Thomas S. Huang Nicu Sebe
Michael S. Lew Vladimir Pavlović
Mathias Kölsch Aphrodite Galata
Branislav Kisačanin (Eds.)

LNCS 3979

Computer Vision in Human-Computer Interaction

ECCV 2006 Workshop on HCI
Graz, Austria, May 2006
Proceedings

Springer

Thomas S. Huang Nicu Sebe
Michael S. Lew Vladimir Pavlović
Mathias Kölsch Aphrodite Galata
Branislav Kisačanin (Eds.)

Computer Vision in Human-Computer Interaction

ECCV 2006 Workshop on HCI
Graz, Austria, May 13, 2006
Proceedings

Springer

Volume Editors

Thomas S. Huang
E-mail: huang@ifp.uiuc.edu

Nicu Sebe
E-mail: nicu@science.uva.nl

Michael S. Lew
E-mail: mlew@liacs.nl

Vladimir Pavlović
E-mail: vladimir@cs.rutgers.edu

Mathias Kölsch
E-mail: kolsch@nps.edu

Aphrodite Galata
E-mail: a.galata@cs.man.ac.uk

Branislav Kisačanin
E-mail: b.kisacanin@ieee.org

Library of Congress Control Number: 2006925105

CR Subject Classification (1998): I.4, I.5, I.3, H.5.2, K.4.2

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

ISSN 0302-9743
ISBN-10 3-540-34202-8 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-34202-1 Springer Berlin Heidelberg New York

Springer is a part of Springer Science+Business Media

springer.com

Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 11754336 06/3142 5 4 3 2 1 0

Lecture Notes in Computer Science 3979

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

Preface

The interests and goals of HCI (human–computer interaction) include understanding, designing, building, and evaluating complex interactive systems involving many people and technologies. Developments in software and hardware technologies are continuously driving applications in supporting our collaborative and communicative needs as social beings, both at work and at play. At the same time, similar developments are pushing the human–computer interface beyond the desktop and into our pockets, streets, and buildings. Developments in mobile, wearable, and pervasive communications and computing technologies provide exciting challenges and opportunities for HCI.

The present volume represents the proceedings of the HCI 2006 Workshop that was held in conjunction with ECCV 2006 (European Conference on Computer Vision) in Graz, Austria. The goal of this workshop was to bring together researchers from the field of computer vision whose work is related to human–computer interaction. We solicited original contributions that address a wide range of theoretical and application issues in human–computer interaction.

We were very pleased by the response and had a difficult task of selecting only 11 papers (out of 27 submitted) to be presented at the workshop. The accepted papers were presented in four sessions, as follows:

Face Analysis

- In their paper "Robust Face Alignment Based On Hierarchical Classifier Network" authors Li Zhang, Haizhou Ai, and Shihong Lao build a hierarchical classifier network that connects face detection and face alignment into a smooth coarse-to-fine procedure. Thus a robust face alignment algorithm on face images with expression and pose changes is introduced. Experiments are reported to show its accuracy and robustness.
- In "EigenExpress Approach in Recognition of Facial Expression using GPU" authors Qi Wu, Mingli Song, Jiajun Bu, and Chun Chen present an efficient facial expression recognition system based on a GPU-based filter for preprocessing and EigenExpress and Modified Hausdorff distance for classification.
- In "Face Representation Method Using Pixel-to-Vertex Map for 3D Model-Based Face Recognition" authors Taehwa Hong, Hagbae Kim, Hyeonjoon Moon, Yongguk Kim, Jongweon Lee, and Seungbin Moon describe a 3D face representation algorithm to reduce the number of vertices and optimize its computation time. They evaluate the performance of the proposed algorithm with the Korean face database collected using a stereo-camera-based 3D face capturing device.
- In "Robust Head Tracking with Particles Based on Multiple Cues Fusion" authors Yuan Li, Haizhou Ai, Chang Huang, and Shihong Lao present a fully automatic and highly robust head tracking algorithm that fuses the face cues from a real-time multiview face detection with color spatiogram and contour

gradient cues under a particle filter framework. Experiments show that this algorithm is highly robust against target position, size, and pose change, as well as unfavorable conditions such as occlusion, poor illumination, and cluttered background.

