Lecture Notes in Computer Science 4299

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
University of California, Los Angeles, CA, 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
Jonathan G. Fiscus (Eds.)

Machine Learning for Multimodal Interaction

Third International Workshop, MLMI 2006
Bethesda, MD, USA, May 1-4, 2006
Revised Selected Papers

Springer

Volume Editors

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

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

Jonathan G. Fiscus
National Institute of Standards and Technology (NIST)
100 Bureau Drive, Gaithersburg, MD 20899-8940, USA
E-mail: jfiscus@nist.gov

Library of Congress Control Number: 2006938909

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 3-540-69267-3 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-69267-6 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: 11965152 06/3142 5 4 3 2 1 0

Preface

This book contains a selection of refereed papers presented at the 3rd Workshop on Machine Learning for Multimodal Interaction (MLMI 2006), held in Bethesda MD, USA during May 1–4, 2006.

The workshop was organized and sponsored jointly by the US National Institute for Standards and Technology (NIST), three projects supported by the European Commission (Information Society Technologies priority of the sixth Framework Programme)—the AMI and CHIL Integrated Projects, and the PASCAL Network of Excellence—and the Swiss National Science Foundation national research collaboration, IM2.

In addition to the main workshop, MLMI 2006 was co-located with the 4th NIST Meeting Recognition Workshop. This workshop was centered on the Rich Transcription 2006 Spring Meeting Recognition (RT-06) evaluation of speech technologies within the meeting domain. Building on the success of previous evaluations in this domain, the RT-06 evaluation continued evaluation tasks in the areas of speech-to-text, who-spoke-when, and speech activity detection.

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, human–computer interaction, and the modeling of discourse and dialog, with an emphasis on the application of machine learning. Out of the submitted full papers, about 50% were accepted for publication in the present volume, after authors had been invited to take review comments and conference feedback into account. The workshop featured invited talks from Roderick Murray-Smith (University of Glasgow), Tsuhan Chen (Carnegie Mellon University) and David McNeill (University of Chicago), and a special session on projects in the area of multimodal interaction including presentations on the VACE, CHIL and AMI projects.

Based on the successes of the first three MLMI workshops, and to strengthen and broaden the base of this workshop series, the MLMI standing committee was formed. The initial membership comprises Samy Bengio (IDIAP), Hervé Bourlard (IDIAP and EPFL), Tsuhan Chen (Carnegie Mellon University), John Garofolo (NIST), Mary Harper (Purdue University), Sharon Oviatt (Natural Interaction Systems), Steve Renals (Edinburgh University), Rainer Stiefelhagen (Universität Karlsruhe), and Alex Waibel (Carnegie Mellon University and Universität Karlsruhe). The committee will provide a permanent link across MLMI workshops. MLMI 2007, the fourth workshop in the series, will take place in

Brno, Czech Republic during June 28–30, 2007, directly after ACL–2007, which takes place in Prague.

Finally, we take this opportunity to thank our Programme 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.

October 2006

Steve Renals
Samy Bengio
Jonathan Fiscus

Organization

Organizing Committee

Samy Bengio	IDIAP Research Institute
Hervé Bourlard	IDIAP Research Institute
Jonathan Fiscus (Co-chair)	NIST
John Garofolo	NIST
Steve Renals (Co-chair)	University of Edinburgh
Vincent Stanford (Demonstrations)	NIST
Alex Waibel (Special Sessions)	CMU and Universität Karlsruhe

Workshop Organization

Patrice Boulanger	NIST
Caroline Hastings	University of Edinburgh
Avril Heron	University of Edinburgh
Jonathan Kilgour	University of Edinburgh
Teresa Vicente	NIST

Program Committee

Marc Al-Hames	Munich University of Technology
Tilman Becker	DFKI
Jean Carletta	University of Edinburgh
Dan Ellis	Columbia University
Corinne Fredouille	University of Avignon
Thomas Hain	University of Sheffield
Mary Harper	Purdue University
Thomas Huang	University of Illinois
Alejandro Jaimes	Fuji Xerox
Samuel Kaski	University of Helsinki
Stephane Marchand-Maillat	University of Geneva
Nelson Morgan	ICSI
Andrei Popescu-Belis	University of Geneva
Mubarak Shah	University of Central Florida
Rainer Stiefelhagen	Universität Karlsruhe
Jean-Philippe Thiran	EPFL
Victor Tom	BAE Systems
Pierre Wellner	IDIAP Research Institute

