The State-of-the-Art in Human-Computer Interaction

Nicu Sebe[1], Michael S. Lew[2], Thomas S. Huang[3]

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

Human computer interaction (HCI) lies at the crossroads of many scientific areas including artificial intelligence, computer vision, face recognition, motion tracking, etc. In recent years there has been a growing interest in improving all aspects of the interaction between humans and computers. It is argued that to truly achieve effective human-computer intelligent interaction (HCII), there is a need for the computer to be able to interact naturally with the user, similar to the way human-human interaction takes place.

Humans interact with each other mainly through speech, but also through body gestures, to emphasize a certain part of the speech and display of emotions. As a consequence, the new interface technologies are steadily driving toward accommodating information exchanges via the natural sensory modes of sight, sound, and touch. In face-to-face exchange, humans employ these communication paths simultaneously and in combination, using one to complement and enhance another. The exchanged information is largely encapsulated in this natural, multimodal format. Typically, conversational interaction bears a central burden in human communication, with vision, gaze, expression, and manual gesture often contributing critically, as well as frequently embellishing attributes such as emotion, mood, attitude, and attentiveness. But the roles of multiple modalities and their interplay remain to be quantified and scientifically understood. What is needed is a science of human-computer communication that establishes a framework for multimodal ”language” and ”dialog”, much like the framework we have evolved for spoken exchange.

Another important aspect is the development of Human-Centered Information Systems. The most important issue here is how to achieve synergism between man and machine. The term ”Human-Centered” 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. What can the scientific/engineering community do to effect a change for the better?

Information systems are ubiquitous in all human endeavors including scientific, medical, military, transportation, and consumer. Individual users use them for learning, searching for 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. An information system consists of two components:

Computer (data/knowledge base, and information processing engine), and humans. It is the intelligent interaction between the two that we are addressing. We aim to identify the important research issues, and to ascertain potentially fruitful future research directions. Furthermore, we shall discuss how an environment can be created which is conducive to carrying out such research.

In many important HCI applications such as computer aided tutoring and learning, it is highly desirable (even mandatory) that the response of the computer take into account the emotional or cognitive state of the human user. Emotions are displayed by visual, vocal, and other physiological means. There is a growing amount of evidence showing that emotional skills are part of what is called "intelligence" [1, 2]. Computers today can recognize much of what is said, and to some extent, who said it. But, they are almost completely in the dark when it comes to how things are said, the affective channel of information. This is true not only in speech, but also in visual communications despite the fact that facial expressions, posture, and gesture communicate some of the most critical information: how people feel. Affective communication explicitly considers how emotions can be recognized and expressed during human-computer interaction.

In most cases today, if you take a human-human interaction, and replace one of the humans with a computer, then the affective communication vanishes. Furthermore, it is not because people stop communicating affect - certainly we have all seen a person expressing anger at his machine. The problem arises because the computer has no ability to recognize if the human is pleased, annoyed, interested, or bored. Note that if a human ignored this information, and continued babbling long after we had yawned, we would not consider that person very intelligent. Recognition of emotion is a key component of intelligence. Computers are presently affect-impaired.

Furthermore, if you insert a computer (as a channel of communication) between two or more humans, then the affective bandwidth may be greatly reduced. Email may be the most frequently used means of electronic communication, but typically all of the emotional information is lost when our thoughts are converted to the digital media.

Research is therefore needed for new ways to communicate affect through computer-mediated environments. Computer-mediated communication today almost always has less affective bandwidth than "being there, face-to-face". The advent of affective wearable computers, which could help amplify affective information as perceived from a person's physiological state, are but one possibility for changing the nature of communication.

The papers in the proceedings present specific aspects of the technologies that support human-computer interaction. Most of the authors are computer vision researchers whose work is related to human-computer interaction.

The paper by Warwick and Gasson [3] discusses the efficacy of a direct connection between the human nervous system and a computer network. The authors give an overview of the present state of neural implants and discuss the possibilities regarding such implant technology as a general purpose human-computer interface for the future.