Gesture and Emotion Recognition

- In "Vision-Based Interpretation of Hand Gestures for Remote Control of a Computer Mouse" authors Antonis A. Argyros and Manolis I. A. Lourakis present a human–computer interaction system that is capable of recognizing hand gestures and of interpreting them to remotely control a computer mouse. This work is based on their previous work on 2D and 3D tracking of colored objects. Two different gestural vocabularies are investigated, based on 2D and 3D hand information, respectively. Experiments are used to compare these vocabularies in terms of efficiency, robustness, reliability, and ease of use.
- In "Computing Emotion Awareness Through Facial Electromyography" authors Egon van den Broek, Marleen Schut, Joyce Westerink, Jan van Herk, and Kees Tuinenbreijer use coarse time windows to discriminate between positive, negative, neutral, and mixed emotions. They use six parameters (i.e., mean, absolute deviation, standard deviation, variance, skewness, and kurtosis) of three facial EMGs: zygomaticus major, corrugator supercilii, and frontalis. The zygomaticus major is shown to discriminate excellently between the four emotion categories and, consequently, can facilitate empathic HCI.

Event Detection

- In "Silhouette-Based Method for Object Classification and Human Action Recognition in Video" authors Yiğithan Dedeoğlu, B. Uğur Töreyin, Uğur Güdükbay, and A. Enis Çetin present an instance-based machine learning algorithm and a system for real-time object classification and human action recognition which makes use of object silhouettes. An adaptive background subtraction model is used for object segmentation. A supervised learning method based on template matching is adopted to classify objects into classes like human, human group, and vehicle, and human actions into predefined classes like walking, boxing, and kicking.
- In "Voice Activity Detection Using Wavelet-Based Multiresolution Spectrum and Support Vector Machines and Audio Mixing Algorithm" authors Wei Xue, Sidan Du, Chengzhi Fang, and Yingxian Ye present a voice activity detection (VAD) algorithm and efficient speech mixing algorithm for a multimedia conference. The proposed VAD uses MFCC of multiresolution spectrum as features and classifies voice by support vector machines (SVM).
- In "Action Recognition in Broadcast Tennis Video Using Optical Flow and Support Vector Machine" authors Guangyu Zhu, Changsheng Xu, Wen Gao, and Qingming Huang present a novel approach to recognize the basic player actions in broadcast tennis video where the player is only about 30 pixels

tall. A new motion descriptor based on optical flow is proposed where the optical flow is treated as spatial patterns of noisy measurements instead of precise pixel displacements. Support vector machine is employed to train the action classifier.

Applications

- In "FaceMouse — A Human-Computer Interface for Tetraplegic People" authors Emanuele Perini, Simone Soria, Andrea Prati, and Rita Cucchiara propose a new human–machine interface particularly conceived for people with severe disabilities (specifically tetraplegic people), that allows them to interact with the computer. They have studied a new paradigm called "derivative paradigm," where the users indicate the direction along which the mouse pointer must be moved. The system that uses this paradigm consists of a common, low-cost webcam and a set of computer vision techniques developed to identify the parts of the user's face and exploit them for moving the pointer.
- In "Object Retrieval by Query with Sensibility Based on the KANSEI-Vocabulary Scale" authors Sunkyoung Baek, Myunggwon Hwang, Miyoung Cho, Chang Choi, and Pankoo Kim propose the KANSEI-Vocabulary Scale by associating human sensibilities with shapes among visual information. They construct the object retrieval system for evaluation of their approach and are able to retrieve object images with the most appropriate shape in terms of the query's sensibility.

We would like to thank the contributing authors and Springer's LNCS team for their help in preparation of the workshop proceedings. There would not be a workshop to begin with had it not been for the invaluable help we received from the Program Committee members (listed later in the book) and their careful reviews of submitted papers. The review process has been facilitated by the Conference Management Toolkit, a free service provided by Microsoft Research (`http://msrcmt.research.microsoft.com/cmt`). We would also like to thank the Chairs of the ECCV 2006 Conference in Graz, Austria, for their support and help. Finally, we would like to thank our corporate sponsor, Delphi Corporation, for generous support of the workshop.

May 2006
Graz, Austria

T.S. Huang
N. Sebe
M.S. Lew
V. Pavlović
M. Kölsch
A. Galata
B. Kisačanin
HCI 2006 Chairs

Organization

HCI 2006 (Workshop on Human–Computer Interaction) was held in conjunction with ECCV 2006 (European Conference on Computer Vision), on 13 May 2006, in Graz, Austria.

Organizing Committee

General Chair	Thomas S. Huang (University of Illinois at Urbana-Champaign, USA)
Program Chairs	Nicu Sebe (University of Amsterdam, Netherlands)
	Michael S. Lew (University of Leiden, Netherlands)
	Vladimir Pavlović (Rutgers University, USA)
	Mathias Kölsch (Naval Postgraduate School, USA)
Publicity Chairs	Aphrodite Galata (University of Manchester, UK)
	Branislav Kisačanin (Delphi Corporation, USA)

Program Committee

K. Aizawa
A. del Bimbo
N. Boujemaa
I. Cohen
J. Cohn
J. Crowley
D. Gatica-Perez
T. Gevers
A. Hanjalic
A. Jaimes
A. Kapoor
M. Nixon
M. Pantic
I. Patras
A. Pentland
S. Sclaroff
Q. Tian
M. Turk
J. Vitria
G. Xu
M. Yang
X. Zhou

Corporate Sponsor

Delphi Corporation, USA
www.delphi.com

Vol. 3927: J. Hespanha, A. Tiwari (Eds.), Hybrid Systems: Computation and Control. XII, 584 pages. 2006.