Sponsoring Projects and Institutions

Institutions:

- US National Institute of Standards and Technology (NIST), http://www.nist.gov/speech/
- 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

Projects:

- AMI, Augmented Multiparty Interaction, http://www.amiproject.org/
- CHIL, Computers in the Human Interaction Loop, http://chil.server.de/
- PASCAL, Pattern Analysis, Statistical Modeling and Computational Learning, http://www.pascal-network.org/
- IM2, Interactive Multimodal Information Management, http://www.im2.ch/

Table of Contents

MLMI'06

I Invited Paper

Model-Based, Multimodal Interaction in Document Browsing 1
Parisa Eslambolchilar and Roderick Murray-Smith

II Multimodal Processing

The NIST Meeting Room Corpus 2 Phase 1 13
Martial Michel, Jerome Ajot, and Jonathan Fiscus

Audio-Visual Processing in Meetings: Seven Questions and Current AMI Answers 24
Marc Al-Hames, Thomas Hain, Jan Cernocky, Sascha Schreiber, Mannes Poel, Ronald Müller, Sebastien Marcel, David van Leeuwen, Jean-Marc Odobez, Sileye Ba, Herve Bourlard, Fabien Cardinaux, Daniel Gatica-Perez, Adam Janin, Petr Motlicek, Stephan Reiter, Steve Renals, Jeroen van Rest, Rutger Rienks, Gerhard Rigoll, Kevin Smith, Andrew Thean, and Pavel Zemcik

A Multimodal Analysis of Floor Control in Meetings 36
Lei Chen, Mary Harper, Amy Franklin, Travis R. Rose, Irene Kimbara, Zhongqiang Huang, and Francis Quek

Combining User Modeling and Machine Learning to Predict Users' Multimodal Integration Patterns 50
Xiao Huang, Sharon Oviatt, and Rebecca Lunsford

Using Audio, Visual, and Lexical Features in a Multi-modal Virtual Meeting Director 63
Marc Al-Hames, Benedikt Hörnler, Christoph Scheuermann, and Gerhard Rigoll

III Image and Video Processing

A Study on Visual Focus of Attention Recognition from Head Pose in a Meeting Room 75
Sileye O. Ba and Jean-Marc Odobez

Multi-person Tracking in Meetings: A Comparative Study 88
Kevin Smith, Sascha Schreiber, Igor Potúcek, Vítzslav Beran, Gerhard Rigoll, and Daniel Gatica-Perez

Gaussian Mixture Models for CHASM Signature Verification 102
Andreas Humm, Jean Hennebert, and Rolf Ingold

Kalman Tracking with Target Feedback on Adaptive Background Learning 114
Aristodemos Pnevmatikakis and Lazaros Polymenakos

Da Vinci's Mona Lisa: A Modern Look at a Timeless Classic 123
Dennis Lin, Jilin Tu, Shyamsundar Rajaram, Zhenqiu Zhang, and Thomas Huang

IV HCI and Applications

The Connector Service-Predicting Availability in Mobile Contexts 129
Maria Danninger, Erica Robles, Leila Takayama, QianYing Wang, Tobias Kluge, Rainer Stiefelhagen, and Clifford Nass

Multimodal Input for Meeting Browsing and Retrieval Interfaces: Preliminary Findings 142
Agnes Lisowska and Susan Armstrong

V Discourse and Dialogue

Gesture Features for Coreference Resolution 154
Jacob Eisenstein and Randall Davis

Syntactic Chunking Across Different Corpora 166
Weiqun Xu, Jean Carletta, and Johanna Moore

Multistream Recognition of Dialogue Acts in Meetings 178
Alfred Dielmann and Steve Renals

Text Based Dialog Act Classification for Multiparty Meetings 190
Matthias Zimmermann, Dilek Hakkani-Tür, Elizabeth Shriberg, and Andreas Stolcke

Detecting Action Items in Multi-party Meetings: Annotation and Initial Experiments 200
Matthew Purver, Patrick Ehlen, and John Niekrasz

