expo, Lausanne, Switzerland, vol. I, pp. 641–644. IEEE Computer Society Press, Los Alamitos (2002)

2. Lew, M.S.: Information theoretic view-based and modular face detection. In: Proceedings of the IEEE Face and Gesture Recognition conference, Killington, VT, pp. 198–203. IEEE Computer Society Press, Los Alamitos (1996)
3. Valstar, M.F., Pantic, M.: Combined Support Vector Machines and Hidden Markov Models for Modeling Facial Action Temporal Dynamics. In: Lew, M., Sebe, N., Huang, T.S., Bakker, E.M. (eds.) HCI 2007. LNCS, vol. 4796, pp. 118–127. Springer, Heidelberg (2007)
4. Chang, C.-C., Wu, C., Aghajan, H.: Pose and Gaze Estimation in Multi-Camera Networks for Non-Restrictive HCI. In: Lew, M., Sebe, N., Huang, T.S., Bakker, E.M. (eds.) HCI 2007. LNCS, vol. 4796, pp. 128–137. Springer, Heidelberg (2007)
5. Oerlemans, A., Thomee, B.: Interactive Feedback for Video Tracking Using Hybird Maximum Likelihood Similarity Measure. In: Lew, M., Sebe, N., Huang, T.S., Bakker, E.M. (eds.) HCI 2007. LNCS, vol. 4796, pp. 79–87. Springer, Heidelberg (2007)
6. Cooper, H., Bowden, R.: Large Lexicon Detection of Sign Language. In: Lew, M., Sebe, N., Huang, T.S., Bakker, E.M. (eds.) HCI 2007. LNCS, vol. 4796, pp. 88–97. Springer, Heidelberg (2007)
7. Sebe, N., Lew, M.S., Huijsmans, N.: Toward Improved Ranking Metrics. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1132–1143 (October 2000)
8. Angelopoulou, A., Psarrou, A., Gupta, G., Garcia Rodríguez, J.: Nonparametric Modelling and Tracking with Active-GNG. In: Lew, M., Sebe, N., Huang, T.S., Bakker, E.M. (eds.) HCI 2007. LNCS, vol. 4796, pp. 98–107. Springer, Heidelberg (2007)
9. Jung, S., Guo, Y., Sawhney, H., Kumar, R.: Multiple Cue Integrated Action Detection. In: Lew, M., Sebe, N., Huang, T.S., Bakker, E.M. (eds.) HCI 2007. LNCS, vol. 4796, pp. 108–117. Springer, Heidelberg (2007)
10. Chu, C.-W., Nevatia, R.: Real Time Body Pose Tracking in an Immersive Training Environment. In: Lew, M., Sebe, N., Huang, T.S., Bakker, E.M. (eds.) HCI 2007. LNCS, vol. 4796, pp. 146–156. Springer, Heidelberg (2007)
11. State, A.: Exact Eye Contact with Virtual Humans. In: Lew, M., Sebe, N., Huang, T.S., Bakker, E.M. (eds.) HCI 2007. LNCS, vol. 4796, pp. 138–145. Springer, Heidelberg (2007)
12. Rajko, S., Qian, G.: Real-time Automatic Kinematic Model Building for Optical Motion Capture Using a Markov Random Field. In: Lew, M., Sebe, N., Huang, T.S., Bakker, E.M. (eds.) HCI 2007. LNCS, vol. 4796, pp. 69–78. Springer, Heidelberg (2007)
13. Vural, E., Cetin, M., Ercil, A., Littlewort, G., Bartlett, M., Movellan, J.: Drowsy Driver Detection Through Facial Movement Analysis. In: Lew, M., Sebe, N., Huang, T.S., Bakker, E.M. (eds.) HCI 2007. LNCS, vol. 4796, pp. 6–18. Springer, Heidelberg (2007)
14. Thomee, B., Huiskes, M.J., Bakker, E., Lew, M.: An Artificial Imagination for Interactive Search. In: Lew, M., Sebe, N., Huang, T.S., Bakker, E.M. (eds.) HCI 2007. LNCS, vol. 4796, pp. 19–28. Springer, Heidelberg (2007)
15. Barreto, A., Zhai, J., Adjouadi, M.: Non-intrusive Physiological Monitoring for Automated Stress Detection in Human-Computer Interaction. In: Lew, M., Sebe, N., Huang, T.S., Bakker, E.M. (eds.) HCI 2007. LNCS, vol. 4796, pp. 29–38. Springer, Heidelberg (2007)
16. Hua, G., Yang, T.-Y., Vasireddy, S.: PEYE: Toward a Visual Motion based Perceptual Interface for Mobile Devices. In: Lew, M., Sebe, N., Huang, T.S., Bakker, E.M. (eds.) HCI 2007. LNCS, vol. 4796, pp. 39–48. Springer, Heidelberg (2007)
17. Song, P., Winkler, S., Gilani, S.O.: Vision-based Projected Tabletop Interface for Finger Interactions. In: Lew, M., Sebe, N., Huang, T.S., Bakker, E.M. (eds.) HCI 2007. LNCS, vol. 4796, pp. 49–58. Springer, Heidelberg (2007)

18. Stødle, D., Bjørndalen, J., Anshus, O.: A System for Hybrid Vision- and Sound-Based Interaction with Distal and Proximal Targets on Wall-Sized, High-Resolution Tiled Displays. In: Lew, M., Sebe, N., Huang, T.S., Bakker, E.M. (eds.) HCI 2007. LNCS, vol. 4796, pp. 59–68. Springer, Heidelberg (2007)
19. Sebe, N., Lew, M.S., Cohen, I., Sun, Y., Gevers, T., Huang, T.S.: Authentic Facial Expression Analysis. In: Proceedings of the International Conference on Automatic Face and Gesture Recognition, Seoul, Korea, pp. 517–522 (May 2004)
20. Lew, M.S., Sebe, N., Djeraba, C., Jain, R.: Content-based Multimedia Information Retrieval: State-of-the-art and Challenges. ACM Transactions on Multimedia Computing, Communication, and Applications 2(1), 1–19 (2006)
21. Sebe, N., Lew, M.S.: Wavelet Based Texture Classification. In: Proceedings of the 15th International Conference on Pattern Recognition, Barcelona, Spain, vol. III, pp. 959–962 (2000)
22. Sidenbladh, H., Black, M.J., Sigal, L.: Implicit Probabilistic Models of Human Motion for Synthesis and Tracking. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) ECCV 2002. LNCS, vol. 2352, pp. 784–800. Springer, Heidelberg (2002)
23. Pantic, M., Pentland, A., Nijholt, A., Huang, T.S.: Human Computing and Machine Understanding of Human Behavior: A Survey. In: ICMI 2006 and IJCAI International Workshops, Banff, Canada, Hyderabad, India, November 3, 2006, January 6, 2007. LNCS, vol. 4451, pp. 47–71. Springer, Heidelberg (2007)