Human-robot interaction (HRI) has recently drawn increased attention. Autonomous mobile robots can recognize and track a user, understand his verbal commands, and take actions to serve him. A major reason that makes HRI distinctive from traditional

HCI is that robots can not only passively receive information from environment but also make decisions and actively change the environment. An interesting approach in this direction is presented by Huang and Weng [4]. Their paper presents a motivational system for HRI which integrates novelty and reinforcement learning. The robot develops its motivational system through its interactions with the world and the trainers. A vision-based gestural guidance interface for mobile robotic platforms is presented by Paquin and Cohen [5]. The interface controls the motion of the robot by using a set of predefined static and dynamic hand gestures inspired by the marshaling code. Images captured by an on-board camera are processed in order to track the operator's hand and head. A similar approach is taken by Nickel and Stiefelhagen [6]. Given the images provided by a calibrated stereo-camera, color and disparity information are integrated into a multi-hypotheses tracking framework in order to find the 3D positions of the respective body parts. Based on the motion of the hands, an HMM-based approach is applied to recognize pointing gestures.

Mixed reality (MR) opens a new direction for human-computer interaction. Combined with computer vision techniques, it is possible to create advanced input devices. Such a device is presented by Tosas and Li [7]. They describe a virtual keypad application which illustrates the virtual touch screen interface idea. Visual tracking and interpretation of the user's hand and finger motion allows the detection of key presses on the virtual touch screen. An interface tailored to create a design-oriented realistic MR workspace is presented by Gheorghe, et al. [8]. An augmented reality human computer interface for object localization is presented by Siegl, et al. [9]. A 3D pointing interface that can perform 3D recognition of arm pointing direction is proposed by Hosoya, et al. [10]. A hand gesture recognition system is also proposed by Licsár and Szirányi [11]. A hand pose estimation approach is discussed by Stenger, et al. [12]. They present an analysis of the design of classifiers for use in a more general hierarchical object recognition approach.

The current down-sizing of computers and sensory devices allows humans to wear these devices in a manner similar to clothes. One major direction of wearable computing research is to smartly assist humans in daily life. Yamazoe, et al. [13] propose a body attached system to capture audio and visual information corresponding to user experience. This data contains significant information for recording/analyzing human activities and can be used in a wide range of applications such as digital diary or interaction analysis. Another wearable system is presented by Tsukizawa, et al. [14].

3D head tracking in a video sequence has been recognized as an essential prerequisite for robust facial expression/emotion analysis, face recognition and model-based coding. The paper by Dornaika and Ahlberg [15] presents a system for real-time tracking of head and facial motion using 3D deformable models. A similar system is presented by Sun, et al [16]. Their goal is to use their real-time tracking system to recognize authentic facial expressions. A pose invariant face recognition approach is proposed by Lee and kim [17]. A 3D head pose estimation approach is proposed by Wang, et al [18]. They present a new method for computing the head pose by using projective invariance of the vanishing point. A multi-view face image synthesis using a factorization model is introduced by Du and Lin [19]. The proposed method can be applied to a several

HCI areas such as view independent face recognition or face animation in a virtual environment.

The emerging idea of ambient intelligence is a new trend in human-computer interaction. An ambient intelligence environment is sensitive to the presence of people and responsive to their needs. The environment will be capable of greeting us when we get home, of judging our mood and adjusting our environment to reflect it. Such an environment is still a vision but it is one that struck a chord in the minds of researchers around the world and become the subject of several major industry initiatives. One such initiative is presented by Kleindienst, et al. [20]. They use speech recognition and computer vision to model new generation of interfaces in the residential environment. An important part of such a system is the localization module. A possible implementation of this module is proposed by Okatani and Takuichi [21]. Another important part of an ambient intelligent system is the extraction of typical actions performed by the user. A solution to this problem is provided by Ma and Lin [22].

Human-computer interaction is a particularly wide area which involves elements from diverse areas such as psychology, ergonomics, engineering, artificial intelligence, databases, etc. This proceedings represents a snapshot of the state of the art in human computer interaction with an emphasis on intelligent interaction via computer vision, artificial intelligence, and pattern recognition methodology. Our hope is that in the not too distant future the research community will have made significant strides in the science of human-computer interaction, and that new paradigms will emerge which will result in natural interaction between humans, computers, and the environment.

