

Rainer Stiefelhagen
John Garofolo (Eds.)

LNCS 4122

Multimodal Technologies for Perception of Humans

**First International Evaluation Workshop on Classification
of Events, Activities and Relationships, CLEAR 2006
Southampton, UK, April 2006, Revised Selected Papers**

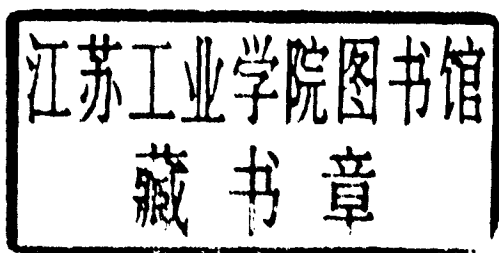


Springer

Rainer Stiefelhagen John Garofolo (Eds.)

Multimodal Technologies for Perception of Humans

First International Evaluation Workshop on Classification
of Events, Activities and Relationships, CLEAR 2006
Southampton, UK, April 6-7, 2006
Revised Selected Papers



Volume Editors

Rainer Stiefel
Universität Karlsruhe (TH)
Institut für Theoretische Informatik
Am Fasanengarten 5, 76131 Karlsruhe, Germany
E-mail: stiefel@ira.uka.de

John Garofolo
National Institute of Standards and Technology
100 Bureau Drive, Stop 8940, Gaithersburg, MD 20899-8940, USA
E-mail: garofolo@nist.gov

Library of Congress Control Number: 2006939517

CR Subject Classification (1998): I.4, I.5, I.2.10, I.3.5, I.2.6, F.2.2

LNCS Sublibrary: SL 6 – Image Processing, Computer Vision, Pattern Recognition, and Graphics

ISSN	0302-9743
ISBN-10	3-540-69567-2 Springer Berlin Heidelberg New York
ISBN-13	978-3-540-69567-7 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 2007
Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 11970897 06/3142 5 4 3 2 1 0

