2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)

4-8 May 2015

Ljubljana, Slovenia

Sponsors

IEEE Catalog Number: CFP15074-ART
ISBN: 978-1-4799-6026-2

11th IEEE International Conference on
Automatic Face and Gesture Recognition
FG2015

2015 11TH IEEE INTERNATIONAL CONFERENCE AND WORKSHOPS ON AUTOMATIC FACE AND GESTURE RECOGNITION (FG)

SPONSORS

PLATINUM SPONSORS

SILVER SPONSOR

BRONZE SPONZORS

COPYRIGHT NOTICE

IEEE Catalog Number: CFP15074-USB

ISBN: 978-1-4799-6025-5

ORGANIZED BY

Home

CONFERENCE
Welcome
People
Main program

WORKSHOPS
CBAR
DeID
FERA
B-Wild
EmoSPACE
UHA3DS
Doctoral consorium

MAIN PROGRAM

TUESDAY, 5 MAY 2015

Opening (8:15-8:30)

Keynote I (8:30-9:30)

Takeo Kanade, CMU, USA

Session 1 (9:30-10:30)

Patrick Flynn, Kevin Bowyer, Jonathon Phillips
Lessons from Collecting a Million Biometric Samples

Jiangning Gao, Adrian Evans
Expression Robust 3D Face Recognition by Matching Multi-component Local Shape Descriptors on the Nasal and Adjoining Cheek Regions

Dong Yi, Zhen Lei, Stan Li
Shared Representation Learning for Heterogenous Face Recognition

Coffee Break (10:30-11:00)

Session 2: My Research Vision for the Next 10 Years (11:00-12:20)

Lunch break (12:20-14:00)

Lunch-break Talks - industry

Session 3: Poster Highlights I (14:00-15:00)

Session 4 (15:00-16:00)

Xiaolong Wang, Guodong Guo, Rohith MV, Chandra Kambhamettu
Leveraging Geometry and Appearance Cues for Recognizing Family Photos

Davoud Shahlaei, Volker Blanz
Realistic Inverse Lighting from a Single 2D Image of a Face, Taken Under Unknown and Complex Lighting

Federico Sukno, Mario Rojas, John Waddington, Paul Whelan
On the Quantitative Analysis of Craniofacial Asymmetry in 3D

Coffee Break (16:00-16:15)

Session 5 (16:15-17:15)

Laszlo Jeni, Jeffrey Cohn, Takeo Kanade
Dense 3D Face Alignment from 2D Videos in Real-Time

Zhefan Ye, Yin Li, Yun Liu, Chanel Bridges, Agata Rozga, James M. Rehg
Detecting Bids for Eye Contact Using A Wearable Camera

Yan Li, Ruiping Wang, Shiguang Shan, Xilin Chen
Hierarchical Hybrid Statistic based Video Binary Code and Its Application to Face Retrieval in TV-Series

Poster Session I (17:15-19:00)

Fernando De La Torre, Wen-Sheng Chu, Xuehan Xiong, Francisco Vincente, Xiaoyu Ding, Jeffrey Cohn
IntraFace

Xiaolong Wang, Chandra Kambhamettu
Age Estimation via Unsupervised Neural Networks

Meng Yang, Weiyang Liu, Linlin Shen
Joint Regularized Nearest Points for Image Set based Face Recognition

Xi Peng, Junzhou Huang, Qiong Hu, Shaoting Zhang, Dimitris Metaxas
Three-Dimensional Head Pose Estimation in-the-Wild

Lilei Zheng, Khalid Idrissi, Christophe Garcia, Stefan Duffner, Atilla Baskurt
Triangular Similarity Metric Learning for Face Verification

Handong Zhao, Zhengming Ding, Yun Fu
Block-wise Constrained Sparse Graph for Face Image Representation

Zhengming Ding, Sungjoo Suh, Jae-Joon Han, Changkyu Choi, Yun Fu
Discriminative Low-Rank Metric Learning for Face Recognition

Arnaud Lienhard, Alice Caplier, Patricia Ladret
Fully Automated Facial Picture Evaluation Using High Level Attributes

Xiangyu Zhu, Junjie Yan, Dong Yi, Zhen Lei, Stan Li
Discriminative 3D Morphable Model Fitting

Ming Shao, Zhengming Ding, Yun Fu
Sparse Low-Rank Fusion based Deep Features for Missing Modality Face Recognition

Deepak Pathak, Sai Nitish Satyavolu, Vinay Namboodiri
Where is my Friend? - Person identification in Social Networks

Harald Hanselmann, Hermann Ney
Speeding up 2D-Warping for Pose-Invariant Face Recognition

Nuri Murat Arar, Hua Gao, Jean-Philippe Thiran
Robust Gaze Estimation Based on Adaptive Fusion of Multiple Cameras

Wen Wang, Ruiping Wang, Shiguang Shan, Xilin Chen
Probabilistic Nearest Neighbor Search for Robust Classification of Face Image Sets

Leonardo Cament, Francisco Galdames, Kevin Bowyer, Claudio Perez
Face Recognition under Pose Variation with Active Shape Model to Adjust Gabor Filter Kernels and to Correct Feature Extraction Location

Ahmed Osman, Jay Turcot, Rana Kaliouby
Supervised Learning Approach to Remote Heart Rate Estimation from Facial Videos

Henry Lo, Joseph Cohen, Wei Ding
Prediction Gradients for Feature Extraction and Analysis from Convolutional Neural Networks

Negar Hassanpour, Liang Chen
A Hierarchical Training and Identification Method using Gaussian Process Models for Face Recognition in Videos

Jie Liang, Jun Zhou, Yongsheng Gao
3D Local Derivative Pattern for Hyperspectral Face Recognition

Neslihan Kose, Ludovic Apvrille, Jean-Luc Dugelay
Facial Makeup Detection Technique Based on Texture and Shape Analysis

Poster Session on Previously Published Journal Papers I (17:15-19:00)

Cigdem Eroglu Erdem, Cigdem Turan, Zafer Aydin
BAUM-2: A Multilingual Audio-Visual Affective Face Database
Multimedia Tools and Applications, pp.1-31, 2014
doi: 10.1007/s11042-014-1986-2

Yi Jin, Jiwen Lu, Qiuqi Ruan
Coupled Discriminative Feature Learning for Heterogeneous Face Recognition
IEEE Transactions on Information Forensics and Security, vol.10, no.3, pp.640-652, March 2015
doi: 10.1109/TIFS.2015.2390414

P. Jonathon Phillips, Alice J. O'Toole
Comparison of Human and Computer Performance Across Face Recognition Experiments
Image and Vision Computing, vol. 32, issue 1, pp. 74–85, January 2014
doi:10.1016/j.imavis.2013.12.002