Vol. 3925: A. Valmari (Ed.), Model Checking Software. X, 307 pages. 2006.

Vol. 3924: P. Sestoft (Ed.), Programming Languages and Systems. XII, 343 pages. 2006.

Vol. 3923: A. Mycroft, A. Zeller (Eds.), Compiler Construction. XIII, 277 pages. 2006.

Vol. 3922: L. Baresi, R. Heckel (Eds.), Fundamental Approaches to Software Engineering. XIII, 427 pages. 2006.

Vol. 3921: L. Aceto, A. Ingólfsdóttir (Eds.), Foundations of Software Science and Computation Structures. XV, 447 pages. 2006.

Vol. 3920: H. Hermanns, J. Palsberg (Eds.), Tools and Algorithms for the Construction and Analysis of Systems. XIV, 506 pages. 2006.

Vol. 3918: W.K. Ng, M. Kitsuregawa, J. Li, K. Chang (Eds.), Advances in Knowledge Discovery and Data Mining. XXIV, 879 pages. 2006. (Sublibrary LNAI).

Vol. 3917: H. Chen, F.Y. Wang, C.C. Yang, D. Zeng, M. Chau, K. Chang (Eds.), Intelligence and Security Informatics. XII, 186 pages. 2006.

Vol. 3916: J. Li, Q. Yang, A.-H. Tan (Eds.), Data Mining for Biomedical Applications. VIII, 155 pages. 2006. (Sublibrary LNBI).

Vol. 3915: R. Nayak, M.J. Zaki (Eds.), Knowledge Discovery from XML Documents. VIII, 105 pages. 2006.

Vol. 3914: A. Garcia, R. Choren, C. Lucena, P. Giorgini, T. Holvoet, A. Romanovsky (Eds.), Software Engineering for Multi-Agent Systems IV. XIV, 255 pages. 2006.

Vol. 3910: S.A. Brueckner, G.D.M. Serugendo, D. Hales, F. Zambonelli (Eds.), Engineering Self-Organising Systems. XII, 245 pages. 2006. (Sublibrary LNAI).

Vol. 3909: A. Apostolico, C. Guerra, S. Istrail, P. Pevzner, M. Waterman (Eds.), Research in Computational Molecular Biology. XVII, 612 pages. 2006. (Sublibrary LNBI).

Vol. 3908: A. Bui, M. Bui, T. Böhme, H. Unger (Eds.), Innovative Internet Community Systems. VIII, 207 pages. 2006.

Vol. 3907: F. Rothlauf, J. Branke, S. Cagnoni, E. Costa, C. Cotta, R. Drechsler, E. Lutton, P. Machado, J.H. Moore, J. Romero, G.D. Smith, G. Squillero, H. Takagi (Eds.), Applications of Evolutionary Computing. XXIV, 813 pages. 2006.

Vol. 3906: J. Gottlieb, G.R. Raidl (Eds.), Evolutionary Computation in Combinatorial Optimization. XI, 293 pages. 2006.

Vol. 3905: P. Collet, M. Tomassini, M. Ebner, S. Gustafson, A. Ekárt (Eds.), Genetic Programming. XI, 361 pages. 2006.

Vol. 3904: M. Baldoni, U. Endriss, A. Omicini, P. Torroni (Eds.), Declarative Agent Languages and Technologies III. XII, 245 pages. 2006. (Sublibrary LNAI).

Vol. 3903: K. Chen, R. Deng, X. Lai, J. Zhou (Eds.), Information Security Practice and Experience. XIV, 392 pages. 2006.

Vol. 3901: P.M. Hill (Ed.), Logic Based Program Synthesis and Transformation. X, 179 pages. 2006.

Vol. 3900: F. Toni, P. Torroni (Eds.), Computational Logic in Multi-Agent Systems. XVII, 427 pages. 2006. (Sublibrary LNAI).

Vol. 3899: S. Frintrop, VOCUS: A Visual Attention System for Object Detection and Goal-Directed Search. XIV, 216 pages. 2006. (Sublibrary LNAI).

Vol. 3898: K. Tuyls, P.J. 't Hoen, K. Verbeeck, S. Sen (Eds.), Learning and Adaption in Multi-Agent Systems. X, 217 pages. 2006. (Sublibrary LNAI).