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

Lecture Notes in Computer Science

For information about Vols. 1–4274

please contact your bookseller or Springer

- Vol. 4377: M. Abe (Ed.), Topics in Cryptology – CT-RSA 2007. XI, 403 pages. 2006.
- Vol. 4371: K. Inoue, K. Satoh, F. Toni (Eds.), Computational Logic in Multi-Agent Systems. X, 315 pages. 2007. (Sublibrary LNAI).
- Vol. 4369: M. Umeda, A. Wolf, O. Bartenstein, U. Geske, D. Seipel, O. Takata (Eds.), Declarative Programming for Knowledge Management. X, 229 pages. 2006. (Sublibrary LNAI).
- Vol. 4368: T. Erlebach, C. Kaklamanis (Eds.), Approximation and Online Algorithms. X, 345 pages. 2006.
- Vol. 4367: K. De Bosschere, D. Kaeli, P. Stenström, D. Whalley, T. Ungerer (Eds.), High Performance Embedded Architectures and Compilers. XI, 307 pages. 2007.
- Vol. 4362: J. van Leeuwen, G.F. Italiano, W. van der Hoek, C. Meinel, H. Sack, F. Plášil (Eds.), SOFSEM 2007: Theory and Practice of Computer Science. XXI, 937 pages. 2006.
- Vol. 4361: H.J. Hoogeboom, G. Păun, G. Rozenberg, A. Salomaa (Eds.), Membrane Computing. IX, 555 pages. 2006.
- Vol. 4357: L. Buttyán, V. Gligor, D. Westhoff (Eds.), Security and Privacy in Ad-Hoc and Sensor Networks. X, 193 pages. 2006.
- Vol. 4355: J. Julliand, O. Kouchnarenko (Eds.), B 2007: Formal Specification and Development in B. XIII, 293 pages. 2006.
- Vol. 4354: M. Hanus (Ed.), Practical Aspects of Declarative Languages. X, 335 pages. 2006.
- Vol. 4353: T. Schwentick, D. Suciu (Eds.), Database Theory – ICDT 2007. XI, 419 pages. 2006.
- Vol. 4352: T.-J. Cham, J. Cai, C. Dorai, D. Rajan, T.-S. Chua, L.-T. Chia (Eds.), Advances in Multimedia Modeling, Part II. XVIII, 743 pages. 2006.
- Vol. 4351: T.-J. Cham, J. Cai, C. Dorai, D. Rajan, T.-S. Chua, L.-T. Chia (Eds.), Advances in Multimedia Modeling, Part I. XIX, 797 pages. 2006.
- Vol. 4349: B. Cook, A. Podelski (Eds.), Verification, Model Checking, and Abstract Interpretation. XI, 395 pages. 2006.
- Vol. 4348: S.T. Taft, R.A. Duff, R.L. Brukardt, E. Ploedereder, P. Leroy (Eds.), Ada 2005 Reference Manual. XXII, 765 pages. 2006.
- Vol. 4347: J. Lopez (Ed.), Critical Information Infrastructures Security. X, 286 pages. 2006.
- Vol. 4345: N. Maglaveras, I. Chouvarda, V. Koutkias, R. Brause (Eds.), Biological and Medical Data Analysis. XIII, 496 pages. 2006. (Sublibrary LNBI).
- Vol. 4344: V. Gruhn, F. Oquendo (Eds.), Software Architecture. X, 245 pages. 2006.
- Vol. 4342: H. de Swart, E. Orlowska, G. Schmidt, M. Roubens (Eds.), Theory and Applications of Relational Structures as Knowledge Instruments II. X, 373 pages. 2006. (Sublibrary LNAI).
- Vol. 4341: P.Q. Nguyen (Ed.), Progress in Cryptology – VIETCRYPT 2006. XI, 385 pages. 2006.
- Vol. 4340: R. Prodan, T. Fahringer, Grid Computing. XXIII, 317 pages. 2006.
- Vol. 4339: E. Ayguadé, G. Baumgartner, J. Ramanujam, P. Sadayappan (Eds.), Languages and Compilers for Parallel Computing. XI, 476 pages. 2006.
- Vol. 4338: P. Kalra, S. Peleg (Eds.), Computer Vision, Graphics and Image Processing. XV, 965 pages. 2006.
- Vol. 4337: S. Arun-Kumar, N. Garg (Eds.), FSTTCS 2006: Foundations of Software Technology and Theoretical Computer Science. XIII, 430 pages. 2006.
- Vol. 4334: B. Beckert, R. Hähnle, P.H. Schmitt (Eds.), Verification of Object-Oriented Software. XXIX, 658 pages. 2006. (Sublibrary LNAI).
- Vol. 4333: U. Reimer, D. Karagiannis (Eds.), Practical Aspects of Knowledge Management. XII, 338 pages. 2006. (Sublibrary LNAI).
- Vol. 4332: A. Bagchi, V. Atluri (Eds.), Information Systems Security. XV, 382 pages. 2006.
- Vol. 4331: G. Min, B. Di Martino, L.T. Yang, M. Guo, G. Ruenger (Eds.), Frontiers of High Performance Computing and Networking – ISPA 2006 Workshops. XXXVII, 1141 pages. 2006.
- Vol. 4330: M. Guo, L.T. Yang, B. Di Martino, H.P. Zima, J. Dongarra, F. Tang (Eds.), Parallel and Distributed Processing and Applications. XVIII, 953 pages. 2006.
- Vol. 4329: R. Barua, T. Lange (Eds.), Progress in Cryptology – INDOCRYPT 2006. X, 454 pages. 2006.
- Vol. 4328: D. Penkler, M. Reitenspiess, F. Tam (Eds.), Service Availability. X, 289 pages. 2006.
- Vol. 4327: M. Baldoni, U. Endriss (Eds.), Declarative Agent Languages and Technologies IV. VIII, 257 pages. 2006. (Sublibrary LNAI).
- Vol. 4326: S. Göbel, R. Malkewitz, I. Iurgel (Eds.), Technologies for Interactive Digital Storytelling and Entertainment. X, 384 pages. 2006.
- Vol. 4325: J. Cao, I. Stojmenovic, X. Jia, S.K. Das (Eds.), Mobile Ad-hoc and Sensor Networks. XIX, 887 pages. 2006.
- Vol. 4320: R. Gotzhein, R. Reed (Eds.), System Analysis and Modeling: Language Profiles. X, 229 pages. 2006.