Jeffrey M. Girard, Jeffrey F. Cohn, Fernando De la Torre
Estimating Smile Intensity: A Better Way
Pattern Recognition Letters, 2014 (preprint available online)
doi:10.1016/j.patrec.2014.10.004

Z. Hammal, J.F. Cohn, D.T. George
Interpersonal Coordination of Head Motion in Distressed Couples
IEEE Transactions on Affective Computing, vol.5, no.2, pp.155-167, April-June 2014
doi: 10.1109/TAFFC.2014.2326408

M. Abadi, R. Subramanian, S.M. Kia, P. Avesani, I. Patras, N. Sebe
DECAF: MEG-based Multimodal Database for Decoding Affective Physiological Responses
IEEE Transactions on Affective Computing, 2015 (preprint available online)
doi: 10.1109/TAFFC.2015.2392932

A. Dhall, R. Goecke, T. Gedeon
Automatic Group Happiness Intensity Analysis

IEEE Transactions on Affective Computing, vol.6, no.1, pp.13-26, Jan.-March 2015
doi: 10.1109/TAFFC.2015.2397456

Demo Session I (17:15-19:00)

Toni Fetzer, Christian Petry
3D Interaction Design: Increasing the Stimulus-Response Correspondence by using Stereoscopic Vision

Laszlo Jeni, Jeffrey Girard, Jeffrey Cohn, Takeo Kanade
Real-Time Dense 3D Face Alignment from 2D Video with Automatic Facial Action Unit Coding

Terence Sim, Li Zhang
Controllable Face Privacy

Martin Savc, Damjan Zazula, Jurij Munda, Bozidar Potocnik
Semi-transparent mirror with hidden camera to assess human emotions

Jixu Chen, Ming-Ching Chang, Peter Tu
A Live Video Analytic System for Affect Analysis in Public Space

Stepan Mracek, Radim Dvorak, Jan Vana, Tomas Novotny, Martin Drahansky
3D Face Recognition Utilizing a Low-Cost Depth Sensor

José Mennesson, Benjamin Allaert, Ioan Marius Bilasco, Nico van der Aa, Alexandre Denis, Samuel Cruz-Lara
Faces and Thoughts : an Empathic Dairy

Oya Celiktutan, Evangelos Sariyanidi, Hatice Gunes
Let me Tell You about Your Personality! Real-time Personality Prediction from Nonverbal Behavioural Cues

Tim den Uyl, Emrah Tasli, Paul Ivan, Mariska Snijdewind
Who do you want to be? Real-time Face Swap

Emrah Tasli, Tim den Uyl, Hugo Boujut, Titus Zaharia
Real-Time Facial Character Animation

Fernando De La Torre, Wen-Sheng Chu, Xuehan Xiong, Francisco Vincente, Jeffrey Cohn
IntraFace

WEDNESDAY, 6 MAY 2015

Keynote II (8:30-9:30)

Ursula Hess, Humboldt University, Berlin

Session 6 (9:30-10:30)

Abhinav Dhall, Jyoti Joshi, Karan Sikka, Roland Goecke, Nicu Sebe
The More the Merrier: Analysing the Affect of a Group of People in Images

Shyam Sundar Rajagopalan, O.V. Ramana Murthy, Roland Goecke, Agata Rozga
Play with Me - Measuring a Child's Engagement in a Social Interaction

Malcolm Dcosta, Dvijesh Shastri, Ricardo Vilalta, Judee Burgoon, Ioannis Pavlidis
Perinasal Indicators of Deceptive Behavior

Coffee Break (10:30-11:00)

Session 7: Special Session on Previously Published Journal Papers (11:00-12:30)

P. Jonathon Phillips, Alice J. O'Toole
Comparison of Human and Computer Performance Across Face Recognition Experiments
Image and Vision Computing, vol. 32, issue 1, pp. 74–85, January 2014
doi:10.1016/j.imavis.2013.12.002

Yi Jin, Jiwen Lu, Qiuqi Ruan
Coupled Discriminative Feature Learning for Heterogeneous Face Recognition
IEEE Transactions on Information Forensics and Security, vol.10, no.3, pp.640-652, March 2015
doi: 10.1109/TIFS.2015.2390414

Jeffrey M. Girard, Jeffrey F. Cohn, Fernando De la Torre
Estimating Smile Intensity: A Better Way
Pattern Recognition Letters, 2014 (preprint available online)
doi:10.1016/j.patrec.2014.10.004

A. Dhall, R. Goecke, T. Gedeon
Automatic Group Happiness Intensity Analysis
IEEE Transactions on Affective Computing, vol.6, no.1, pp.13-26, Jan.-March 2015
doi: 10.1109/TAFFC.2015.2397456

Z. Hammal, J.F. Cohn, D.T. George
Interpersonal Coordination of Head Motion in Distressed Couples

IEEE Transactions on Affective Computing, vol.5, no.2, pp.155-167, April-June 2014
doi: 10.1109/TAFFC.2014.2326408

M. Abadi, R. Subramanian, S.M. Kia, P. Avesani, I. Patras, N. Sebe
DECAF: MEG-based Multimodal Database for Decoding Affective Physiological Responses
IEEE Transactions on Affective Computing, 2015 (preprint available online)
doi: 10.1109/TAFFC.2015.2392932

Lunch break (12:30-14:00)

Lunch-break Talks - industry

Session 8: Poster Highlights II (14:00-15:00)

Session 9 (15:00-16:00)

Chongliang Wu, Shangfei Wang, Qiang Ji
Multi-Instance Hidden Markov Model For Facial Expression Recognition

Mahdi Jampour, Thomas Mauthner, Horst Bischof
Pairwise Linear Regression: An Efficient and Fast Multi-view Facial Expression Recognition

Tobias Gehrig, Ziad Al-Halah, Hazım Ekenel, Rainer Stiefelhagen
Action Unit Intensity Estimation using Hierarchical Partial Least Squares

Coffee Break (16:00-16:15)

Session 10 (16:15-17:15)

Robert Walecki, Ognjen Rudovic, Vladimir Pavlovic , Maja Pantic
Variable-state Latent Conditional Random Fields for Facial Expression Recognition and Action Unit Detection

Xing Zhang, Lijun Yin, Jeffrey Cohn
Three Dimensional Binary Edge Feature Representation for Pain Expression Analysis

Peng Liu, Lijun Yin
Spontaneous Facial Expression Analysis Based on Temperature Changes and Head Motions

Poster Session II (17:15-19:00)

Heng Zhang, Vishal Patel, Sumit Shekhar, Rama Chellappa
Domain Adaptive Sparse Representation-Based Classification