Vol. 3897: B. Preneel, S. Tavares (Eds.), Selected Areas in Cryptography. XI, 371 pages. 2006.

Vol. 3896: Y. Ioannidis, M.H. Scholl, J.W. Schmidt, F. Matthes, M. Hatzopoulos, K. Boehm, A. Kemper, T. Grust, C. Boehm (Eds.), Advances in Database Technology - EDBT 2006. XIV, 1208 pages. 2006.

Vol. 3895: O. Goldreich, A.L. Rosenberg, A.L. Selman (Eds.), Theoretical Computer Science. XII, 399 pages. 2006.

Vol. 3894: W. Grass, B. Sick, K. Waldschmidt (Eds.), Architecture of Computing Systems - ARCS 2006. XII, 496 pages. 2006.

Vol. 3893: L. Atzori, D.D. Giusto, R. Leonardi, F. Pereira (Eds.), Visual Content Processing and Representation. IX, 224 pages. 2006.

Vol. 3892: A. Carbone, N.A. Pierce (Eds.), DNA Computing. XI, 440 pages. 2006.

Vol. 3891: J.S. Sichman, L. Antunes (Eds.), Multi-Agent-Based Simulation VI. X, 191 pages. 2006. (Sublibrary LNAI).

Vol. 3890: S.G. Thompson, R. Ghanea-Hercock (Eds.), Defence Applications of Multi-Agent Systems. XII, 141 pages. 2006. (Sublibrary LNAI).

Vol. 3889: J. Rosca, D. Erdogmus, J.C. Príncipe, S. Haykin (Eds.), Independent Component Analysis and Blind Signal Separation. XXI, 980 pages. 2006.

Vol. 3888: D. Draheim, G. Weber (Eds.), Trends in Enterprise Application Architecture. IX, 145 pages. 2006.

Vol. 3887: J.R. Correa, A. Hevia, M. Kiwi (Eds.), LATIN 2006: Theoretical Informatics. XVI, 814 pages. 2006.

Vol. 3886: E.G. Bremer, J. Hakenberg, E.-H.(S.) Han, D. Berrar, W. Dubitzky (Eds.), Knowledge Discovery in Life Science Literature. XIV, 147 pages. 2006. (Sublibrary LNBI).

Vol. 3885: V. Torra, Y. Narukawa, A. Valls, J. Domingo-Ferrer (Eds.), Modeling Decisions for Artificial Intelligence. XII, 374 pages. 2006. (Sublibrary LNAI).

Vol. 3884: B. Durand, W. Thomas (Eds.), STACS 2006. XIV, 714 pages. 2006.

Vol. 3882: M.L. Lee, K.-L. Tan, V. Wuwongse (Eds.), Database Systems for Advanced Applications. XIX, 923 pages. 2006.

Vol. 3881: S. Gibet, N. Courty, J.-F. Kamp (Eds.), Gesture in Human-Computer Interaction and Simulation. XIII, 344 pages. 2006. (Sublibrary LNAI).

Vol. 3880: A. Rashid, M. Aksit (Eds.), Transactions on Aspect-Oriented Software Development I. IX, 335 pages. 2006.

Vol. 3879: T. Erlebach, G. Persinao (Eds.), Approximation and Online Algorithms. X, 349 pages. 2006.

¥359.00元

Lecture Notes in Computer Science

For information about Vols. 1–3878

please contact your bookseller or Springer

Vol. 3987: M. Hazas, J. Krumm, T. Strang (Eds.), Location- and Context-Awareness. X, 289 pages. 2006.

Vol. 3984: M. Gavrilova, O. Gervasi, V. Kumar, C.J. K. Tan, D. Taniar, A. Laganà, Y. Mun, H. Choo (Eds.), Computational Science and Its Applications - ICCSA 2006, Part V. XXV, 1045 pages. 2006.

Vol. 3983: M. Gavrilova, O. Gervasi, V. Kumar, C.J. K. Tan, D. Taniar, A. Laganà, Y. Mun, H. Choo (Eds.), Computational Science and Its Applications - ICCSA 2006, Part IV. XXVI, 1191 pages. 2006.

Vol. 3982: M. Gavrilova, O. Gervasi, V. Kumar, C.J. K. Tan, D. Taniar, A. Laganà, Y. Mun, H. Choo (Eds.), Computational Science and Its Applications - ICCSA 2006, Part III. XXV, 1243 pages. 2006.

Vol. 3981: M. Gavrilova, O. Gervasi, V. Kumar, C.J. K. Tan, D. Taniar, A. Laganà, Y. Mun, H. Choo (Eds.), Computational Science and Its Applications - ICCSA 2006, Part II. XXVI, 1255 pages. 2006.