- Vol. 4319: L.-W. Chang, W.-N. Lie (Eds.), *Advances in Image and Video Technology*. XXVI, 1347 pages. 2006.
- Vol. 4318: H. Lipmaa, M. Yung, D. Lin (Eds.), *Information Security and Cryptology*. XI, 305 pages. 2006.
- Vol. 4317: S.K. Madria, K.T. Claypool, R. Kannan, P. Uppuluri, M.M. Gore (Eds.), *Distributed Computing and Internet Technology*. XIX, 466 pages. 2006.
- Vol. 4316: M.M. Dalkilic, S. Kim, J. Yang (Eds.), *Data Mining and Bioinformatics*. VIII, 197 pages. 2006. (Sublibrary LNBI).
- Vol. 4313: T. Margaria, B. Steffen (Eds.), *Leveraging Applications of Formal Methods*. IX, 197 pages. 2006.
- Vol. 4312: S. Sugimoto, J. Hunter, A. Rauber, A. Morishima (Eds.), *Digital Libraries: Achievements, Challenges and Opportunities*. XVIII, 571 pages. 2006.
- Vol. 4311: K. Cho, P. Jacquet (Eds.), *Technologies for Advanced Heterogeneous Networks II*. XI, 253 pages. 2006.
- Vol. 4309: P. Inverardi, M. Jazayeri (Eds.), *Software Engineering Education in the Modern Age*. VIII, 207 pages. 2006.
- Vol. 4308: S. Chaudhuri, S.R. Das, H.S. Paul, S. Tirthapura (Eds.), *Distributed Computing and Networking*. XIX, 608 pages. 2006.
- Vol. 4307: P. Ning, S. Qing, N. Li (Eds.), *Information and Communications Security*. XIV, 558 pages. 2006.
- Vol. 4306: Y. Avrithis, Y. Kompatsiaris, S. Staab, N.E. O'Connor (Eds.), *Semantic Multimedia*. XII, 241 pages. 2006.
- Vol. 4305: A.A. Shvartsman (Ed.), *Principles of Distributed Systems*. XIII, 441 pages. 2006.
- Vol. 4304: A. Sattar, B.-H. Kang (Eds.), *AI 2006: Advances in Artificial Intelligence*. XXVII, 1303 pages. 2006. (Sublibrary LNAI).
- Vol. 4303: A. Hoffmann, B.-H. Kang, D. Richards, S. Tsumoto (Eds.), *Advances in Knowledge Acquisition and Management*. XI, 259 pages. 2006. (Sublibrary LNAI).
- Vol. 4302: J. Domingo-Ferrer, L. Franconi (Eds.), *Privacy in Statistical Databases*. XI, 383 pages. 2006.
- Vol. 4301: D. Pointcheval, Y. Mu, K. Chen (Eds.), *Cryptography and Network Security*. XIII, 381 pages. 2006.
- Vol. 4300: Y.Q. Shi (Ed.), *Transactions on Data Hiding and Multimedia Security I*. IX, 139 pages. 2006.
- Vol. 4299: S. Renals, S. Bengio, J.G. Fiscus (Eds.), *Machine Learning for Multimodal Interaction*. XII, 470 pages. 2006.
- Vol. 4297: Y. Robert, M. Parashar, R. Badrinath, V.K. Prasanna (Eds.), *High Performance Computing - HiPC 2006*. XXIV, 642 pages. 2006.
- Vol. 4296: M.S. Rhee, B. Lee (Eds.), *Information Security and Cryptology - ICISC 2006*. XIII, 358 pages. 2006.
- Vol. 4295: J.D. Carswell, T. Tezuka (Eds.), *Web and Wireless Geographical Information Systems*. XI, 269 pages. 2006.
- Vol. 4294: A. Dan, W. Lamersdorf (Eds.), *Service-Oriented Computing - ICSC 2006*. XIX, 653 pages. 2006.
- Vol. 4293: A. Gelbukh, C.A. Reyes-Garcia (Eds.), *MI-CAI 2006: Advances in Artificial Intelligence*. XXVIII, 1232 pages. 2006. (Sublibrary LNAI).
- Vol. 4292: G. Bebis, R. Boyle, B. Parvin, D. Koracin, P. Remagnino, A. Nefian, G. Meenakshisundaram, V. Pascucci, J. Zara, J. Molineros, H. Theisel, T. Malzbender (Eds.), *Advances in Visual Computing, Part II*. XXXII, 906 pages. 2006.
- Vol. 4291: G. Bebis, R. Boyle, B. Parvin, D. Koracin, P. Remagnino, A. Nefian, G. Meenakshisundaram, V. Pascucci, J. Zara, J. Molineros, H. Theisel, T. Malzbender (Eds.), *Advances in Visual Computing, Part I*. XXXI, 916 pages. 2006.
- Vol. 4290: M. van Steen, M. Henning (Eds.), *Middleware 2006*. XIII, 425 pages. 2006.
- Vol. 4289: M. Ackermann, B. Berendt, M. Grobelnik, A. Hotho, D. Mladenić, G. Semeraro, M. Spiliopoulou, G. Stumme, V. Svátek, M. van Someren (Eds.), *Semantics, Web and Mining*. X, 197 pages. 2006. (Sublibrary LNAI).
- Vol. 4288: T. Asano (Ed.), *Algorithms and Computation*. XX, 766 pages. 2006.
- Vol. 4287: C. Mao, T. Yokomori (Eds.), *DNA Computing*. XII, 440 pages. 2006.
- Vol. 4286: P.G. Spirakis, M. Mavronicolas, S.C. Konogiannis (Eds.), *Internet and Network Economics*. XI, 401 pages. 2006.
- Vol. 4285: Y. Matsumoto, R.W. Sproat, K.-F. Wong, M. Zhang (Eds.), *Computer Processing of Oriental Languages*. XVII, 544 pages. 2006. (Sublibrary LNAI).
- Vol. 4284: X. Lai, K. Chen (Eds.), *Advances in Cryptology - ASIACRYPT 2006*. XIV, 468 pages. 2006.
- Vol. 4283: Y.Q. Shi, B. Jeon (Eds.), *Digital Watermarking*. XII, 474 pages. 2006.
- Vol. 4282: Z. Pan, A. Cheok, M. Haller, R.W.H. Lau, H. Saito, R. Liang (Eds.), *Advances in Artificial Reality and Tele-Existence*. XXIII, 1347 pages. 2006.
- Vol. 4281: K. Barkaoui, A. Cavalcanti, A. Cerone (Eds.), *Theoretical Aspects of Computing - ICTAC 2006*. XV, 371 pages. 2006.
- Vol. 4280: A.K. Datta, M. Gradinariu (Eds.), *Stabilization, Safety, and Security of Distributed Systems*. XVII, 590 pages. 2006.
- Vol. 4279: N. Kobayashi (Ed.), *Programming Languages and Systems*. XI, 423 pages. 2006.
- Vol. 4278: R. Meersman, Z. Tari, P. Herrero (Eds.), *On the Move to Meaningful Internet Systems 2006: OTM 2006 Workshops, Part II*. XLV, 1004 pages. 2006.
- Vol. 4277: R. Meersman, Z. Tari, P. Herrero (Eds.), *On the Move to Meaningful Internet Systems 2006: OTM 2006 Workshops, Part I*. XLV, 1009 pages. 2006.
- Vol. 4276: R. Meersman, Z. Tari (Eds.), *On the Move to Meaningful Internet Systems 2006: CoopIS, DOA, GADA, and ODBASE, Part II*. XXXII, 752 pages. 2006.
- Vol. 4275: R. Meersman, Z. Tari (Eds.), *On the Move to Meaningful Internet Systems 2006: CoopIS, DOA, GADA, and ODBASE, Part I*. XXXI, 1115 pages. 2006.