Tung-Ying Lee, Tzu-Shan Chang, Shang-Hong Lai
Correcting Radial and Perspective Distortion by Using Face Shape Information

Makarand Tapaswi, Martin Bäuml, Rainer Stiefelhagen
Improved Weak Labels using Contextual Cues for Person Identification in Videos

Miriam Redi, Nikhil Rasiwasia, Gaurav Aggarwal, Alejandro Jaimes
The Beauty of Capturing Faces: Rating the Quality of Digital Portraits

Jun Li, Shasha Li, Jiani Hu, Weihong Deng
Adaptive LPQ: An Efficient Descriptor for Blurred Face Recognition

Zhongjun Wu, Jiayu Li, Jiani Hu, Weihong Deng
Pose-Invariant Face Recognition Using 3D Multi-Depth Generic Elastic Models

Iryna Anina, Ziheng Zhou, Guoying Zhao, Matti Pietikainen
OuluVS2: a multi-view audiovisual database for non-rigid mouth motion analysis

Naimul Khan, Xiaoming Nan, Azhar Quddus, Edward Rosales, Ling Guan
On Video Based Face Recognition Through Adaptive Sparse Dictionary

Bin Yang, Junjie Yan, Zhen Lei, Stan Li
Fine-grained Evaluation on Face Detection in the Wild

Ross Beveridge, Hao Zhang, Bruce Draper, Patrick Flynn, Zhenhua Feng, Patrik Huber, Josef Kittler, Zhiwu Huang, Shaoxin Li, Yan Li, Meina Kan, Ruiping Wang, Shiguang Shan, Xilin Chen, Haoxiang Li, Gang Hua, Vitomir Struc, Janez Krizaj, Changxing Ding, Dacheng Tao, Jonathon Phillips
Report on the FG 2015 Video Person Recognition Evaluation

Andreas Lanitis, Nicolas Tsapatsoulis, Kleanthis Soteriou, Daiki Kuwahara, Shigeo Morishima
FG2015 Age Progression Evaluation

Sharifa Alghowinem, Roland Goecke, Jeffrey Cohn, Michael Wagner, Gordon Parker, Michael Breakspear
Cross-Cultural Detection of Depression from Nonverbal Behaviour

Heng Zhang, Vishal Patel, Rama Chellappa
Robust Multimodal Recognition via Multitask Multivariate Low-Rank Representations

Matthaeus Schumacher, Volker Blanz
Exploration of the Correlations of Attributes and Features in Faces

Rui Li, Jared Curhan, M.Ehsan Hoque

Predicting Video-Conferencing Conversation Outcomes Based on Modeling Facial Expression Synchronization

Mojtaba Khomami Abadi, Juan Abdon Miranda Correa, Julia Wache, Heng Yang, Ioannis Patras, Nicu Sebe
Inference of Personality Traits and Affect Schedule by Analysis of Spontaneous Reactions to Affective Videos

Radu Vieriu, Sergey Tulyakov, Stanislau Semeniuta, Enver Sangineto, Nicu Sebe
Facial Expression Recognition under a Wide Range of Head Poses

Taleb Alashkar, Boulbaba Ben Amor, Stefano Berretti, Mohamed Daoudi
Analyzing Trajectories on Grassmann Manifold for Early Emotion Detection from Depth Videos

Xudong Yang, Di Huang, Yunhong Wang, Liming Chen
Automatic 3D Facial Expression Recognition using Geometric Scattering Representation

Zahoor Zafrulla, Himanshu Sahni, Abdelkareem Bedri, Pavleen Thukral, Thad Starner
Hand Detection in American Sign Language Depth Data Using Domain-Driven Random Forest Regression

Poster Session on Previously Published Journal Papers II (17:15-19:00)

Ognjen Rudovic, Vladimir Pavlovic, Maja Pantic
Context-Sensitive Dynamic Ordinal Regression for Intensity Estimation of Facial Action Units
IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.37, no.5, pp.944-958, May 1 2015
doi: 10.1109/TPAMI.2014.2356192

Sezer Ulukaya, Cigdem Eroglu Erdem
Gaussian Mixture Model based Estimation of the Neutral Face Shape for Emotion Recognition
Digital Signal Processing, vol. 32, pp. 11–23, September 2014
doi:10.1016/j.dsp.2014.05.013

Weihong Deng, Jiani Hu, Jiwen Lu, Jun Guo
Transform-Invariant PCA: A Unified Approach to Fully Automatic Face Alignment, Representation, and Recognition
IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.36, no.6, pp.1275-1284, June 2014
doi: 10.1109/TPAMI.2013.194

Federico M. Sukno, John L. Waddington, Paul F. Whelan
3-D Facial Landmark Localization With Asymmetry Patterns and Shape Regression from Incomplete Local Features
IEEE Transactions on Cybernetics, 2014 (preprint available online)
doi: 10.1109/TCYB.2014.2359056

Michal Kawulok, Jolanta Kawulok, Jakub Nalepa, Bogdan Smolka
Self-adaptive Algorithm for Segmenting Skin Regions
EURASIP Journal on Advances in Signal Processing, vol. 2014, 2014
doi:10.1186/1687-6180-2014-170

Joseph Roth, Xioming Liu, Dimitris Metaxas
On Continuous User Authentication via Typing Behavior
IEEE Transactions on Image Processing, vol. 23, no.10, pp. 4611-4624, October 2014
doi: 10.1109/TIP.2014.2348802

Jacob Foytik, Vijayan K. Asari
A Two-Layer Framework for Piecewise Linear Manifold-Based Head Pose Estimation
International Journal of Computer Vision, vol. 101, issue 2, pp. 270-287, January 2013
doi: 10.1007/s11263-012-0567-y

Demo Session II (17:15-19:00)

Same as Demo Session I

THURSDAY, 7 MAY 2015

Keynote III (8:30-9:30)

Louis-Philippe Morency, CMU, USA

Session 11 (9:30-10:30)

Ajjen Joshi, Camille Monnier, Margrit Betke, Stan Sclaroff
A Random Forest Approach to Segmenting and Classifying Gestures

Pavlo Molchanov, Shalini Gupta, Kihwan Kim, Kari Pulli
Multi-sensor System for Driver's Hand-Gesture Recognition

Nataraj Jammalamadaka, Andrew Zisserman, C. V. Jawahar
Human Pose Search using Deep Poselets

Coffee Break (10:30-11:00)

Session 12: Panel Session: The Promise and Perils of Found Data (11:00-12:20)

Lunch break (12:20-14:00)

Lunch-break Talks - industry

Session 13: Poster Highlights III (14:00-15:00)

Session 14 (15:00-16:00)

