Available online at www.sciencedirect.com
ScienceDirect
Image and Vision Computing 25 (2007) 1833–1835
www.elsevier.com/locate/imavis

Editorial

The age of human computer interaction

Historians will refer to our time as the Age of Information. While information is indeed important, we think that the creative exchange of information, the interactive process between humans and computers will be the foremost challenge of the 21st century. Human computer interaction (HCI) lies at the crossroads of many academic areas including artificial intelligence, psychology, computer vision, literature, philosophy, mathematics, and sociology. It is at the curious intersections between these areas where the true advancements in HCI lie and someday may lead to the ultimate goal of intelligent and intuitive HCI.

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.

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. Researchers argue that users will not be merely externally using applications, but actually living in applications. An IBM research initiative is presented by Kleindienst et al. [3]. They use speech recognition and computer vision to model a new generation of interfaces in the residential environment. Important goals are also to be open, flexible, modular, and provide distant access. At the ATR Media Information Sciences laboratories, Yamazoe et al. [9] describe a new system for wearing the computing system. They 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. The current down-sizing of computers and sensory devices allows humans to wear these devices in a manner similar to

0262-8856/$ - see front matter
doi:10.1016/j.imavis.2007.06.007

clothes. One major direction of their research is to intelligently assist humans in daily life.

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 frequently arises because the computer has no ability to recognize if the human is pleased, annoyed, interested, or bored. Recognition of emotion is a key component of intelligence. The advent of affective-aware computers are but one possibility for changing the nature of communication. Sebe et al. [5] introduce the problem of authentic facial emotion recognition, create a new authentic emotion test set, and adapt several learning algorithms to perform automatic emotion detection. The premise behind authentic emotion detection is that most computer vision test sets were created using "acted" or "faked" expressions such as "please smile for the camera." For many humans the "acted" smile may be quite different than the one corresponding to an authentic internal emotion of happiness. Therefore, "authentic" test sets are needed if we are to perform representative affective algorithm comparisons.

Understanding the situation of a human through visual analysis is one of the core technologies in intuitive HCI methods. Ideally, we would like the computer to understand the user in the same way that another human would, simply via visual imagery. We are working on the fundamental science of visually understanding humans through extracting information automatically about the head, hands, and other body. Visual information extraction of the head pose is described by Wang and Sung [8]. They present a new method for computing the head pose by using projective invariance of the vanishing point. In addition to the information about the head orientation, they also extract information about the dominant facial features such as the eyes and mouth. The information contained in the head is not limited to merely the locations of facial features but also reveals meaningful information about the emotional state of the person.

Beyond automatic detection of the head, it is certain that the hands convey significant information in daily interactions. In the work by Nickel and Stiefelhagen [4], 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. A hand pose estimation approach is discussed by Stenger et al. [7]. Specifically, they are able to segment the hand and extract the pose information from cluttered scenes using a general hierarchical object recognition approach. Further, they present an analysis several pertinent classifiers in the context of hand pose detection. An augmented reality human computer interface for object localization is presented by Siegl et al. [6]. The overall goal of their work is to facilitate 3D environment interaction by going beyond traditional 2D tools. Their method promotes and implements the notion of a 3D cursor which is guided by pointing within a stereo camera visual analysis system.

Human–computer interaction is a particularly wide area which involves elements from diverse areas. The papers within this special issue represent a snapshot of the state of the art in human computer interaction with an emphasis on intelligent interaction via computer vision and artificial intelligence. Our hope is that in the next decade 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.

Acknowledgements

We would like to thank Leiden University, University of Amsterdam, the Dutch National Science Foundation (NWO), the BSIK/BRICKS/FOCUS program, and the University of Illinois at Urbana-Champaign for thier support.

References

[1] P. Salovey, J. Mayer, Emotional intelligence, Imagination, Cognition, and Personality 9 (1990) 185–211.
[2] D. Goleman, Emotional Intelligence, Bantam Books, New York, 1995.
[3] Kleindienst, Macek, Seredi, Sedivy, Interaction framework for home environment using speech and vision .
[4] Nickel, Stiefelhagen, Visual recognition of pointing gestures for human–robot interaction.
[5] Sebe, Lew, Sun, Cohen, Gevers, Huang, Authentic facial expression analysis.
[6] Siegl, Hanheide, Wrede, Pinz, An augmented reality human computer interface for object localization in a cognitive vision system.
[7] Stenger, Thayananthan, Torr, Cipolla, Estimating 3D hand pose using hierarchical multi-label classification.
[8] Wang, Sung, EM enhancement of 3D head pose estimated by point at infinity.
[9] Yamazoe, Utsumi, Hosaka, Yachida, A body-mounted camera system for head pose estimation and user-view image synthesis.

M. Lew*
Leiden University
E-mail address: mlew@liacs.nl

N. Sebe
University of Amsterdam

T.S. Huang
University of Illinois at Urbana-Champaign

* Corresponding author.