Vol. 3980: M. Gavrilova, O. Gervasi, V. Kumar, C.J. K. Tan, D. Taniar, A. Laganà, Y. Mun, H. Choo (Eds.), Computational Science and Its Applications - ICCSA 2006, Part I. LXXV, 1199 pages. 2006.

Vol. 3979: T.S. Huang, N. Sebe, M.S. Lew, V. Pavlović, M. Kölsch, A. Galata, B. Kisačanin (Eds.), Computer Vision in Human-Computer Interaction. XII, 121 pages. 2006.

Vol. 3978: B. Hnich, M. Carlsson, F. Fages, F. Rossi (Eds.), Recent Advances in Constraints. VIII, 179 pages. 2006. (Sublibrary LNAI).

Vol. 3970: T. Braun, G. Carle, S. Fahmy, Y. Koucheryavy (Eds.), Wired/Wireless Internet Communications. XIV, 350 pages. 2006.

Vol. 3968: K.P. Fishkin, B. Schiele, P. Nixon, A. Quigley (Eds.), Pervasive Computing. XV, 402 pages. 2006.

Vol. 3964: M. Ü. Uyar, A.Y. Duale, M.A. Fecko (Eds.), Testing of Communicating Systems. XI, 373 pages. 2006.

Vol. 3960: R. Vieira, P. Quaresma, M.d.G.V. Nunes, N.J. Mamede, C. Oliveira, M.C. Dias (Eds.), Computational Processing of the Portuguese Language. XII, 274 pages. 2006. (Sublibrary LNAI).

Vol. 3959: J.-Y. Cai, S. B. Cooper, A. Li (Eds.), Theory and Applications of Models of Computation. XV, 794 pages. 2006.

Vol. 3958: M. Yung, Y. Dodis, A. Kiayias, T. Malkin (Eds.), Public Key Cryptography - PKC 2006. XIV, 543 pages. 2006.

Vol. 3956: G. Barthe, B. Gregoire, M. Huisman, J.-L. Lanet (Eds.), Construction and Analysis of Safe, Secure, and Interoperable Smart Devices. IX, 175 pages. 2006.

Vol. 3955: G. Antoniou, G. Potamias, C. Spyropoulos, D. Plexousakis (Eds.), Advances in Artificial Intelligence. XVII, 611 pages. 2006. (Sublibrary LNAI).

Vol. 3954: A. Leonardis, H. Bischof, A. Pinz (Eds.), Computer Vision – ECCV 2006, Part IV. XVII, 613 pages. 2006.

Vol. 3953: A. Leonardis, H. Bischof, A. Pinz (Eds.), Computer Vision – ECCV 2006, Part III. XVII, 649 pages. 2006.

Vol. 3952: A. Leonardis, H. Bischof, A. Pinz (Eds.), Computer Vision – ECCV 2006, Part II. XVII, 661 pages. 2006.

Vol. 3951: A. Leonardis, H. Bischof, A. Pinz (Eds.), Computer Vision – ECCV 2006, Part I. XXXV, 639 pages. 2006.

Vol. 3950: J.P. Müller, F. Zambonelli (Eds.), Agent-Oriented Software Engineering VI. XVI, 249 pages. 2006.

Vol. 3947: Y.-C. Chung, J.E. Moreira (Eds.), Advances in Grid and Pervasive Computing. XXI, 667 pages. 2006.

Vol. 3946: T.R. Roth-Berghofer, S. Schulz, D.B. Leake (Eds.), Modeling and Retrieval of Context. XI, 149 pages. 2006. (Sublibrary LNAI).

Vol. 3945: M. Hagiya, P. Wadler (Eds.), Functional and Logic Programming. X, 295 pages. 2006.

Vol. 3944: J. Quiñonero-Candela, I. Dagan, B. Magnini, F. d'Alché-Buc (Eds.), Machine Learning Challenges. XIII, 462 pages. 2006. (Sublibrary LNAI).

Vol. 3942: Z. Pan, R. Aylett, H. Diener, X. Jin, S. Göbel, L. Li (Eds.), Technologies for E-Learning and Digital Entertainment. XXV, 1396 pages. 2006.

Vol. 3939: C. Priami, L. Cardelli, S. Emmott (Eds.), Transactions on Computational Systems Biology IV. VII, 141 pages. 2006. (Sublibrary LNBI).

Vol. 3936: M. Lalmas, A. MacFarlane, S. Rüger, A. Tombros, T. Tsikrika, A. Yavlinsky (Eds.), Advances in Information Retrieval. XIX, 584 pages. 2006.

Vol. 3935: D. Won, S. Kim (Eds.), Information Security and Cryptology - ICISC 2005. XIV, 458 pages. 2006.

Vol. 3934: J.A. Clark, R.F. Paige, F.A. C. Polack, P.J. Brooke (Eds.), Security in Pervasive Computing. X, 243 pages. 2006.