Preface

During winter and spring 2006, the first international CLEAR evaluation took place, which targets the evaluation of systems for the perception of people, their identities, activities, interactions and relationships in human–human interaction scenarios as well as related scenarios. As part of the evaluation, a two-day workshop was held during April 6–7, 2006, in Southampton, UK, in which the participating systems were presented and the evaluation results discussed in detail.

This book contains the system description papers that were presented at the CLEAR 2006 workshop as well as an overview of the evaluation tasks and the results that were obtained in each of these by the various participants. The book also includes two invited papers about related evaluation activities that were presented at the workshop.

The CLEAR evaluation campaign and workshop was jointly organized by the Universität Karlsruhe, Germany and the US National Institute of Standards and Technology (NIST).

CLEAR 2006 was supported by the European Integrated Project CHIL—Computers in the Human Interaction Loop—as well as the US DTO—Disruptive Technology Office - VACE - Video Analysis Content Extraction—program, which jointly organized part of their perceptual technology evaluations within CLEAR 2006 for the first time. CLEAR 2006 was thus sponsored by the European Commission (Information Society Technologies priority of the Sixth Framework Programme) and the US DTO.

CLEAR 2006 was also organized in cooperation with the NIST RT - Rich Transcription Meeting Recognition evaluation, which focused more on the evaluation of content-related technologies, such as speech and video text recognition. CLEAR and RT shared some of their evaluation data sets, so that, for example, the speaker-localization results generated for CLEAR could be used for the far-field speech-to-text task in RT06. This was facilitated through the harmonization of the 2006 CLEAR and RT evaluation deadlines.

The current evaluation tasks that were conducted in CLEAR 2006 consisted of part of the evaluation tasks related to human activity analysis that were part of the CHIL and VACE perceptual technology evaluation activities. These current CLEAR 2006 evaluation tasks can be categorized as follows:

- Tracking tasks (faces/persons/vehicles, 2D/3D, acoustic/visual/audio-visual)
- Person identification tasks (acoustic, visual, audio-visual)
- Head pose estimation (single-view studio data, multi-view lecture data)
- Acoustic scene analysis (acoustic event detection, acoustic environment classification)

Most of these tasks were evaluated on multi-modal multi-site recordings of seminars and meetings provided by the CHIL and VACE projects, as well as on

surveillance data provided by the UK Home Office i-LIDS- Imagery Library for Intelligent Detection Systems program.

Participation in the CLEAR 2006 evaluation and workshop was also open to any site interested in participating in at least one of the evaluation tasks. As a benefit for participating, participating sites would receive the necessary development and evaluation data sets, including scoring tools etc. without cost.

