Lecture Notes in Computer Science 3361

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

Samy Bengio Hervé Bourlard (Eds.)

Machine Learning for Multimodal Interaction

First International Workshop, MLMI 2004
Martigny, Switzerland, June 21-23, 2004
Revised Selected Papers

Springer

Volume Editors

Samy Bengio
Hervé Bourlard
IDIAP Research Institute
Rue du Simplon 4, P.O. Box 592, 1920 Martigny, Switzerland
E-mail: {bengio,bourlard}@idiap.ch

Library of Congress Control Number: 2004118425

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

ISSN 0302-9743
ISBN 3-540-24509-X 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

springeronline.com

© Springer-Verlag Berlin Heidelberg 2005
Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 11384212 06/3142 5 4 3 2 1 0

Preface

This book contains a selection of refereed papers presented at the 1st Workshop on Machine Learning for Multimodal Interaction (MLMI 2004), held at the "Centre du Parc," Martigny, Switzerland, during June 21–23, 2004. The workshop was organized and sponsored jointly by three European projects,

- AMI, Augmented Multiparty Interaction, http://www.amiproject.org
- PASCAL, Pattern Analysis, Statistical Modeling and Computational Learning, http://www.pascal-network.org
- M4, Multi-modal Meeting Manager, http://www.m4project.org

as well as the Swiss National Centre of Competence in Research (NCCR):

- IM2: Interactive Multimodal Information Management, http://www.im2.ch

MLMI 2004 was thus sponsored by the European Commission and the Swiss National Science Foundation.

Given the multiple links between the above projects and several related research areas, it was decided to organize 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 in meetings. The motivation for creating such a forum, which could be perceived as a number of papers from different research disciplines, evolved from a real need that arose from these projects and the strong motivation of their partners for such a multidisciplinary workshop. This assessment was indeed confirmed by the success of this first MLMI workshop, which attracted more than 200 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 M4, AMI and IM2 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 60% were accepted for publication in this volume, after the authors were invited to take review comments and conference feedback into account.

In this book, and following the structure of the workshop, the papers were divided into the following sections:

1. HCI and Applications
2. Structuring and Interaction
3. Multimodal Processing
4. Speech Processing
5. Dialogue Management
6. Vision and Emotion

In the spirit of MLMI 2004 and its associated projects, all the oral presentations were recorded, and synchronized with additional material (such as presentation slides) and are now available, with search facilities, at: **http://mmm.idiap.ch/mlmi04/**

Based on the success of MLMI 2004, a series of MLMI workshop is now being planned, with the goal of involving a larger community, as well as a larger number of European projects working in similar or related areas. MLMI 2005 will be organized by the University of Edinburgh and held on 11–13 July 2005, also in collaboration with the NIST (US National Institute of Standards and Technology), while MLMI 2006 will probably be held in the US, probably in conjunction with a NIST evaluation.

Finally, we take this opportunity to thank our Program Committee members for an excellent job, as well as the sponsoring projects and funding agencies. We also thank all our administrative support, especially Nancy Robyr who played a key role in the management and organization of the workshop, as well as in the follow-up of all the details resulting in this book.

December 2004

Samy Bengio
Hervé Bourlard

Organization

General Chairs

Samy Bengio	IDIAP Research Institute, Switzerland
Hervé Bourlard	IDIAP Research Institute and EPFL, Switzerland

Program Committee

Jean Carletta	University of Edinburgh, UK
Daniel Gatica-Perez	IDIAP Research Institute, Switzerland
Phil Green	University of Sheffield, UK
Hynek Hermansky	IDIAP Research Institute, Switzerland
Jan Larsen	Technical University of Denmark
Nelson Morgan	ICSI, Berkeley, USA
Erkki Oja	Helsinki University of Technology, Finland
Barbara Peskin	ICSI, Berkeley, USA
Thierry Pun	University of Geneva, Switzerland
Steve Renals	University of Edinburgh, UK
John Shawe-Taylor	University of Southampton, UK
Jean-Philippe Thiran	EPFL Lausanne, Switzerland
Luc Van Gool	ETHZ Zurich, Switzerland
Pierre Wellner	IDIAP Research Institute, Switzerland
Steve Whittaker	University of Sheffield, UK

Sponsoring Projects and Institutions

Projects:

- Augmented Multiparty Interaction (AMI), http://www.amiproject.org
- Pattern Analysis, Statistical Modeling and Computational Learning (PASCAL), http://www.pascal-network.org
- Multi-modal Meeting Manager (M4), http://www.m4project.org
- Interactive Multimodal Information Management (IM2), http://www.im2.ch

Institutions:

- European Commission
- Swiss National Science Foundation, through the National Centres of Competence in Research (NCCR) program

