Lecture Notes in Computer Science 4122

*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*

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

Springer

Volume Editors

Rainer Stiefelhagen
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

# 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 | Universität Karlsruhe, Germany |
| John Garofolo | National Institute of Standards and Technology (NIST), USA |

## Workshop Organization

| | |
|---|---|
| Rachel Bowers | NIST |
| Margit Rödder | Universität Karlsruhe, Germany |

## Evaluation Task Organizers

| | |
|---|---|
| Keni Bernardin | Universität Karlsruhe, Germany |
| Maurizio Omologo | ITC-IRST, Trento, Italy |
| John Garofolo | NIST |
| Hazim Ekenel | Universität Karlsruhe, Germany |
| Djamel Mostefa | ELDA, Paris, France |
| Aristodemos Pnevmatikakis | Athens Information Technology, Greece |
| Ferran Marques | Universitat Politecnica de Catalunya, Barcelona, Spain |
| Ramon Morros | Universitat Politècnica de Catalunya, Spain |
| Michael Voit | Universität Karlsruhe, Germany |
| Andrey Temko | Universitat Politècnica de Catalunya, Spain |
| Rob Malkin | 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
*Rainer Stiefelhagen, Keni Bernardin, Rachel Bowers, John Garofolo, Djamel Mostefa, and Padmanabhan Soundararajan*

## 3D Person Tracking

3D Audiovisual Person Tracking Using Kalman Filtering and Information Theory . . . . . 45
*Nikos Katsarakis, George Souretis, Fotios Talantzis, Aristodemos Pnevmatikakis, and Lazaros Polymenakos*

A Generative Approach to Audio-Visual Person Tracking . . . . . 55
*Roberto Brunelli, Alessio Brutti, Paul Chippendale, Oswald Lanz, Maurizio Omologo, Piergiorgio Svaizer, and Francesco Tobia*

An Audio-Visual Particle Filter for Speaker Tracking on the CLEAR'06 Evaluation Dataset . . . . . 69
*Kai Nickel, Tobias Gehrig, Hazim K. Ekenel, John McDonough, and Rainer Stiefelhagen*

Multi- and Single View Multiperson Tracking for Smart Room Environments . . . . . 81
*Keni Bernardin, Tobias Gehrig, and Rainer Stiefelhagen*

UPC Audio, Video and Multimodal Person Tracking Systems in the Clear Evaluation Campaign . . . . . 93
*Alberto Abad, Cristian Canton-Ferrer, Carlos Segura, José Luis Landabaso, Dušan Macho, Josep Ramon Casas, Javier Hernando, Montse Pardàs, and Climent Nadeu*

A Joint System for Single-Person 2D-Face and 3D-Head Tracking in CHIL Seminars . . . . . 105
*Gerasimos Potamianos and Zhenqiu Zhang*

Speaker Tracking in Seminars by Human Body Detection . . . . . 119
*Bo Wu, Vivek Kumar Singh, Ram Nevatia, and Chi-Wei Chu*

TUT Acoustic Source Tracking System 2006 . . . . . 127
*Pasi Pertilä, Teemu Korhonen, Tuomo Pirinen, and Mikko Parviainen*

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
*Ming Liu, Hao Tang, Huazhong Ning, and Thomas Huang*

ISL Person Identification Systems in the CLEAR Evaluations ..... 249
*Hazım Kemal Ekenel and Qin Jin*

Audio, Video and Multimodal Person Identification in a Smart Room ..... 258
*Jordi Luque, Ramon Morros, Ainara Garde, Jan Anguita, Mireia Farrus, Dušan Macho, Ferran Marqués, Claudi Martínez, Verónica Vilaplana, and Javier Hernando*

## Head Pose Estimation

Head Pose Estimation on Low Resolution Images ..... 270
*Nicolas Gourier, Jérôme Maisonnasse, Daniela Hall, and James L. Crowley*

Evaluation of Head Pose Estimation for Studio Data ..... 281
*Jilin Tu, Yun Fu, Yuxiao Hu, and Thomas Huang*

Neural Network-Based Head Pose Estimation and Multi-view Fusion ... 291
*Michael Voit, Kai Nickel, and Rainer Stiefelhagen*

Head Pose Estimation in Seminar Room Using Multi View Face Detectors ..... 299
*Zhenqiu Zhang, Yuxiao Hu, Ming Liu, and Thomas Huang*

Head Pose Detection Based on Fusion of Multiple Viewpoint Information ..... 305
*Cristian Canton-Ferrer, Josep Ramon Casas, and Montse Pardàs*

## Acoustic Scene Analysis

CLEAR Evaluation of Acoustic Event Detection and Classification Systems ..... 311
*Andrey Temko, Robert Malkin, Christian Zieger, Dušan Macho, Climent Nadeu, and Maurizio Omologo*

The CLEAR 2006 CMU Acoustic Environment Classification System ... 323
*Robert G. Malkin*

## Other Evaluations

2D Multi-person Tracking: A Comparative Study in AMI Meetings ..... 331
*Kevin Smith, Sascha Schreiber, Igor Potúcek, Vítezslav Beran, Gerhard Rigoll, and Daniel Gatica-Perez*

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

**Author Index** ..... 359