This first CLEAR evaluation and workshop – around 60 people participated in the workshop – was clearly a big success. Overall, nine major evaluation tasks, including more than 20 subtasks, were evaluated. Sixteen different institutions participated in the evaluation, including eight participants from the CHIL program, five participants from the VACE program and three external participants. We were also pleased to have a number of representatives from related evaluation programs and projects give presentations. They were:

- David Cher, (SILOGIC, FR) Topic: Evaluations in ETISEO
- James Ferryman (University of Reading, UK) Topic: PETS Evaluation and Perspective
- Daniel Gatica-Perez (IDIAP, CHE) Topic: Technology Evaluations in AMI
- Mats Ljungqvist (European Commission) Topic: EU-Funded Research Initiatives
- Jonathon Phillips (NIST) Topic: Do Evaluations and Challenge Problems Hinder Creativity?
- Alan Smeaton (Dublin City University, IRL) Topic: TrecVid

Based on the success of CLEAR 2006, it was decided to organize CLEAR 2007 during May 8–9, 2007 in Baltimore, USA. This will again be organized in conjunction with and be collocated with the NIST RT 2007 evaluations, May 10–11, 2007.

Finally, we would like to take this opportunity to thank the sponsoring projects and funding agencies, all the participants of the evaluation and the workshop, the invited speakers and everybody involved in the organization of the evaluations and the workshop.

September 2006

Rainer Stiefelhagen
John Garofolo

Organization

Chairs

Rainer Stiefelhagen
John Garofolo

Universität Karlsruhe, Germany
National Institute of Standards and Technology
(NIST), USA

Workshop Organization

Rachel Bowers
Margit Rödder

NIST
Universität Karlsruhe, Germany

Evaluation Task Organizers

Keni Bernardin
Maurizio Omologo
John Garofolo
Hazim Ekenel
Djamel Mostefa
Aristodemos Pnevmatikakis
Ferran Marques

Universität Karlsruhe, Germany
ITC-IRST, Trento, Italy
NIST
Universität Karlsruhe, Germany
ELDA, Paris, France
Athens Information Technology, Greece
Universitat Politècnica de Catalunya,
Barcelona, Spain
Universitat Politècnica de Catalunya, Spain
Universität Karlsruhe, Germany
Universitat Politècnica de Catalunya, Spain
Carnegie Mellon University, Pittsburgh, PA

Sponsoring Projects and Institutions

Projects:

- CHIL, Computers in the Human Interaction Loop, <http://chil.server.de>
- VACE, Video Analysis Content Extraction,
<https://control.nist.gov/dto/twiki/bin/view/Main/WebHome>

Institutions:

- European Commission, through the Multimodal Interfaces objective of the Information Society Technologies (IST) priority of the Sixth Framework Programme
- US National Institute of Standards and Technology (NIST),
<http://www.nist.gov/speech>

Table of Contents

Overview

The CLEAR 2006 Evaluation	1
<i>Rainer Stiefelhagen, Keni Bernardin, Rachel Bowers, John Garofolo, Djamel Mostefa, and Padmanabhan Soundararajan</i>	

3D Person Tracking

3D Audiovisual Person Tracking Using Kalman Filtering and Information Theory	45
<i>Nikos Katsarakis, George Souretis, Fotios Talantzis, Aristodemos Pnevmatikakis, and Lazaros Polymenakos</i>	
A Generative Approach to Audio-Visual Person Tracking	55
<i>Roberto Brunelli, Alessio Brutti, Paul Chippendale, Oswald Lanz, Maurizio Omologo, Piergiorgio Svaizer, and Francesco Tobia</i>	
An Audio-Visual Particle Filter for Speaker Tracking on the CLEAR'06 Evaluation Dataset	69
<i>Kai Nickel, Tobias Gehrig, Hazim K. Ekenel, John McDonough, and Rainer Stiefelhagen</i>	
Multi- and Single View Multiperson Tracking for Smart Room Environments	81
<i>Keni Bernardin, Tobias Gehrig, and Rainer Stiefelhagen</i>	
UPC Audio, Video and Multimodal Person Tracking Systems in the Clear Evaluation Campaign	93
<i>Alberto Abad, Cristian Canton-Ferrer, Carlos Segura, José Luis Landabaso, Dušan Macho, Josep Ramon Casas, Javier Hernando, Montse Pardàs, and Climent Nadeu</i>	
A Joint System for Single-Person 2D-Face and 3D-Head Tracking in CHIL Seminars	105
<i>Gerasimos Potamianos and Zhenqiu Zhang</i>	
Speaker Tracking in Seminars by Human Body Detection	119
<i>Bo Wu, Vivek Kumar Singh, Ram Nevatia, and Chi-Wei Chu</i>	
TUT Acoustic Source Tracking System 2006	127
<i>Pasi Pertilä, Teemu Korhonen, Tuomo Pirinen, and Mikko Parviainen</i>	

