

Editorial

Special Section on Multimodal Interfaces and Applications

THE USE OF a keyboard and mouse as human computer interfaces has been the traditional method of text and data entry. However, there has been a growing interest in multimodal methods in human-computer interactions in order to achieve more natural, usable interfaces to technology. The emphasis has been on acquisition, integration and recognition of different modes of information; audio and video, speech and gesture, etc. Combining these modes for robust, reliable applications has been of interest to many areas of computing.

In order to draw attention to this area, a special session, Multimodal Interfaces and Applications, was organized at ICME2001 (IEEE International Conference on Multimedia and Expo) in Tokyo, Japan, in August 2001. Based on this special session, a series of high level submissions was requested for a special section in this journal.

As a result of the reviewing process, we finally decided to accept two high-quality papers for this special section. These papers present unique methods for integrating multiple information cues in video and audio analyses. M. Yeasin *et al.* focus on tracking body parts of multiple people in an unconstrained environment. They developed an MHT (multiple hypothesis tracking) based algorithm that integrates color and motion information so that the body parts are accurately be located regardless of noisy false detections of trajectories. J. W. Fisher and T. Darrell apply a signal level fusion technique to detecting audio and visual signals arriving from a common source. They introduced a probabilistic multimodal generation model for deriving an information theoretic measure of cross-modal correspondence, where non-parametric statistical density modelling techniques can characterize the mutual information between signals from different domains. By comparing the

mutual information between audio and visual signals, a person who is speaking a given utterance can successfully be identified regardless of other persons' different utterances or non-speech events.

This area is a multidisciplinary area and continued advances in this area will be dominated by teams working from perspectives. While we received many submissions, we hope that the authors of unsuccessful submissions will continue to research in this area and consider submitting their results to the IEEE TRANSACTIONS ON MULTIMEDIA (TMM) in the future. We would like to thank all the reviewers for their comments, which significantly contribute to the quality of this special section. We also thank Dr. Tsuhan Chen, the Editor-in-Chief of IEEE TMM. Dr. Chen gave us many timely, reasonable suggestions and comments. Without his encouragement to us, we would not be able to complete the organizing process of this special section. Finally, we thank Deborah L. Tomaro, IEEE Signal Processing Society Publications Office and Jo-Ellen B. Snyder, Publications Manager, IEEE Signal Processing Society, for their jobs for assisting us in handling the manuscripts, review results and communications.

JUN OHYA, *Guest Editor*
Waseda University
Tokyo, Japan
ohya@waseda.jp

RICHARD REILLY, *Guest Editor*
University College Dublin
Dublin, Ireland

CHUNG-SHENG LI, *Guest Editor*
IBM T. J. Watson Research Center
Hawthorne, NY 10532 USA