Table of Contents

MLMI 2004

I HCI and Applications

Accessing Multimodal Meeting Data: Systems, Problems and Possibilities
Simon Tucker, Steve Whittaker 1

Browsing Recorded Meetings with Ferret
Pierre Wellner, Mike Flynn, Maël Guillemot 12

Meeting Modelling in the Context of Multimodal Research
Dennis Reidsma, Rutger Rienks, Nataša Jovanović 22

Artificial Companions
Yorick Wilks 36

Zakim – A Multimodal Software System for Large-Scale Teleconferencing
Max Froumentin 46

II Structuring and Interaction

Towards Computer Understanding of Human Interactions
Iain McCowan, Daniel Gatica-Perez, Samy Bengio, Darren Moore, Hervé Bourlard 56

Multistream Dynamic Bayesian Network for Meeting Segmentation
Alfred Dielmann, Steve Renals 76

Using Static Documents as Structured and Thematic Interfaces to Multimedia Meeting Archives
Denis Lalanne, Rolf Ingold, Didier von Rotz, Ardhendu Behera, Dalila Mekhaldi, Andrei Popescu-Belis 87

An Integrated Framework for the Management of Video Collection
Nicolas Moënne-Loccoz, Bruno Janvier, Stéphane Marchand-Maillet, Eric Bruno 101

The NITE XML Toolkit Meets the ICSI Meeting Corpus: Import, Annotation, and Browsing
Jean Carletta, Jonathan Kilgour 111

III Multimodal Processing

S-SEER: Selective Perception in a Multimodal Office Activity Recognition System
Nuria Oliver, Eric Horvitz 122

Mapping from Speech to Images Using Continuous State Space Models
Tue Lehn-Schiøler, Lars Kai Hansen, Jan Larsen 136

An Online Algorithm for Hierarchical Phoneme Classification
Ofer Dekel, Joseph Keshet, Yoram Singer 146

Towards Predicting Optimal Fusion Candidates: A Case Study on Biometric Authentication Tasks
Norman Poh, Samy Bengio 159

Mixture of SVMs for Face Class Modeling
Julien Meynet, Vlad Popovici, Jean Philippe Thiran 173

AV16.3: An Audio-Visual Corpus for Speaker Localization and Tracking
Guillaume Lathoud , Jean-Marc Odobez, Daniel Gatica-Perez 182

IV Speech Processing

The 2004 ICSI-SRI-UW Meeting Recognition System
Chuck Wooters, Nikki Mirghafori, Andreas Stolcke, Tuomo Pirinen, Ivan Bulyko, Dave Gelbart, Martin Graciarena, Scott Otterson, Barbara Peskin, Mari Ostendorf 196

On the Adequacy of Baseform Pronunciations and Pronunciation Variants
Mathew Magimai-Doss, Hervé Bourlard 209

Tandem Connectionist Feature Extraction for Conversational Speech Recognition
Qifeng Zhu, Barry Chen, Nelson Morgan, Andreas Stolcke 223

Long-Term Temporal Features for Conversational Speech Recognition
Barry Chen, Qifeng Zhu, Nelson Morgan 232

Speaker Indexing in Audio Archives Using Gaussian Mixture Scoring Simulation
Hagai Aronowitz, David Burshtein, Amihood Amir 243

Speech Transcription and Spoken Document Retrieval in Finnish
Mikko Kurimo, Ville Turunen, Inger Ekman 253

A Mixed-Lingual Phonological Component Which Drives the Statistical Prosody Control of a Polyglot TTS Synthesis System
Harald Romsdorfer, Beat Pfister, René Beutler 263

V Dialogue Management

Shallow Dialogue Processing Using Machine Learning Algorithms (or Not)
Andrei Popescu-Belis, Alexander Clark, Maria Georgescul, Denis Lalanne, Sandrine Zufferey 277

ARCHIVUS: A System for Accessing the Content of Recorded Multimodal Meetings
Agnes Lisowska, Martin Rajman, Trung H. Bui 291

VI Vision and Emotion

Piecing Together the Emotion Jigsaw
Roddy Cowie, Marc Schröder 305

Emotion Analysis in Man-Machine Interaction Systems
T. Balomenos, A. Raouzaiou, S. Ioannou, A. Drosopoulos, K. Karpouzis, S. Kollias 318

A Hierarchical System for Recognition, Tracking and Pose Estimation
Philipp Zehnder, Esther Koller-Meier, Luc Van Gool 329

Automatic Pedestrian Tracking Using Discrete Choice Models and Image Correlation Techniques
Santiago Venegas-Martinez, Gianluca Antonini, Jean Philippe Thiran, Michel Bierlaire 341

A Shape Based, Viewpoint Invariant Local Descriptor
Mihai Osian, Tinne Tuytelaars, Luc Van Gool 349

Author Index 361