Tracking Multiple Speakers with Probabilistic Data Association
Filters 137
Tobias Gehrig and John McDonough

2D Face Detection and Tracking

2D Person Tracking Using Kalman Filtering and Adaptive Background
Learning in a Feedback Loop 151
Aristodemos Pnevmatikakis and Lazaros Polymenakos

PittPatt Face Detection and Tracking for the CLEAR 2006
Evaluation 161
Michael C. Nechyba and Henry Schneiderman

Person Tracking on Surveillance Data

The AIT Outdoors Tracking System for Pedestrians and Vehicles 171
*Aristodemos Pnevmatikakis, Lazaros Polymenakos, and
Vasileios Mylonakis*

Evaluation of USC Human Tracking System for Surveillance Videos 183
Bo Wu, Xuefeng Song, Vivek Kumar Singh, and Ram Nevatia

Vehicle Tracking

Multi-feature Graph-Based Object Tracking 190
Murtaza Taj, Emilio Maggio, and Andrea Cavallaro

Multiple Vehicle Tracking in Surveillance Videos 200
*Yun Zhai, Phillip Berkowitz, Andrew Miller, Khurram Shafique,
Aniket Vartak, Brandyn White, and Mubarak Shah*

Robust Appearance Modeling for Pedestrian and Vehicle Tracking 209
Wael Abd-Almageed and Larry S. Davis

Robust Vehicle Blob Tracking with Split/Merge Handling 216
Xuefeng Song and Ram Nevatia

Person Identification

A Decision Fusion System Across Time and Classifiers for Audio-Visual
Person Identification 223
*Andreas Stergiou, Aristodemos Pnevmatikakis, and
Lazaros Polymenakos*

The CLEAR'06 LIMSI Acoustic Speaker Identification System for
CHIL Seminars 233
Claude Barras, Xuan Zhu, Jean-Luc Gauvain, and Lori Lamel

Person Identification Based on Multichannel and Multimodality Fusion	241
<i>Ming Liu, Hao Tang, Huazhong Ning, and Thomas Huang</i>	
ISL Person Identification Systems in the CLEAR Evaluations.....	249
<i>Hazim Kemal Ekenel and Qin Jin</i>	
Audio, Video and Multimodal Person Identification in a Smart Room	258
<i>Jordi Luque, Ramon Morros, Ainara Garde, Jan Anguita, Mireia Farrus, Dušan Macho, Ferran Marqués, Claudi Martínez, Verónica Vilaplana, and Javier Hernando</i>	

Head Pose Estimation

Head Pose Estimation on Low Resolution Images	270
<i>Nicolas Gouriér, Jérôme Maisonnasse, Daniela Hall, and James L. Crowley</i>	
Evaluation of Head Pose Estimation for Studio Data	281
<i>Jilin Tu, Yun Fu, Yuxiao Hu, and Thomas Huang</i>	
Neural Network-Based Head Pose Estimation and Multi-view Fusion ...	291
<i>Michael Voit, Kai Nickel, and Rainer Stiefelhofen</i>	
Head Pose Estimation in Seminar Room Using Multi View Face Detectors	299
<i>Zhenqiu Zhang, Yuxiao Hu, Ming Liu, and Thomas Huang</i>	
Head Pose Detection Based on Fusion of Multiple Viewpoint Information	305
<i>Cristian Canton-Ferrer, Josep Ramon Casas, and Montse Pardàs</i>	

Acoustic Scene Analysis

CLEAR Evaluation of Acoustic Event Detection and Classification Systems	311
<i>Andrey Temko, Robert Malkin, Christian Zieger, Dušan Macho, Climent Nadeu, and Maurizio Omologo</i>	
The CLEAR 2006 CMU Acoustic Environment Classification System ...	323
<i>Robert G. Malkin</i>	

Other Evaluations

2D Multi-person Tracking: A Comparative Study in AMI Meetings	331
<i>Kevin Smith, Sascha Schreiber, Igor Potúcek, Vítězslav Beran, Gerhard Rigoll, and Daniel Gatica-Perez</i>	

Head Pose Tracking and Focus of Attention Recognition Algorithms
in Meeting Rooms 345
 Sileye O. Ba and Jean-Marc Odobez

Author Index 359

The CLEAR 2006 Evaluation

Rainer Stiefelhagen¹, Keni Bernardin¹, Rachel Bowers², John Garofolo²,
Djamel Mostefa³, and Padmanabhan Soundararajan⁴

¹ Interactive Systems Lab, Universität Karlsruhe, 76131 Karlsruhe, Germany
{stiefel, keni}@ira.uka.de

² National Institute of Standards and Technology (NIST), Information Technology
Lab - Information Access Division, Speech Group
{rachel.bowers, garofolo}@nist.gov

