Lecture Notes in Computer Science 3869

Commenced Publication in 1973
Founding and Former Series Editors:
Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

Editorial Board

David Hutchison
Lancaster University, UK

Takeo Kanade
Carnegie Mellon University, Pittsburgh, PA, USA

Josef Kittler
University of Surrey, Guildford, UK

Jon M. Kleinberg
Cornell University, Ithaca, NY, USA

Friedemann Mattern
ETH Zurich, Switzerland

John C. Mitchell
Stanford University, CA, USA

Moni Naor
Weizmann Institute of Science, Rehovot, Israel

Oscar Nierstrasz
University of Bern, Switzerland

C. Pandu Rangan
Indian Institute of Technology, Madras, India

Bernhard Steffen
University of Dortmund, Germany

Madhu Sudan
Massachusetts Institute of Technology, MA, USA

Demetri Terzopoulos
New York University, NY, USA

Doug Tygar
University of California, Berkeley, CA, USA

Moshe Y. Vardi
Rice University, Houston, TX, USA

Gerhard Weikum
Max-Planck Institute of Computer Science, Saarbruecken, Germany

Steve Renals Samy Bengio (Eds.)

Machine Learning for Multimodal Interaction

Second International Workshop, MLMI 2005
Edinburgh, UK, July 11-13, 2005
Revised Selected Papers

Springer

Volume Editors

Steve Renals
University of Edinburgh, Centre for Speech Technology Research
2 Buccleuch Place, Edinburgh EH8 9LW, UK
E-mail: s.renals@ed.ac.uk

Samy Bengio
IDIAP Research Institute
Rue du Simplon 4, Case Postale 592, 1920 Martigny, Switzerland
E-mail: bengio@idiap.ch

Library of Congress Control Number: 2006920577

CR Subject Classification (1998): H.5.2-3, H.5, I.2.6, I.2.10, I.2, I.7, K.4, I.4

LNCS Sublibrary: SL 3 – Information Systems and Application, incl. Internet/Web and HCI

ISSN 0302-9743
ISBN-10 3-540-32549-2 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-32549-9 Springer Berlin Heidelberg New York

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.

Springer is a part of Springer Science+Business Media

springer.com

© Springer-Verlag Berlin Heidelberg 2006
Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 11677482 06/3142 5 4 3 2 1 0

Preface

This book contains a selection of refereed papers presented at the Second Workshop on Machine Learning for Multimodal Interaction (MLMI 2005), held in Edinburgh, Scotland, during 11–13 July 2005.

The workshop was organized and sponsored jointly by two European integrated projects, three European Networks of Excellence and a Swiss national research network:

- AMI, Augmented Multiparty Interaction, http://www.amiproject.org/
- CHIL, Computers in the Human Interaction Loop, http://chil.server.de/
- HUMAINE, Human–Machine Interaction Network on Emotion, http://emotion-research.net/
- PASCAL, Pattern Analysis, Statistical Modeling and Computational Learning, http://www.pascal-network.org/
- SIMILAR, human–machine interfaces similar to human–human communication, http://www.similar.cc/
- IM2, Interactive Multimodal Information Management, http://www.im2.ch/

In addition to the main workshop, MLMI 2005 hosted the NIST (US National Institute of Standards and Technology) Meeting Recognition Workshop. This workshop (the third such sponsored by NIST) was centerd on the Rich Transcription 2005 Spring Meeting Recognition (RT-05) evaluation of speech technologies within the meeting domain. Building on the success of the RT-04 spring evaluation, the RT-05 evaluation continued the speech-to-text and speaker diarization evaluation tasks and added two new evaluation tasks: speech activity detection and source localization.

MLMI 2005 was thus sponsored by the European Commission (Information Society Technologies priority of the Sixth Framework Programme), the Swiss National Science Foundation and the US National Institute of Standards and Technology.

Given the multiple links between the above projects and several related research areas, and the success of the first MLMI 2004 workshop, it was decided to organize once again a joint workshop bringing together researchers from the different communities working around the common theme of advanced machine learning algorithms for processing and structuring multimodal human interaction. The motivation for creating such a forum, which could be perceived as a number of papers from different research disciplines, evolved from an actual need that arose from these projects and the strong motivation of their partners for such a multidisciplinary workshop. This assessment was confirmed this year by a significant increase in the number of sponsoring research projects, and by the success of the workshop itself, which attracted about 170 participants.

The conference program featured invited talks, full papers (subject to careful peer review, by at least three reviewers), and posters (accepted on the basis of

abstracts) covering a wide range of areas related to machine learning applied to multimodal interaction — and more specifically to multimodal meeting processing, as addressed by the various sponsoring projects. These areas included:

- Human–human communication modeling
- Speech and visual processing
- Multimodal processing, fusion and fission
- Multimodal dialog modeling
- Human–human interaction modeling
- Multimodal data structuring and presentation
- Multimedia indexing and retrieval
- Meeting structure analysis
- Meeting summarizing
- Multimodal meeting annotation
- Machine learning applied to the above

Out of the submitted full papers, about 50% were accepted for publication in the present volume, after having been invited to take review comments and conference feedback into account.