References

1. Salovey, P., Mayer, J.: Emotional intelligence. Imagination, Cognition, and Personality **9** (1990) 185–211
2. Goleman, D.: Emotional Intelligence. Bantam Books, New York (1995)
3. Warwick, K., Gasson, M.: Practical interface experiments with implant technology. In: International Workshop on Human-Computer Interaction, Lecture Notes in Computer Science, vol. 3058, Springer (2004) 6–16
4. Huang, X., Weng, J.: Motivational system for human-robot interaction. In: International Workshop on Human-Computer Interaction, Lecture Notes in Computer Science, vol. 3058, Springer (2004) 17–27
5. Paquin, V., Cohen, P.: A vision-based gestural guidance interface for mobile robotic platforms. In: International Workshop on Human-Computer Interaction, Lecture Notes in Computer Science, vol. 3058, Springer (2004) 38–46
6. Nickel, K., Stiefelhagen, R.: Real-time person tracking and pointing gesture recognition for human-robot interaction. In: International Workshop on Human-Computer Interaction, Lecture Notes in Computer Science, vol. 3058, Springer (2004) 28–37
7. Tosas, M., Li, B.: Virtual touch screen for mixed reality. In: International Workshop on Human-Computer Interaction, Lecture Notes in Computer Science, vol. 3058, Springer (2004) 47–57
8. Gheorghe, L., Ban, Y., Uehara, K.: Exploring interactions specific to mixed reality 3D modeling systems. In: International Workshop on Human-Computer Interaction, Lecture Notes in Computer Science, vol. 3058, Springer (2004) 113–123

9. Siegl, H., Schweighofer, G., Pinz, A.: An AR human computer interface for object localization in a cognitive vision framework. In: International Workshop on Human-Computer Interaction, Lecture Notes in Computer Science, vol. 3058, Springer (2004) 167–177
10. Hosoya, E., Sato, H., Kitabata, M., Harada, I., Nojima, H., Onozawa, A.: Arm-pointer: 3D pointing interface for real-world interaction. In: International Workshop on Human-Computer Interaction, Lecture Notes in Computer Science, vol. 3058, Springer (2004) 70–80
11. Licsár, A., Szirányi, T.: Hand gesture recognition in camera-projector system. In: International Workshop on Human-Computer Interaction, Lecture Notes in Computer Science, vol. 3058, Springer (2004) 81–91
12. Stenger, B., Thayananthan, A., Torr, P., Cipolla, R.: Hand pose estimation using hierarchical detection. In: International Workshop on Human-Computer Interaction, Lecture Notes in Computer Science, vol. 3058, Springer (2004) 102–112
13. Yamazoe, H., Utsumi, A., Tetsutani, N., Yachida, M.: A novel wearable system for capturing user view images. In: International Workshop on Human-Computer Interaction, Lecture Notes in Computer Science, vol. 3058, Springer (2004) 156–166
14. Tsukizawa, S., Sumi, K., Matsuyama, T.: 3D digitization of a hand-held object with a wearable vision sensor. In: International Workshop on Human-Computer Interaction, Lecture Notes in Computer Science, vol. 3058, Springer (2004) 124–134
15. Dornaika, F., Ahlberg, J.: Model-based head and facial motion tracking. In: International Workshop on Human-Computer Interaction, Lecture Notes in Computer Science, vol. 3058, Springer (2004) 211–221
16. Sun, Y., Sebe, N., Lew, M., Gevers, T.: Authentic emotion detection in real-time video. In: International Workshop on Human-Computer Interaction, Lecture Notes in Computer Science, vol. 3058, Springer (2004) 92–101
17. Lee, H.S., Kim, D.: Pose invariant face recognition using linear pose transformation in feature space. In: International Workshop on Human-Computer Interaction, Lecture Notes in Computer Science, vol. 3058, Springer (2004) 200–210
18. Wang, J.G., Sung, E., Venkateswarlu, R.: EM enhancement of 3D head pose estimated by perspective invariance. In: International Workshop on Human-Computer Interaction, Lecture Notes in Computer Science, vol. 3058, Springer (2004) 178–188
19. Du, Y., Lin, X.: Multi-view face image synthesis using factorization model. In: International Workshop on Human-Computer Interaction, Lecture Notes in Computer Science, vol. 3058, Springer (2004) 189–199
20. Kleindienst, J., Macek, T., Serédi, L., Šedivý, J.: Djinn: Interaction framework for home environment using speech and vision. In: International Workshop on Human-Computer Interaction, Lecture Notes in Computer Science, vol. 3058, Springer (2004) 145–155
21. Okatani, I., Takuichi, N.: Location-based information support system using multiple cameras and LED light sources with the compact battery-less information terminal (CoBIT). In: International Workshop on Human-Computer Interaction, Lecture Notes in Computer Science, vol. 3058, Springer (2004) 135–144
22. Ma, G., Lin, X.: Typical sequences extraction and recognition. In: International Workshop on Human-Computer Interaction, Lecture Notes in Computer Science, vol. 3058, Springer (2004) 58–69