Vol. 3933: F. Bonchi, J.-F. Boulicaut (Eds.), Knowledge Discovery in Inductive Databases. VIII, 251 pages. 2006.

Vol. 3931: B. Apolloni, M. Marinaro, G. Nicosia, R. Tagliaferri (Eds.), Neural Nets. XIII, 370 pages. 2006.

Vol. 3930: D.S. Yeung, Z.-Q. Liu, X.-Z. Wang, H. Yan (Eds.), Advances in Machine Learning and Cybernetics. XXI, 1110 pages. 2006. (Sublibrary LNAI).

Vol. 3929: W. MacCaull, M. Winter, I. Düntsch (Eds.), Relational Methods in Computer Science. VIII, 263 pages. 2006.

Vol. 3928: J. Domingo-Ferrer, J. Posegga, D. Schreckling (Eds.), Smart Card Research and Advanced Applications. XI, 359 pages. 2006.

Table of Contents

Computer Vision in Human-Computer Interaction

Robust Face Alignment Based on Hierarchical Classifier Network

Li Zhang[1], Haizhou Ai[1], and Shihong Lao[2]

[1] Department of Computer Science, Tsinghua University, Beijing 100084, China
[2] Sensing and Control Technology Lab, Omron Corporation, Kyoto 619-0283, Japan
ahz@mail.tsinghua.edu.cn

Abstract. Robust face alignment is crucial for many face processing applications. As face detection only gives a rough estimation of face region, one important problem is how to align facial shapes starting from this rough estimation, especially on face images with expression and pose changes. We propose a novel method of face alignment by building a hierarchical classifier network, connecting face detection and face alignment into a smooth coarse-to-fine procedure. Classifiers are trained to recognize feature textures in different scales from entire face to local patterns. A multi-layer structure is employed to organize the classifiers, which begins with one classifier at the first layer and gradually refines the localization of feature points by more classifiers in the following layers. A Bayesian framework is configured for the inference of the feature points between the layers. The boosted classifiers detects facial features discriminately from its local neighborhood, while the inference between the layers constrains the searching space. Extensive experiments are reported to show its accuracy and robustness.

1 Introduction

Face alignment, whose objective is to localize the feature points on face images such as the contour points of eyes, noses, mouths and outlines, plays a fundamental role in many face processing tasks. The shape and texture of the feature points acquired by the alignment provide very helpful information for applications such as face recognition, modeling and synthesis. However, since the shape of the face may vary largely in practical images due to differences in age, expression and etc, a robust alignment algorithm, especially against errant initialization and face shape variation, is still a goal to achieve.

There have been many studies on face alignment in the recent decade, most of which were based on Active Shape Model (ASM) and Active Appearance Model (AAM), proposed by Cootes et al [1]. In all these improvements, local or global texture features are employed to guide an iterative optimization of label points under the constraint of a statistical shape model. Many different types of features such as Gabor[2], Haar wavelet[3], and machine learning methods such as Ada-Boosting[4, 5], k-NN[6] have been employed to replace the gradient

T.S. Huang et al. (Eds.): HCI/ECCV 2006, LNCS 3979, pp. 1–11, 2006.

feature and simple gaussian model in the classical ASM methods, improving the robustness of the texture feature. Besides, different methods of optimization such as weighted least-square[7, 8], statistical inference[9, 10] and optical flows[11] have been carried out to improve the efficiency of convergence.

However, most of these methods do not pay much attention to the initialization of the alignment, which strongly affects the performance of the alignment. As face detection algorithms only give a rough position of the face, it is often difficult to estimate all the feature points properly in initialization, especially for face images with expression and pose changes. With a bad initialization, the iterative optimization of both ASM and AAM will be stuck in local minima, and the alignment will fail.

To overcome this deficiency, we propose a novel method by building a hierarchical local texture classifier network, connecting face detection and face alignment into a smooth coarse-to-fine procedure. The algorithm is motivated by the following idea: since the texture patterns of feature points are often distinctive from their neighboring non-feature texture, localization of these feature textures can be considered as a pattern recognition task, like face detection task. Considering the face detection as the coarsest texture classifier at the first layer, a hierarchical structure can be settled to gradually refines the localization of feature points by more classifiers in the following layers. A Bayesian framework is configured for the inference of the feature points between the layers. Both the classifiers in different scales and the inter-layer inference are helpful for avoiding the local-minima problems.

There have been a previous work [5] trying to solve face alignment using face detection technique. Boosted classifiers, which have been widely used to recognize the face pattern, were introduced to recognize smaller texture patterns for every facial feature point. Compared with this work, our method not only employed boosted classifiers but also connect the face detection and alignment organically, further clarifying the coarse-to-fine relation between the two tasks.