³ Evaluations and Language Resources Distribution Agency (ELDA), Paris, France
mostefa@elda.org

⁴ Computer Science and Engineering, University of South Florida, Tampa, FL, USA
psoundar@cse.usf.edu

Abstract. This paper is a summary of the first CLEAR evaluation on CClassification of Events, Activities and Relationships - which took place in early 2006 and concluded with a two day evaluation workshop in April 2006. CLEAR is an international effort to evaluate systems for the multimodal perception of people, their activities and interactions. It provides a new international evaluation framework for such technologies. It aims to support the definition of common evaluation tasks and metrics, to coordinate and leverage the production of necessary multimodal corpora and to provide a possibility for comparing different algorithms and approaches on common benchmarks, which will result in faster progress in the research community. This paper describes the evaluation tasks, including metrics and databases used, that were conducted in CLEAR 2006, and provides an overview of the results. The evaluation tasks in CLEAR 2006 included person tracking, face detection and tracking, person identification, head pose estimation, vehicle tracking as well as acoustic scene analysis. Overall, more than 20 subtasks were conducted, which included acoustic, visual and audio-visual analysis for many of the main tasks, as well as different data domains and evaluation conditions.

1 Introduction

CClassification of Events, Activities and Relationships (CLEAR) is an international effort to evaluate systems that are designed to analyze people's identities, activities, interactions and relationships in human-human interaction scenarios, as well as related scenarios. The first CLEAR evaluation has been conducted from around December 2005, when the first development data and scoring scripts were disseminated, until April 2006, when a two-day evaluation workshop took place in Southampton, UK, during which the evaluation results and system details of all participants were discussed.

1.1 Motivation

Many researchers, research labs and in particular a number of current major research projects worldwide – including the European projects CHIL, Computers in the Human Interaction Loop [1], and AMI, “Augmented Multi-party Interaction” [2], as well as the US programs VACE, “Video Analysis Content extraction” [3], and CALO, “Cognitive Assistant that Learns and Organizes” [4] – are working on technologies to analyze people, their activities, and their interaction. However, common evaluation standards for such technologies are missing. Until now, most researchers and research projects use their own different data sets, annotations, task definitions, metrics and evaluation procedures. As a consequence, comparability of the research algorithms and systems is virtually impossible. Furthermore, this leads to a costly multiplication of data production and evaluation efforts for the research community as a whole.

CLEAR was created to address this problem. Its goal is to provide a common international evaluation forum and framework for such technologies, and to serve as a forum for the discussion and definition of related common benchmarks, including the definition of common metrics, tasks and evaluation procedures.

The outcomes for the research community that we expect from such a common evaluation forum are

- the definition of widely adopted common metrics and tasks
- a greater availability of resources by sharing the data collection and annotation burden
- provision of challenging multimodal data sets for the development of robust perceptual technologies
- comparability of systems and approaches and
- thus faster progress in developing better, more robust technology.

1.2 Background

The CLEAR 2006 evaluation has emerged out of the existing evaluation efforts of the European Integrated project CHIL, which has in previous years conducted a number of evaluations on multimodal perceptual technologies, including tasks such as person tracking and identification, head pose estimation, gesture recognition and acoustic event detection, as well as the technology evaluation efforts in the US VACE program, which conducted several similar evaluations in face, person and vehicle tracking. For CLEAR 2006, the technology evaluations of CHIL and VACE were combined for the first time, and the evaluations were also open to any site interested in participating.

In order to broaden the participation and discussion of evaluation tasks and metrics, representatives from other related projects and evaluation efforts (AMI[2], NIST RT evaluations[5], NIST People-ID evaluations, PETs[6], Trec-Vid[7], ETISEO[8]) were actively invited to participate in the preparation of the workshop as well as to present an overview about their related activities at the workshop.

1.3 Scope and Evaluation Tasks in 2006

The CLEAR 2006 evaluation and workshop was organized in conjunction with the National Institute of Standards and Technology (NIST) Rich Transcription (RT) 2006 evaluation [5]. While the evaluations conducted in RT focus on content-related technologies, such as speech and text recognition, CLEAR is more about context-related multimodal technologies such as person tracking, person identification, head pose estimation, analyzing focus of attention, interaction, activities and events. CLEAR 2006 and RT06 in particular shared some of their evaluation data sets, so that for example the speaker-localization results generated for CLEAR could be used for the far-field speech-to-text task in RT06. Also the evaluation deadlines of CLEAR and RT 2006 were harmonized so that this would be possible. This is an important first step towards developing a comprehensive multimedia evaluation program.