Yasushi Makihara, Al Mansur, Daigo Muramatsu, Zasim Uddin, Yasushi Yagi
Multi-view Discriminant Analysis with Tensor Representation and Its Application to Cross-view Gait Recognition

Yelin Kim, Jixu Chen, Ming-Ching Chang, Xin Wang, Emily Mower Provost, Siwei Lyu
Modeling Transition Patterns Between Events for Temporal Human Action Segmentation and Classification

Iftekhar Naim, Iftekhar Tanveer, Daniel Gildea, M. Ehsan Hoque
Automated Prediction and Analysis of Job Interview Performance: The Role of What You Say and How You Say It

Coffee Break (16:00-16:15)

Session 15: Special Session on Evaluations and Mouth Motion Analysis (16:15-17:15)

Jiwen Lu, Junlin Hu, Venice Erin Liong, Xiuzhuang Zhou, Andre Bottino, Ihtesham Ul Islam, Tiago Vieira, Xiaoqian Qin, Xiaoyang Tan, Songcan Chen, Shahar Mahpod, Yosi Keller, Lilei Zheng, Khalid Idrissi, Christophe Garcia, Stefan Duffner, Atilla Baskurt, Modesto Castrillón-Santana, Javier Lorenzo-Navarro
The FG 2015 Kinship Verification in the Wild Evaluation

Shiyang Cheng, Ioannis Marras, Stefanos Zafeiriou, Maja Pantic
Active Nonrigid ICP Algorithm

Epameinondas Antonakos, Anastasios Roussos, Stefanos Zafeiriou
A Survey on Mouth Modeling and Analysis for Sign Language Recognition

Poster Session III (17:15-19:00)

Yale Song, Daniel McDuff, Deepak Vasisht, Ashish Kapoor
Exploiting Sparsity and Co-occurrence Structure for Action Unit Recognition

Arnaud Dapogny, Kevin Bailly, Severine Dubuisson
Dynamic facial expression recognition by joint static and multi-time gap transition classification

Nesrine Fourati, Catherine Pelachaud
Multi-level classification of emotional body expression

Songfan Yang, Le An Mehran Kafai, Bir Bhanu
To Skip or not to Skip? A Dataset of Spontaneous Affective Response of Online Advertising (SARA) for Audience Behavior Analysis

Shan Wu, Shangfei Wang, Jun Wang
Enhanced facial expression recognition by age

Jeffrey Girard, Jeffrey Cohn, Laszlo Jeni, Simon Lucey, Fernando De La Torre
How much training data for facial action unit detection?

Thomas Vandal, Daniel McDuff, Rana Kaliouby
Event Detection: Ultra Large-scale Clustering of Facial Expressions

Samuel Berlemont, Grégoire Lefebvre, Stefan Duffner, Christophe Garcia
Siamese Neural Network based Similarity Metric for Inertial Gesture Classification and Rejection

Kaoning Hu, Lijun Yin
Multiple Feature Representations from Multi-Layer Geometric Shape for Hand Gesture Analysis

Yu Kong, Behnam Satarboroujeni, Yun Fu
Hierarchical 3D Kernel Descriptors for Action Recognition Using Depth Sequences

Jeong-jik Seo, Jisoo Son, Hyungil Kim, Wesley De Neve, Young Man Ro
Efficient and Effective Human Action Recognition in Video through Motion Boundary Description with a Compact Set of Trajectories

Toni Fetzer, Christian Petry, Frank Deinzer, Karsten Huffstadt
3D Interaction Design: Increasing the Stimulus-Response Correspondence by using Stereoscopic Vision

Seyed Morteza Safdarnejad, Xiaoming Liu, Lalita Udpa, Brooks Andrus, John Wood, Dean Craven
Sports Videos in the Wild (SVW): A Video Dataset for Sports Analysis

Toi Nguyen, Lan Le, Hai Tran, Rémy Mullot, Vincent Courboulay
A New Hand Representation Based on Kernels for Hand Posture Recognition

Philip Krejov, Andrew Gilbert, Richard Bowden
Combining Discriminative and Model Based Approaches for Hand Pose Estimation

Hanjie Wang, Xiujuan Chai, Yu Zhou, Xilin Chen
Fast Sign Language Recognition Benefited From Low Rank Approximation

Jun Shiraishi, Takeshi Saitoh
Optical Flow based Lip Reading using Non Rectangular ROI and Head Motion Reduction

Doctoral Consortium Poster Session (17:15-19:00)

Shyam Sundar Rajagopalan, O.V. Ramana Murth, Roland Goecke, Agata Rozga
Play with Me - Measuring a Child's Engagement in a Social Interaction

Peng Liu, Lijun Yin
Spontaneous Facial Expression Analysis Based on Temperature Changes and Head Motions

Zahoor Zafrulla, Himanshu Sahni, Abdelkareem Bedri, Pavleen Thukral, Thad Starner
Hand Detection in American Sign Language Depth Data Using Domain-Driven Random Forest Regression

Malcolm Dcosta, Dvijesh Shastri, Ioannis Pavlidis
Perinasal Indicators of Malevolence

Shiyang Cheng, Ioannis Marras, Stefanos Zafeiriou, Maja Pantic
Active Nonrigid ICP Algorithm

Negar Hassanpour, Liang Chen
A Hierarchical Training and Identification Method using Gaussian Process Models for Face Recognition in Videos

Jiangning Gao, Adrian N. Evans
Expression Robust 3D Face Recognition by Matching Multi-component Local Shape Descriptors on the Nasal and Adjoining Cheek Regions

Radu-Laurentiu Vieriu, Sergey Tulyakov, Stanislau Semeniuta, Enver Sangineto, Nicu Sebe
Facial Expression Recognition under a Wide Range of Head Poses

Yelin Kim, Jixu Chen, Ming-Ching Chang, Xin Wang, Emily Mower Provost, Siwei Lyu
Modeling Transition Patterns Between Events for Temporal Human Action Segmentation and Classification

Demo Session III (17:15-19:00)

Same as Demo Session I

ORGANIZED BY

A New Hand Representation Based on Kernels for Hand Posture Recognition

Van-Toi NGUYEN[1,2,3] Thi-Lan LE[1] Thanh-Hai TRAN[1] Rémy MULLOT[2]
Vincent COURBOULAY[2]
[1] International Research Institute MICA, HUST-CNRS/UMI-2954-GRENOBLE INP
and Hanoi University of Science & Technology, Vietnam
[2] L3i Laboratory, the University of La Rochelle, France
[3] The University of Information and Communication Technology under Thai Nguyen University, Vietnam

