Guest Editorial: Human–Computer Interaction: Real-Time Vision Aspects of Natural User Interfaces

Zoran Živković · Nicu Sebe · Hamid Aghajan ·
Branislav Kisačanin

Published online: 20 December 2012
© Springer Science+Business Media New York 2012

Computer vision for human–computer interaction has been in our living rooms for a few years now. We have witnessed an exponential growth in capability of vision systems for computer games, from Sony’s EyeToy for PlayStation 2 in 2003, Eye for PlayStation 3 in 2007, to Microsoft’s Kinect for XBOX 360 in 2010. In addition to enhancing our interaction with computer games, Kinect has had a deeper role in changing our expectations for human–machine interfaces of the future. We may need to wait a few more years to have computers like those in the science fiction movie *Minority Report*, but the games and hacks using Kinect show us that such interfaces are getting closer to reality. We now expect our TVs, smart phones, and tablets to have such capability and soon that will be a reality.

To that end, the research in both academia and industry continues to make great strides, as demonstrated by the presentations at the Workshop on Human–Computer Interaction: Real-Time Vision Aspects of Natural User Interfaces, which the present guest editors organized at the 2011 ICCV in Barcelona. After the workshop was over, we had an open call for papers for authors to write journal articles on their work related to this topic, and the result of this effort is this Special Issue of the IJCV. The papers presented in this special issue

span a wide spectrum of problems in HCI, from hand gesture and body pose recognition, 3D face analysis, tracking of eyes and body poses, to incremental learning and 3D displays that track the user and adjust their content to achieve the depth sensation. The progress in each of these and other related problems is essential for future human–computer interfaces.

The following are summaries of the papers contained in this special issue.

- K. P. Pramod, P. Vadakkepat, and A. P. Loh addressed the complex background problem in hand posture recognition in their paper “Attention Based Detection and Recognition of Hand Postures Against Complex Backgrounds.” A Bayesian model of attention is utilized to detect and segment hand postures by considering the shape, texture and color features. The algorithm provided reliable performance against cluttered natural environments including skin colored complex backgrounds. The paper also contributed a new complex background hand posture dataset shot with 40 subjects, having various ethnic origins.
- C. Ellis, S. Z. Masood, M. F. Tappen, J. J. LaViola Jr., and R. Sukthankar examine intertwined issues of latency and recognition accuracy in their paper “Exploring the Trade-off Between Accuracy and Observational Latency in Action Recognition.” Focusing on the problem of recognizing actions from 3D skeletal observations, this paper proposes a recognition framework that can optimize both the accuracy of the recognition system and the number of frames that must be viewed before the action can be recognized.
- In their paper “Random forests for real time 3D face analysis,” G. Fanelli, M. Dantone, J. Gall, A. Fossati, and L. Van Gool present a framework for head pose

Z. Živković (✉)
NXP Semiconductors, Eindhoven, The Netherlands
e-mail: zoran.z.zivkovic@googlemail.com

N. Sebe
University of Trento, Trento, Italy

H. Aghajan
Stanford university, Stanford, USA and University of Ghent,
Ghent, Belgium

B. Kisačanin
Texas Instruments, Dallas, USA

estimation and facial features detection from depth data of varying quality. The algorithm uses random forests to directly map image patches into probabilistic votes in the desired continuous output spaces, achieving robust and accurate estimates on challenging datasets, including a novel annotated head pose database acquired with a Kinect.

- When introduced, the cross-ratio (CR) method for remote eye gaze tracking offered many attractive features including no calibration and freedom of head movement. However, due to strong simplifying assumptions, current CR-based methods still require calibration and are also sensitive to large head movements. In “Improving Head Movement Tolerance of CR Based Eye Trackers” Flavio L. Coutinho and Carlos H. Morimoto present and evaluate two new CR-based methods that successfully improve the robustness of the method to head motion.
- N. Lehment, M. Kaiser, and G. Rigoll examine the problem of finding accurate and robust observation likelihood approximations for stochastic human pose tracking in their paper “Using Segmented 3D Point Clouds for Accurate Likelihood Approximation in Human Pose Tracking.” The proposed approach uses Euclidean distances between segmented 3D point clouds to assign likelihood scores to hypothetical human body poses. The segmentation of data is shown to improve accuracy while enabling an efficient and simple integration of collision and occlusion handling.
- Many HCI systems could benefit from a fast incremental learning framework. The paper, “Euler Principal Component Analysis,” by S. Liwicki, G. Tzimiropoulos, S. Zafeiriou, and M. Pantic introduces a kernel principal component analysis (PCA) algorithm which combines efficiency with robustness. The authors substitute the standard error measure of PCA (i.e., ℓ_2 norm) with a robust dissimilarity that is fast to compute and retains standard PCA’s favorable properties. Their extensive evaluation shows promising results for face reconstruction, object tracking and background modeling.
- C. Malleson and J. Collomosse propose a new form of 3D display in “Virtual Volumetric Graphics on Commodity Displays Using 3D Viewer Tracking.” Many 3D displays rely upon stereo disparity. The proposed display enables interactive exploration of 3D scenes by varying a virtual camera position according to the tracked 3D position of the user. No wearable or specialized hardware is required; only a 2D commodity display and a pair of web cameras. The authors describe the calibration, tracking and visualization aspects of the system; quantitatively evaluating accuracy and latency, and reporting user experiments that trial the display.

We would like to express our sincere gratitude to all of the authors for their work and patience with the review process, to the reviewers for their valuable comments and feedback, and to the editorial staff at Springer for their expedient handling of this special issue.