In the present book, and following the structure of the workshop, the papers are divided into the following sections:

1. Invited Papers
2. Multimodal Processing
3. HCI and Applications
4. Discourse and Dialog
5. Emotion
6. Visual Processing
7. Speech and Audio Processing
8. NIST Meeting Recognition Evaluation

Based on the successes of MLMI 2004 and MLMI 2005, it was decided to organize MLMI 2006 in the USA, in collaboration with NIST (US National Institute of Standards and Technology), again in conjunction with the NIST meeting recognition evaluation.

Finally, we take this opportunity to thank our Program Committee members, the sponsoring projects and funding agencies, and those responsible for the excellent management and organization of the workshop and the follow-up details resulting in the present book.

November 2005

Steve Renals
Samy Bengio

Organization

General Chairs

Steve Renals	University of Edinburgh
Samy Bengio	IDIAP Research Institute

Local Organization

Caroline Hastings	University of Edinburgh
Avril Heron	University of Edinburgh
Bartosz Dobrzelecki	University of Edinburgh
Jean Carletta	University of Edinburgh
Mike Lincoln	University of Edinburgh

Program Committee

Marc Al-Hames	Munich University of Technology
Tilman Becker	DFKI
Hervé Bourlard	IDIAP Research Institute
Jean Carletta	University of Edinburgh
Franciska de Jong	University of Twente
John Garofolo	NIST
Thomas Hain	University of Sheffield
Lori Lamel	LIMSI
Benoit Macq	UCL-TELE
Johanna Moore	University of Edinburgh
Laurence Nigay	CLIPS-IMAG
Barbara Peskin	ICSI
Thierry Pun	University of Geneva
Marc Schröder	DFKI
Rainer Stiefelhagen	Universitaet Karlsruhe

NIST Meeting Recognition Workshop Organization

Jon Fiscus	NIST
John Garofolo	NIST

Sponsoring Projects and Institutions

Projects:

- AMI, Augmented Multiparty Interaction, http://www.amiproject.org/
- CHIL, Computers in the Human Interaction Loop, http://chil.server.de/
- HUMAINE, Human–Machine Interaction Network on Emotion, http://emotion-research.net/
- SIMILAR, human–machine interfaces similar to human–human communication, http://www.similar.cc/
- PASCAL, Pattern Analysis, Statistical Modeling and Computational Learning, http://www.pascal-network.org/
- IM2, Interactive Multimodal Information Management, http://www.im2.ch/

Institutions :

- European Commission, through the Multimodal Interfaces objective of the Information Society Technologies (IST) priority of the Sixth Framework Programme.
- Swiss National Science Foundation, through the National Center of Competence in Research (NCCR) program.
- US National Institute of Standards and Technology (NIST), http://www.nist.gov/speech/

Table of Contents

I Invited Papers

Gesture, Gaze, and Ground
David McNeill 1

Toward Adaptive Information Fusion in Multimodal Systems
Xiao Huang, Sharon Oviatt 15

II Multimodal Processing

The AMI Meeting Corpus: A Pre-announcement
Jean Carletta, Simone Ashby, Sebastien Bourban, Mike Flynn, Mael Guillemot, Thomas Hain, Jaroslav Kadlec, Vasilis Karaiskos, Wessel Kraaij, Melissa Kronenthal, Guillaume Lathoud, Mike Lincoln, Agnes Lisowska, Iain McCowan, Wilfried Post, Dennis Reidsma, Pierre Wellner 28

VACE Multimodal Meeting Corpus
Lei Chen, R. Travis Rose, Ying Qiao, Irene Kimbara, Fey Parrill, Haleema Welji, Tony Xu Han, Jilin Tu, Zhongqiang Huang, Mary Harper, Francis Quek, Yingen Xiong, David McNeill, Ronald Tuttle, Thomas Huang 40

Multimodal Integration for Meeting Group Action Segmentation and Recognition
Marc Al-Hames, Alfred Dielmann, Daniel Gatica-Perez, Stephan Reiter, Steve Renals, Gerhard Rigoll, Dong Zhang 52

Detection and Resolution of References to Meeting Documents
Andrei Popescu-Belis, Denis Lalanne 64

Dominance Detection in Meetings Using Easily Obtainable Features
Rutger Rienks, Dirk Heylen 76

Can Chimeric Persons Be Used in Multimodal Biometric Authentication Experiments?
Norman Poh, Samy Bengio 87

III HCI and Applications

Analysing Meeting Records: An Ethnographic Study and Technological Implications
Steve Whittaker, Rachel Laban, Simon Tucker 101

Browsing Multimedia Archives Through Intra- and Multimodal Cross-Documents Links
Maurizio Rigamonti, Denis Lalanne, Florian Evéquoz, Rolf Ingold 114

The "FAME" Interactive Space
F. Metze, P. Gieselmann, H. Holzapfel, T. Kluge, I. Rogina, A. Waibel, M. Wölfel, J. Crowley, P. Reignier, D. Vaufreydaz, F. Bérard, B. Cohen, J. Coutaz, S. Rouillard, V. Arranz, M. Bertrán, H. Rodriguez 126