There are also some other previous works employing hierarchical approaches. C.Liu et al.[10] introduced a hierarchical data driven Markov chain Monte Carlo (HDDMCMC) method to deduce the inter-layer correlation, while F. Jiao et al. [2] used Gabor wavelet in multi-frequency to characterize the texture feature. Many other methods used a multi-resolution strategy in their implementations, most of which simply take the alignment result in the low-resolution images as the initialization for high-resolution images. Compared with these methods, our method emphasizes both the inference of the inter-layer correlation and the local texture model, so as to achieve both robustness and efficiency.

The rest of the paper is organized as follows: In Section 2, classifiers based on boosted Haar-like feature are introduced to model the likelihood of feature texture patterns. In Section 3, hierarchical structure and algorithm of feature point selection are presented to organize these classifiers, and a Bayesian framework is configured for shape parameter inference between the layers. Extensive experiments are reported in Section 4. Finally, conclusions are drawn in Section 5.

2 Feature Texture Classifier

In order to localize the exact position of a feature point, its texture pattern should be modeled. Classical ASM method only uses a vector of gradient perpendicular to the contour to represent the feature, and characterizes it with Principle Component Analysis (PCA)[1]. Since this 1D profile feature and PCA are too simple, the localization can fall into local minima.

In our work, to make the local texture model more discriminative against non-feature texture patterns, we propose to learn the local texture classifiers by boosting weak classifiers based on Haar-like rectangle features. The boosted classifier is capable of capturing complicated texture pattern, such as human faces [12] and facial features[5]. Through Real AdaBoost learning [13], sample weights are adapted to select and combine weak classifiers into a strong one as,

$$conf(x) = \sum_{t=1}^{T} h_t(x) - b \tag{1}$$

Thus, for each feature point i we will have a strong classifier $Conf_i(x)$. Given a feature pattern x, the strong classifier gives highly discriminative output values which can detect the corresponding feature point.

The classifiers can be trained to recognize facial feature texture in different scales from entire face to local patterns. And each classifier can be reinforced

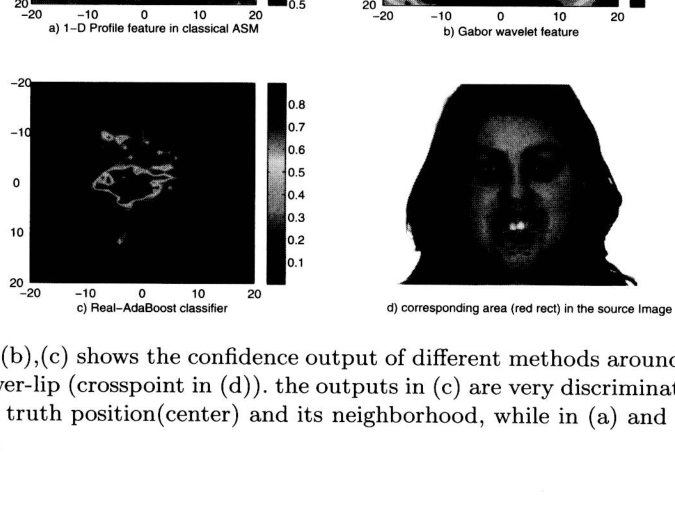


Fig. 1. (a),(b),(c) shows the confidence output of different methods around the feature point of lower-lip (crosspoint in (d)). the outputs in (c) are very discriminative between the ground truth position(center) and its neighborhood, while in (a) and (b) they are ambiguous.

by using a cascade of several classifiers [12]. For large pattern like entire face, an exhaustive searching on the image can find any possible candidate of the feature pattern by maximizing the likelihood output of the classifier. However, for localization of small patterns like facial feature, the large image space will make exhaustive searching time-consuming, and the output of classifiers would not be reliable in such large space. Hence, it is also important to constrain the searching space and maintain the geometry shape formed by the feature points, which we will discuss in Section 3.

3 Hierarchical Classifier Network

To connect the task of face detection and face alignment,let us consider them as two opposite bounds of a continuum (Fig.2). Shown on the left, the face detection task can explore large image space to find face regions. Its advantage is the high robustness against the complexity of background and the variation of face texture, while the detailed facial feature points are not aligned accurately. On the right, the face alignment task can localize each facial feature point accurately, given a good initialization. Inspired by the observation, we propose a coarse-to-fine structure combining the two tasks together to localize the facial shape starting from simple face detection rectangle.