Overlap in Meetings: ASR Effects and Analysis by Dialog Factors, Speakers, and Collection Site 212
Özgür Çetin and Elizabeth Shriberg

VI Speech and Audio Processing

A Speaker Localization System for Lecture Room Environment 225
Mikko Parviainen, Tuomo Pirinen, and Pasi Pertilä

Robust Speech Activity Detection in Interactive Smart-Room Environments 236
Dušan Macho, Climent Nadeu, and Andrey Temko

Automatic Cluster Complexity and Quantity Selection: Towards Robust Speaker Diarization 248
Xavier Anguera, Chuck Wooters, and Javier Hernando

Speaker Diarization for Multi-microphone Meetings Using Only Between-Channel Differences 257
Jose M. Pardo, Xavier Anguera, and Chuck Wooters

Warped and Warped-Twice MVDR Spectral Estimation With and Without Filterbanks 265
Matthias Wölfel

Robust Heteroscedastic Linear Discriminant Analysis and LCRC Posterior Features in Meeting Data Recognition 275
Martin Karafiát, František Grézl, Petr Schwarz, Lukáš Burget, and Jan Černocký

Juicer: A Weighted Finite-State Transducer Speech Decoder 285
Darren Moore, John Dines, Mathew Magimai Doss, Jithendra Vepa, Octavian Cheng, and Thomas Hain

Speech-to-Speech Translation Services for the Olympic Games 2008 297
Sebastian Stüker, Chengqing Zong, Jürgen Reichert, Wenjie Cao, Muntsin Kolss, Guodong Xie, Kay Peterson, Peng Ding, Victoria Arranz, Jian Yu, and Alex Waibel

VII NIST Meeting Recognition Evaluation

The Rich Transcription 2006 Spring Meeting Recognition Evaluation 309
Jonathan G. Fiscus, Jerome Ajot, Martial Michel, and John S. Garofolo

The IBM RT06s Evaluation System for Speech Activity Detection in CHIL Seminars 323
Etienne Marcheret, Gerasimos Potamianos, Karthik Visweswariah, and Jing Huang

A Lightweight Speech Detection System for Perceptive Environments 336
Dominique Vaufreydaz, Rémi Emonet, and Patrick Reignier

Robust Speaker Diarization for Meetings: ICSI RT06S Meetings Evaluation System 346
Xavier Anguera, Chuck Wooters, and Jose M. Pardo

Technical Improvements of the E-HMM Based Speaker Diarization System for Meeting Records 359
Corinne Fredouille and Grégory Senay

The AMI Speaker Diarization System for NIST RT06s Meeting Data 371
David A. van Leeuwen and Marijn Huijbregts

The 2006 Athens Information Technology Speech Activity Detection and Speaker Diarization Systems 385
Elias Rentzeperis, Andreas Stergiou, Christos Boukis, Aristodemos Pnevmatikakis, and Lazaros C. Polymenakos

Speaker Diarization: From Broadcast News to Lectures 396
Xuan Zhu, Claude Barras, Lori Lamel, and Jean-Luc Gauvain

The ISL RT-06S Speech-to-Text System 407
Christian Fügen, Shajith Ikbal, Florian Kraft, Kenichi Kumatani, Kornel Laskowski, John W. McDonough, Mari Ostendorf, Sebastian Stüker, and Matthias Wölfel

The AMI Meeting Transcription System: Progress and Performance 419
Thomas Hain, Lukas Burget, John Dines, Giulia Garau, Martin Karafiat, Mike Lincoln, Jithendra Vepa, and Vincent Wan

The IBM Rich Transcription Spring 2006 Speech-to-Text System for Lecture Meetings 432
Jing Huang, Martin Westphal, Stanley Chen, Olivier Siohan, Daniel Povey, Vit Libal, Alvaro Soneiro, Henrik Schulz, Thomas Ross, and Gerasimos Potamianos

The ICSI-SRI Spring 2006 Meeting Recognition System 444
Adam Janin, Andreas Stolcke, Xavier Anguera, Kofi Boakye, Özgür Çetin, Joe Frankel, and Jing Zheng

The LIMSI RT06s Lecture Transcription System 457
Lori Lamel, Eric Bilinski, Gilles Adda, Jean-Luc Gauvain, and Holger Schwenk

Author Index 469