Development of Peripheral Feedback to Support Lectures
Janienke Sturm, Rahat Iqbal, Jacques Terken 138

Real-Time Feedback on Nonverbal Behaviour to Enhance Social Dynamics in Small Group Meetings
Olga Kulyk, Jimmy Wang, Jacques Terken 150

IV Discourse and Dialogue

A Multimodal Discourse Ontology for Meeting Understanding
John Niekrasz, Matthew Purver 162

Generic Dialogue Modeling for Multi-application Dialogue Systems
Trung H. Bui, Job Zwiers, Anton Nijholt, Mannes Poel 174

Toward Joint Segmentation and Classification of Dialog Acts in Multiparty Meetings
Matthias Zimmermann, Yang Liu, Elizabeth Shriberg, Andreas Stolcke 187

V Emotion

Developing a Consistent View on Emotion-Oriented Computing
Marc Schröder, Roddy Cowie 194

Multimodal Authoring Tool for Populating a Database of Emotional Reactive Animations
Alejandra García-Rojas, Mario Gutiérrez, Daniel Thalmann, Frédéric Vexo 206

VI Visual Processing

A Testing Methodology for Face Recognition Algorithms
Aristodemos Pnevmatikakis, Lazaros Polymenakos 218

Estimating the Lecturer's Head Pose in Seminar Scenarios - A Multi-view Approach
Michael Voit, Kai Nickel, Rainer Stiefelhagen 230

Foreground Regions Extraction and Characterization Towards Real-Time Object Tracking
José Luis Landabaso, Montse Pardàs 241

Projective Kalman Filter: Multiocular Tracking of 3D Locations Towards Scene Understanding
C. Canton-Ferrer, J.R. Casas, A.M. Tekalp, M. Pardàs 250

VII Speech and Audio Processing

Least Squares Filtering of Speech Signals for Robust ASR
Vivek Tyagi, Christian Wellekens 262

A Variable-Scale Piecewise Stationary Spectral Analysis Technique Applied to ASR
Vivek Tyagi, Christian Wellekens, Hervé Bourlard 274

Accent Classification for Speech Recognition
Arlo Faria 285

Hierarchical Multi-stream Posterior Based Speech Recognition System
Hamed Ketabdar, Hervé Bourlard, Samy Bengio 294

Variational Bayesian Methods for Audio Indexing
Fabio Valente, Christian Wellekens 307

Microphone Array Driven Speech Recognition: Influence of Localization on the Word Error Rate
Matthias Wölfel, Kai Nickel, John McDonough 320

Automatic Speech Recognition and Speech Activity Detection in the CHIL Smart Room
Stephen M. Chu, Etienne Marcheret, Gerasimos Potamianos 332

The Development of the AMI System for the Transcription of Speech in Meetings
Thomas Hain, Lukas Burget, John Dines, Iain McCowan, Giulia Garau, Martin Karafiat, Mike Lincoln, Darren Moore, Vincent Wan, Roeland Ordelman, Steve Renals 344

Improving the Performance of Acoustic Event Classification by Selecting and Combining Information Sources Using the Fuzzy Integral
Andrey Temko, Dušan Macho, Climent Nadeu 357

VIII NIST Meeting Recognition Evaluation

The Rich Transcription 2005 Spring Meeting Recognition Evaluation
Jonathan G. Fiscus, Nicolas Radde, John S. Garofolo, Audrey Le, Jerome Ajot, Christophe Laprun 369

Linguistic Resources for Meeting Speech Recognition
Meghan Lammie Glenn, Stephanie Strassel 390

Robust Speaker Segmentation for Meetings: The ICSI-SRI Spring 2005 Diarization System
Xavier Anguera, Chuck Wooters, Barbara Peskin, Mateu Aguiló 402

Speech Activity Detection on Multichannels of Meeting Recordings
Zhongqiang Huang, Mary P. Harper 415

NIST RT'05S Evaluation: Pre-processing Techniques and Speaker Diarization on Multiple Microphone Meetings
Dan Istrate, Corinne Fredouille, Sylvain Meignier, Laurent Besacier, Jean François Bonastre 428

The TNO Speaker Diarization System for NIST RT05s Meeting Data
David A. van Leeuwen 440

The 2005 AMI System for the Transcription of Speech in Meetings
Thomas Hain, Lukas Burget, John Dines, Giulia Garau, Martin Karafiat, Mike Lincoln, Iain McCowan, Darren Moore, Vincent Wan, Roeland Ordelman, Steve Renals 450

Further Progress in Meeting Recognition: The ICSI-SRI Spring 2005 Speech-to-Text Evaluation System
Andreas Stolcke, Xavier Anguera, Kofi Boakye, Özgür Çetin, František Grézl, Adam Janin, Arindam Mandal, Barbara Peskin, Chuck Wooters, Jing Zheng 463

Speaker Localization in CHIL Lectures: Evaluation Criteria and Results
Maurizio Omologo, Piergiorgio Svaizer, Alessio Brutti, Luca Cristoforetti 476

Author Index 489