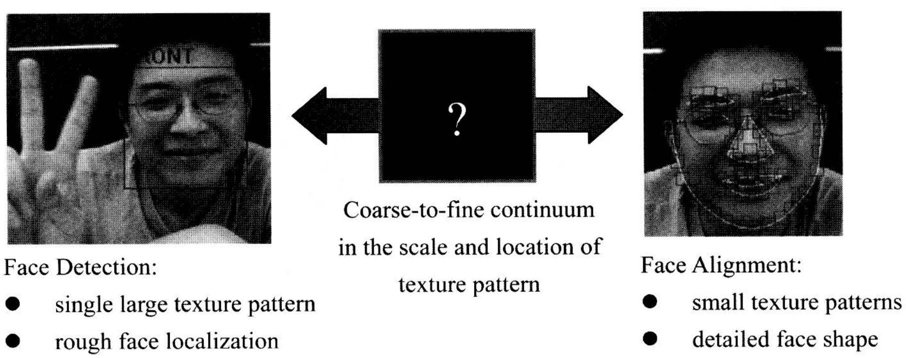


Fig. 2. Continuum of Face Detection and Face Alignment

3.1 Hierarchical Structure

The facial shape denoted by a vector $S^{(K)} = [x_1, y_1, \dots, x_N, y_N]$, can be modeled by shape pose parameter p, q, as

$$S^{(K)} = T_q(\overline{S} + U \cdot p) \tag{2}$$

where p is the parameter of the point distribution model constructed by PCA with average shape $\overline{S}$ and eigenvectors U; $T_q(s)$ is the geometrical transformation based on 4 parameters: scale, rotation, and translation [1].

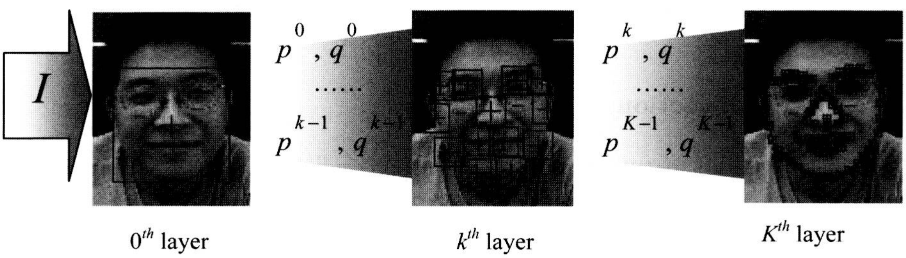
Given an image I as input, this coarse-to-fine procedure is illustrated in Fig.3: At the first layer, one single classifier is trained as a face detector to localize the central point (x_c, y_c), rotation angle θ and size W of the face. The face size is then normalized to L pixels in width, and p^0, q^0 are estimated through a MAP(max-a-posterior) inference,

$$\underset{p^0,q^0}{\text{argmax}}\, P(x_c, y_c, \theta, W|p, q)P(p^0)P(q^0) \tag{3}$$

Then for the following layers, k^{th} layer for example, the further alignment task is divided into several sub-tasks localizing the feature points defined by point set $S^{(k)} = [x_i^{(k)}, y_i^{(k)}], i = 1..n_k$. And the geometry shape of $S^{(k)}$ is modeled by the parameters $p^{(k)}$ and $q^{(k)}$, as

$$S^{(k)} = A^{(k)} \cdot T_{q^{(k)}}(\overline{S} + U \cdot p^{(k)}) \tag{4}$$

which represents the new feature points as a linear combination of original feature points using $A^{(k)}$.



0^{th}**layer** $p^{(0)}$‹$q^{(0)}$ are estimated from one classifier as a face detector;
k^{th}**layer** $p^{(k)}$,$q^{(k)}$ are estimated from both the local search and $p^{(k-1)}$,$q^{(k-1)}$;
K^{th}**layer** $p^{(K)}$,$q^{(K)}$ are estimated, localizing all the feature points in $S^{(K)}$.

0^{th}**layer** $p^{(0)}, q^{(0)}$ are estimated from one classifier as a face detector;
k^{th}**layer** $p^{(k)}$,$q^{(k)}$ are estimated from both the local search and $p^{(k-1)}$,$q^{(k-1)}$;
K^{th}**layer** $p^{(K)}$,$q^{(K)}$ are estimated, localizing all the feature points in $S^{(K)}$.

Fig. 3. The hierarchical structure and coarse-to-fine procedure of face alignment

For each feature point in $S^{(k)}$ a classifier is trained with $\frac{L}{2^k} \times \frac{L}{2^k}$ sized rectangle features to distinguish the feature from non-feature. We can find each feature point independently by maximizing the corresponding likelihood output of the classifier. However, not only the likelihood of the texture but also the geometry shape of the feature points should be considered, and the feature point set should be constrained by the parameter of the shape model. Therefore, to align the face, the localization of the feature point set $S^{(k)}$ can be formulated by Bayesian inference given both the texture in this layer and the parameters estimated in the previous layer, which is