***Abstract*— Hand posture recognition is an extremely active research topic in Computer Vision and Robotics, with many applications ranging from automatic sign language recognition to human-system interaction. Recently, a new descriptor for object representation based on the kernel method (KDES) has been proposed. While this descriptor has been shown to be efficient for hand posture representation, across-the-board use of KDES for hand posture recognition has some drawbacks. This paper proposes three improvements to KDES to make it more robust to scale change, rotation, and differences in the object structure. First, the gradient vector inside the gradient kernel is normalized, making gradient KDES invariant to rotation. Second, patches with adaptive size are created, to make hand representation more robust to changes in scale. Finally, for patch-level features pooling, a new pyramid structure is proposed, which is more suitable for hand structure. These innovations are tested on three datasets; the results bring out an increase in recognition rate (as compared to the original method) from 84.4% to 91.2%.**

I. INTRODUCTION

Vision-based hand posture recognition plays an important role in natural human-machine interaction. This paper focuses on the recognition of hand postures, an issue that is relevant to (i) static recognition, as hand postures can directly replace some remote control devices, using a one-to-one correspondence between hand postures and commands; but also to (ii) dynamic recognition, as a dynamic hand gesture can be defined as a sequence of hand postures, hence the usefulness of the identification of key hand postures for dynamic hand gesture recognition.

Challenges to vision-based hand posture recognition include the following: (i) as with all other problems in computer vision, vision-based hand posture recognition is affected by changes in lighting condition, cluttered backgrounds, and changes in scale; (ii) additionally, the hand is a highly deformable object; there is a considerable number of mutually similar hand postures; (iii) applications using hand posture generally require real-time, user-independent recognition. A number of hand recognition methods have been proposed to address these challenges [1], [2], [3], [4]. These methods can be divided into two main categories depending on whether they are based on an implicit or explicit representation of the hand.

The methods belonging to the explicit category often require good hand segmentation results ([5], [6]) to extract hand components: typically fingers. For example, in [5], the region of the hand is detected by applying a color segmentation technique on simple uniform background. Then, the palm morphological characteristics and finger features that allow identifying the raised fingers are extracted based on the Self-Growing and Self-Organized Neural Gas network. This method could not be applicable for real applications because the segmentation of the hand in a real environment with cluttered backgrounds is not always good. The approaches proposed in [7], [8] do not require a good segmentation, but the computation time is too great for many applications. In [8], hand postures are matched with predesigned bunch graph models. Each node in this model contains local features. The classification task is done by finding the region that has the best match between bunch graph and target image. The bunch graph is slid on the image. At each position of the bunch graph, the position of each node is refined to find the best one considering the distortions of the nodes.

The methods of the second category are more flexible since they just require a hand region as input [9], [10], [11]. Hand region is usually defined by a bounding box. In [9], good results were obtained when applying SIFT features with BoW (Bag of Words) and SVM (Support Vector Machine) into hand posture recognition. This method, however, does not work well when working with low resolution due to the limited number of detected keypoints.

In [11], it was proposed to apply kernel descriptor method (KDES) [12] for hand posture recognition. This method proved effective even in case of low resolution as well as complex background. Nevertheless, the use of original KDES still has some limitations. In this paper, we propose a new hand posture representation based on KDES with following properties:

- *Patch-level features are invariant to rotation*: At patch level, the original KDES computes the gradient based features without considering the orientation. For this reason, the generated features are sensitive to rotation. To remedy this, we propose to compute the dominant orientation of the patch and normalize all gradient vec-

This research is funded by Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant number FWO.102.2013.08.

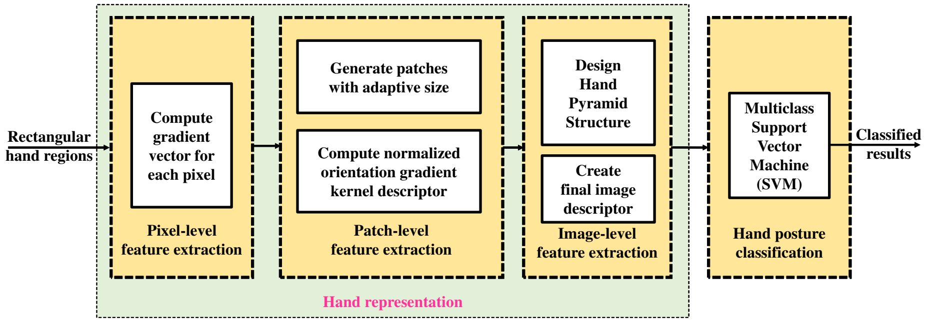

Fig. 1: The framework of proposed hand postures recognition method.

tors in the patch to this orientation. Patch-level features will thus be invariant to rotation.

- *Robust to scale change*: The original KDES computes features over patches of fixed size. At two scales, the number of patches to be considered and the corresponding patch descriptions will be different. We propose a strategy to generate patches with adaptive size. This makes constant the number of patches and robust patch description. As a result, image-level feature is invariant to scale change.
- *Suitable to the specific structure of the hand*: At image level, the original KDES organizes a spatial pyramid structure of patches to build the final description of the image. However, we observe that the hand is an object with a specific structure. We then design a new pyramid structure that better reflects the structure of the hand.

To evaluate the proposed method, we use three datasets (Triesch dataset [7], NUS II dataset [10]) and [11]). We perform different experiments in order to demonstrate the recognition performance, according to each proposed improvement, and compare with the state-of-the-art methods.

The remainder of the paper is organized as follows. The section II introduces the proposed method for hand posture recognition using the kernel method. Each step of the method is explained in detail, indicating the main improvement in each step. The experimental results are presented in the section III. The conclusions and directions for future work are given in the section IV.

II. PROPOSED METHOD FOR HAND POSTURE RECOGNITION

A. General framework of hand posture recognition using the Kernel method

The proposed framework of hand posture recognition using kernel is presented in Fig. 1. It comprises two main steps:

- *Hand representation*: This step takes a hand region image (from now on called *image*, for short) as input and returns a descriptor of the hand candidate. It is composed of multiple sub-steps:
  - Pixel-level feature extraction: At this level, a gradient vector is computed for each pixel of the image.
  - Patch-level feature extraction: At this level, we firstly have to generate a set of patches then compute patch-level features. Different from [12], depending on image resolution, we create patches with adaptive size instead of fixed size. This adaptive size ensures the number of patches to be considered unchanged. In addition, it makes the patch descriptor more robust to scale change. For each patch, we compute patch features as follows. Given an image patch, we compute a gradient descriptor based on the original idea proposed in [12]. However, unlike [12], we first compute the dominant orientation of the patch, and then normalize all gradient vectors to this orientation. This normalization is done inside the gradient kernel allowing the descriptor to be invariant to rotation.
  - Image-level feature extraction: At this step, we propose a modification w.r.t to [12]: To combine patch features, we propose a pyramid structure *specific to hand postures* instead of a general pyramid structure. This specific pyramid structure makes the descriptor more suitable for hand representation. Given an image, the final representation is built based on features extracted from lower levels using efficient match kernels (EMK) proposed in [12]. First, we have to compute the feature vector for each cell of the hand pyramid structure, and then concatenate them into a final descriptor.
- *Hand posture classification*: Once the hand is represented by a descriptor vector, any classifier could be applied for the classification task. In this paper following the strategy originally proposed [12], we will use Multi-class SVM. In the following sections, we focus to present in detail the successive steps in hand representation.