The evaluation tasks in CLEAR 2006 can be broken down into four categories:

- tracking tasks (faces/persons/vehicles, 2D/3D, acoustic/visual/audio-visual)
- person identification tasks (acoustic, visual, audio-visual)
- head pose estimation (single view studio data, multi-view lecture data)
- acoustic scene analysis (events, environments)

These tasks and their various subtasks will be described in Section 3.

Due to the short time frame for preparing the joint technology evaluations in CLEAR, it was decided that the evaluations tasks that had already been defined in VACE and CLEAR, respectively, would be kept as they were, and thus were run independently in parallel, with their slightly differing annotations and on different data sets. As a consequence there were, for example, several 3D person tracking tasks (CHIL) as well as 2D person tracking tasks (VACE) in CLEAR 2006. As a first step of harmonizing evaluation tasks, the participants from CHIL and VACE had, however, agreed on common metrics for multiple object tracking (see section 3.3). The aim for upcoming evaluations is to further harmonize metrics and benchmarks.

1.4 Contributors

CLEAR 2006 would not have been possible without the help and effort of many people and institutions worldwide. CLEAR 2006 was supported by the projects CHIL [1] and VACE [3]. The organizers of CLEAR are the Interactive Systems Labs of the Universität Karlsruhe, Germany (UKA), and the US National Institute of Standards and Technology (NIST), with the support of contractors University of South Florida (USF) and VideoMining Inc. The participants and contributors to the CLEAR 2006 evaluations included: the Research and Education Society in Information Technologies at Athens Information Technology, Athens, Greece, (AIT), the Interactive Systems Labs at Carnegie Mellon University, Pittsburgh, PA, USA, (CMU) the Evaluations and Language resources Distribution Agency, Paris, France (ELDA), the IBM T.J. Watson Research

Center, RTE 134, Yorktown Heights, USA (IBM), the Project PRIMA of the Institut National de Recherche en Informatique et en Automatique, Grenoble, France (INRIA), the Centro per la ricerca scientifica e tecnologica at the Instituto Trentino di Cultura, Trento, Italy (ITC-IRST), the Laboratoire d'Informatique pour la mécanique et les sciences de l'ingénieur at the Centre national de la recherche scientifique, Paris, France (LIMSI), Pittsburgh Pattern Recognition, Inc., Pittsburgh, PA, USA (PPATT), the department of Electronic Engineering of the Queen Mary University of London, UK, (QMUL) the Institute of Signal Processing of the Technical University of Tampere, Finland, (TUT), the Breckman Institute for Advanced Science and Tech. at the University of Illinois Urbana Champaign, USA (UIUC) the Institute for Robotics and Intelligent Systems of the University of Southern California, USA, (USC).

UKA and ITC-IRST provided recordings of seminars (lectures), which were used for the 3D single person tracking tasks the face detection task and for person recognition. AIT, IBM and UPC provided several recordings of “interactive” seminars (basically small interactive meetings), which were used for the multi-person tracking tasks, for face detection, for the person identification tasks and for acoustic event detection. INRIA provided the Pointing'04 database for head pose detection. UKA provided 26 seminar recordings with head pose annotations for the lecturer, which data was used for the second head pose estimation task. UPC, ITC and CMU provided different databases with annotated acoustic events used for acoustic event classification. Visual and acoustic annotations of the CHIL seminar and interactive seminar data were mainly done by ELDA, in collaboration with UKA, CMU, AIT, IBM, ITC-irst and UPC. ELDA also packaged and distributed the data coming from CHIL. The data coming from VACE was derived from a single source for the surveillance data - i-LIDS. The meeting room data was a collection derived from data collected at CMU, University of Edinburgh (EDI), NIST, TNO, and Virginia Tech (VT). The discussion and definition of the individual tasks and evaluation procedures were moderated by “task-leaders”. The task-leaders coordinated all aspects surrounding the running of their given tasks. These were Keni Bernardin (UKA, 3D single- and multi-person tracking), Maurizio Omologo (ITC-irst, 3D acoustic single-person tracking), John Garofolo/Rachel Bowers (NIST, 2D Multi-person tracking tasks, VACE 2D face tracking, vehicle tracking), Hazim Ekenel (UKA, visual person identification), Djamel Mostefa (ELDA, acoustic identification), Aristodemos Pnevmatikakis (AIT, audio-visual identification), Ferran Marques and Ramon Morros (both UPC, CHIL 2D Face detection), Michael Voit (UKA, head pose estimation), Andrey Temko (UPC, acoustic event detection). The tasks leaders were also responsible for scoring the evaluation submissions, which in addition were also centrally scored by ELDA.

This paper aims at giving an overview of the CLEAR 2006 evaluation. In the remainder of this paper we will therefore give a brief overview of the data sets used (Section 2) and the various evaluation tasks (Section 3). In Section 4 we present an overview of the results and discuss some of the outcomes and potential implications for further evaluations.