B. Hand representation

1) Extraction of pixel-level features: According to [12], [11], a number of features can be computed at the pixel level, such as pixel values, texture, and gradient. In [11], it was argued that gradient is the best feature for hand posture recognition; accordingly, in this paper we use the gradient at pixel level.

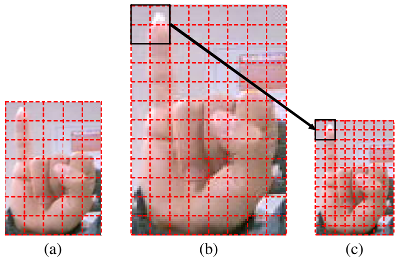

Fig. 2: An example of the uniform patch in the original KDES and the adaptive patch in our method. (a,b) two images of the same hand posture with different sizes are divided using a uniform patch; (b, c): two images of the same hand posture with different sizes are divided using the adaptive patch.

The gradient vector at a pixel z is defined by its magnitude $m(z)$ and orientation $\theta(z)$. In [12], the orientation $\widetilde{\theta}(z)$ is defined as follows:

$$\widetilde{\theta}(z) = [sin(\theta(z))\ cos(\theta(z))] \tag{1}$$

2) Extraction of patch-level features:

a) Generate a set of patches with adaptive size from an image: In the original work [12], the author generated patches with a fixed size for all images in the dataset even the dataset contains images with different resolutions. For low-resolution images, the number of generated patches will be very limited, producing a poor representation of the image. Beside, the feature vectors of two images of the same hand posture at two scales will be highly different. Consequently, the original KDES is not invariant to scale change.

Fig. 2(a,b) illustrates this problem. Fig. 2(a) and (b) are two images of the same hand posture at two scales. Fig. 2(a) has a size of 40×56 while Fig. 2(b) is two times bigger (64×96). When we use a uniform patch of size 16×16 and uniform grid 8×8, Fig. 2(b) has 77 patches while Fig. 2(a) has only 24 patches. A patch of Fig. 2(a) contains more real area of hand than a patch of Fig. 2 (b). Obviously, the feature vectors of patches are very different. The above analysis motivates us to make an adaptive patch size in order to get a similar number of patches along both horizontal and vertical axes. Suppose that the given number of patches is $np_x \times np_y$ (np_x patches along the horizontal axis and np_y patches along the vertical axis). The number of grid cells $ngrid_x \times ngrid_y$ is defined as: $ngrid_x = np_x + 1, ngrid_y = np_y + 1$. With an image has size of $w \times h$, the adaptive grid cell size along horizontal axis $gridsize_x = \frac{w}{ngrid_x}$ and the adaptive grid cell size along vertical axis $gridsize_y = \frac{h}{ngrid_y}$. The adaptive patch has the size of $patchsize_x \times patchsize_y$ where $patchsize_x = 2gridsize_x$ and $patchsize_y = 2gridsize_y$. A patch is constructed from 4 cells of the grid. The overlap of two adjacent patches along the horizontal or vertical axes is a region of two cells of the grid. By this way, the size of the patches is directly proportional to the size of the image. Fig. 2(b,c) illustrates the advantage of the proposed adaptive patch and the representation of images based on patch-level features will be robust to scale change.

b) Compute patch-level feature: Patch-level features are computed based on the idea of the kernel method. Derived from a match kernel representing the similarity of two patches, we can extract the feature vector for the patch using an approximate patch-level feature map, given a designed patch level match kernel function.

The gradient match kernel is constructed from three kernels that are gradient magnitude kernel $k_{\widetilde{m}}$, orientation kernel k_o and position kernel k_p. In [12], gradient match kernel is defined as follows:

$$K_{gradient}(P,Q) = \sum_{z \in P} \sum_{z' \in Q} k_{\widetilde{m}}(z,z') k_o(\widetilde{\theta}(z), \widetilde{\theta}(z')) k_p(z,z') \tag{2}$$

where P and Q are patches of two different images that we need to measure the similarity. z and z' denote the 2D position of a pixel in the image patch P and Q respectively. $\theta(z)$ and $\theta(z')$ are gradient orientations at pixel z and z' in the patch P and Q respectively.

Directly using the gradient orientation $\widetilde{\theta}(z)$ in orientation kernel, the patch level features extracted from the match kernel will not be invariant to rotation. We then propose to normalize gradient orientation before applying in match kernel. Specifically, inspired by the idea of SIFT descriptor [13], we compute a dominant orientation of the patch and normalize all gradient vectors to this orientation. We propose two ways to determine the dominant orientation $\overline{\theta}(P)$ of the patch P. First, we use the dominant orientation of the patch as proposed in [13]. Second, we compute a vector sum of all the gradient vectors in the patch. The normalized gradient angle of a pixel z in P thus becomes:

$$\omega(z) = \theta(z) - \overline{\theta}(P) \tag{3}$$

Then, according (1), the normalized orientation of a gradient vector will be:

$$\widetilde{\omega}(z) = [sin(\omega(z))\ cos(\omega(z))] \tag{4}$$

Finally, we define the gradient match kernel with the normalized orientation as follows:

$$K_{gradient}(P,Q) = \sum_{z \in P} \sum_{z' \in Q} k_{\widetilde{m}}(z,z') k_o(\widetilde{\omega}(z), \widetilde{\omega}(z')) k_p(z,z') \tag{5}$$

The gradient magnitude kernel $k_{\widetilde{m}}$ is defined as: $k_{\widetilde{m}}(z,z') = \widetilde{m}(z)\widetilde{m}(z')$ where $\widetilde{m}(z) = \frac{m(z)}{\sqrt{\sum_{z \in P} m(z)^2 + \epsilon_g}}$, ϵ_g is a small constant, $m(z)$ is the gradient magnitude at a pixel z.

Both the orientation kernel k_o and the position kernel k_p are Gaussian kernels $k(x,x') = \exp(-\gamma \|x - x'\|^2)$. The factor γ is defined individually for k_o and k_p that are denoted by γ_o and γ_p respectively.

Now, given the definition of match kernel, how to extract feature vector for a patch. Let $\varphi_o(.)$ and $\varphi_p(.)$ the feature maps for the gradient orientation kernel k_o and position kernel k_p respectively. Then, the approximate feature over image patch P is constructed as:

$$\overline{F}_{gradient}(P) = \sum_{z \in P} \widetilde{m}(z)\phi_o(\widetilde{\omega}(z)) \otimes \phi_p(z) \quad (6)$$

where $\otimes$ is the Kronecker product, $\phi_o(\widetilde{\omega}(z))$ and $\phi_p(z)$ are approximate feature maps for the kernel k_o and k_p, respectively.

The approximate feature maps are computed based on a basis method of kernel descriptor. The basic idea of representation based on kernel methods is to compute the approximate explicit feature map for kernel match function. In other word, the kernel match functions are approximated based on explicit feature maps. This enables efficient learning methods for linear kernels to be applied to the non-linear kernel. This approach was introduced in [14], [15], [16], [12].

One of the methods for approximating explicit features has been presented in [16]. In the following, we review this method briefly. Given a match kernel function $k(x, y)$, the feature map $\varphi(.)$ for the kernel $k(x, y)$ is a function mapping a vector x into a feature space so as:

$$k(x, y) = \varphi(x)^\top \varphi(y) \quad (7)$$

Suppose that we have a set of basis vectors $B = \{\varphi(v_i)\}_{i=1}^{D}$, the approximation of feature map $\varphi(x)$ can be:

$$\phi(x) = Gk_B(x) \quad (8)$$

where G is defined by: $G^\top G = K_{BB}^{-1}$ where K_{BB} is $D \times D$ matrix with $\{K_{BB}\}_{ij} = k(v_i, v_j)$. k_B is a $D \times 1$ vector with $\{k_B\}_i = k(x, v_i)$.

To extract approximate features $\phi_o(\widetilde{\omega}(z))$, $\phi_p(z)$ from match kernels, compact basis vectors need to be generated by learning. The compact basis vectors are learned from sufficient basis vectors using kernel principal component analysis. Where, the sufficient basis vectors are sampled uniformly and densely from support region using a fine grid so as these basis vectors make an accurate approximation to match kernels. We use the shared set of basis vectors and match kernel parameters from [12] that were learned using a subset of ImageNet. Let the learned set of d_o basis vectors is $B_o = \{\varphi_o(x_1), \varphi_o(x_2), ..., \varphi_o(x_{d_o})\}$ and the set of d_p basis vectors is $B_p = \{\varphi_p(y_1), \varphi_p(y_2), ..., \varphi_p(y_{d_p})\}$ considering k_o and k_p kernels respectively. Where x_i are sampled normalized gradient vectors and y_i are normalized $2D$ position of pixels in an image patch.

The Kronecker product causes high dimension of the feature vector $\overline{F}_{gradient}(P)$. To reduce the dimension of $\overline{F}_{gradient}$, the kernel principal component analysis is applied into the joint basis vectors $\{\varphi_o(x_i) \otimes \varphi_p(y_j)\}_{i=1..d_o, j=1..d_p}$. Let t-*th* component ${\alpha^t}_{ij}$ is learned through kernel principal component analysis, following [12], the resulting gradient kernel descriptor for match kernel in (5) has the form:

$$\widetilde{F}^t_{gradient}(P) = \sum_{i=1}^{d_o}\sum_{j=1}^{d_p} \alpha^t_{ij} \sum_{z \in P} \widetilde{m}(z) k_o(\widetilde{\omega}(z), x_i) k_p(z, y_j) \quad (9)$$

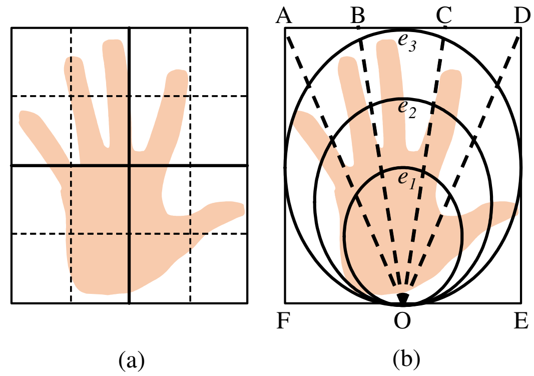

Fig. 3: (a) General spatial pyramid structure used in [16]. (b) The proposed hand pyramid structure.

3) Extraction of image-level features : Once patch-level features are computed for each patch, the remaining work is computing a feature vector representing the whole image. In [16], the authors proposed a spatial pyramid structure by dividing the image into cells using horizontal and vertical lines at several layers (Fig.3(a)). This structure is general without taking the specific shape of objects into account. In our work, as the hand is an object with a specific structure, we propose a new pyramid structure specifically for the hand. In the following, we present in detail each step to build the final descriptor of the image.

a) Design a hand specific pyramid structure for patch-level features pooling: Fig. 3.b shows the proposed hand pyramid structure. The main idea is to exploit characteristics of hand postures. Let the hand posture image have a size of $w \times h$. We remark that the regions at images corners often do not contain hands. For this reason, we only consider the area inside the inscribed ellipse of the hand image rectangle bounding box (e_3). The lines along the fingers converge at the lowest center point of the palm, near the wrist (O). Based on the structure of the hand, the ellipses (e_1, e_2, e_3) and the lines (OA, OB, OC, OD) are used to divide the hand region into parts that contain different components of the hand such as palm and fingers where $AB = BC = CD$. The detail of designed structure is described as: O is the midpoint of $FE(OF = OE)$. The ellipse e_1 is the inscribed ellipse of the rectangle that has a size of $(\frac{1}{2}w \times \frac{1}{2}h)$. The line FE is a tangent line of the ellipse e_1. The contact between the line FE and the ellipse e_1 is O. The ellipses are upright. In the similarity, the ellipsis e_2 is the inscribed ellipsis of the rectangle that has size of $(\frac{3}{4}w \times \frac{3}{4}h)$. In a layer, we define a cell as being a full region limited by these ellipses and lines. In our work, the hand pyramid structure has 3 layers, (see Fig.4(b)).

- Layer 1: This layer contains only one cell defined by the biggest inscribed ellipse e_3.
- Layer 2: In [12], this layer has four rectangular cells. Unlike this, we create eight cells: three cells created from 3 ellipses and five cells created from the intersection of four lines with the biggest ellipse.

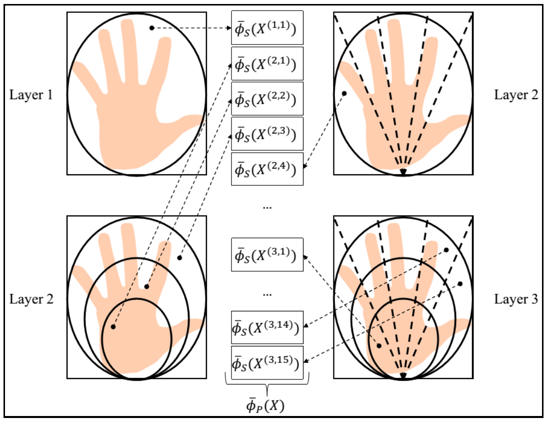

Fig. 4: Construction of image-level feature concatenating feature vectors of cells in layers of hand pyramid structure.

- Layer 3: This layer has 15 cells generated from the intersection between lines and three ellipses.

b) Create the final descriptor of the whole image: Let C be a cell that has a set of patch-level features $X = \{x_1, ..., x_p\}$ then the feature map on this set of vectors is defined as:

$$\overline{\phi}_S(X) = \frac{1}{|X|} \sum_{x \in X} \phi(x) \tag{10}$$

Where $\phi(x)$ is approximate feature maps (8) for the kernel $k(x, y)$ with the set of basis vector that is generated by constrained singular value decomposition method (CKSVD) [16]. The feature vector on the set of patches, $\overline{\phi}_S(X)$, is extracted explicitly.

Given an image, let L be the number of spatial layers to be considered. In our case $L = 3$. The number of cells in layer l-th is (n_l). $X(l, t)$ is set of patch-level features falling within the spatial cell (l, t) (cell t-th in the l-th level). A patch is fallen in a cell when its centroid belongs to the cell. The feature map on the hand pyramid structure is:

$$\begin{aligned} \overline{\phi}_P(X) \;\; = \;\; & [w^{(1)}\overline{\phi}_S(X^{(1,1)}); ...; w^{(l)}\overline{\phi}_S(X^{(l,t)}); \\ & ...; w^{(L)}\overline{\phi}_S(X^{(L,n_L)})] \end{aligned} \tag{11}$$

In (11), $w^{(l)} = \frac{\frac{1}{n_l}}{\sum_{l=1}^{L} \frac{1}{n_l}}$ is the weight associated with level l. Fig. 4 shows image-level feature extraction on the proposed hand pyramid structure. Until now, we obtain the final representation of the whole image, which we call image-level feature vector. This vector will be the input of a Multiclass SVM for training and testing.

III. EXPERIMENTAL RESULTS

In order to evaluate the performance of our hand representation method, we use three available hand posture datasets [10], [7], [11]. Tab. I gives information on these datasets after pre-processing. In these datasets, hand posture images are captured in complex natural backgrounds.

We perform two experiments. The first experiment aims at comparing the performance of our method with the state of the art methods. The objective of the second experiment is to analyze the effect of our three improvements.

TABLE I: Three datasets used in our experiments

Name of dataset	# hand postures	# training images	#testing images	image resolution
Nguyen et al. [11]	21	4636	4690	$(25 \div 138) \times (37 \div 135)$
NUS II [10]	10	200	1000	$(27 \div 97) \times (57 \div 110)$
Jochen Triesch [7]	10	60	660	$(31 \div 108) \times (42 \div 113)$

In the first experiment, among different approaches proposed for hand posture recognition, we choose two methods presented in [11] and in [9] because they are closely related to our work and proved robust for hand posture recognition.

TABLE II: Average accuracy (%) obtained for three datasets with manual hand segmentation

Dataset	Method [9]	Method [11]	Our method
Nguyen et al. [11]	34.5	84.4	**91.2**
NUS II [10]	43.2	95.3	**97.1**
Jochen Triesch [7]	60.8	95.7	**96.7**

TABLE III: Average accuracy obtained (%) for the dataset [11] with automatic hand segmentation

	Method in [9]	Method in [11]	Our method
Accuracy (%)	20.6	74.0	**80.0**

Tab. II shows obtained accuracy of three methods on three datasets with perfect hand detection while Tab. III illustrates the accuracy of the three methods for the dataset [11] with automatic hand detection. For automatic hand detection, we apply the hand detection method proposed in [17] to detect internal center region of the hand then simply expand in order to obtain whole hand region. We keep only the true detections (based on the condition of the PASCAL VOC challenge that is based on Jaccard index) and discard the false detections since we focus on evaluating the hand posture recognition method. We select randomly from the automatic detection results on dataset [11] 100 examples per posture for testing and 100 examples per posture for training. We can observe that our method outperforms the two state of the art methods on all datasets for both manual and automatic hand detection. The recognition accuracy with automatic hand detection is, of course, lower than with manual detection, but remains relatively good (80%). This suggests that we can combine our recognition method with the hand detection method in order to build a complete human-robot interaction using hand postures. However, the performance of the method depends on the characteristics of the data to which it is applied. With the dataset [11], since this dataset contains images of the same hand postures in different scales, our method has proved its robustness. Our method gets 7% better than the original method based on kernel descriptor. For the two others datasets, the improvement in recognition accuracy is smaller. The method presented in [9] shows limitations when applied to these datasets, due to the small number of detected key points. Tab. IV shows the main diagonal of the confusion matrix obtained from the method in [11] and our method with the same dataset [11]. Our method improves the recognition

TABLE IV: Main diagonal of the confusion matrix (%) with the method in [11] and our method for 21 hand posture classes in the dataset [11]

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
Method [11]	100	69.4	91.2	95.6	73.2	90.4	66.1	45.2	74.3	84.1	83.4	100	95.8	93.6	98.6	100	91.6	92.2	70.1	93.1	67.2
Our method	100	71.1	99.6	93.0	80.8	96.1	**83.3**	**74.2**	82.9	92.7	92.6	100	100	100	100	100	94.7	99.0	77.7	92.6	86.9

TABLE V: Main diagonal of the confusion matrix (%) obtained with our method for 21 hand posture classes in the dataset [11] for three testing cases

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
Method [11]	100	69.4	91.2	95.6	73.2	90.4	66.1	45.2	74.3	84.1	83.4	100	95.8	93.6	98.6	100	91.6	92.2	70.1	93.1	67.2
Case 1	100	81.0	95.1	93.4	82.1	90.0	76.6	73.7	82.4	86.8	88.2	100	100	97.9	100	100	94.7	97.1	81.7	93.5	90.0
Case 2	100	75.9	99.1	93.4	78.1	100	84.9	76.5	79.7	81.4	91.3	100	100	100	100	100	93.8	100	75.0	92.6	90.0
Case 3	100	71.1	99.6	93.0	80.8	96.1	83.3	74.2	82.9	92.7	92.6	100	100	100	100	100	94.7	99.0	77.7	92.6	86.9

accuracy for almost all hand posture classes (19 over 21). Especially, for class #7 and #8, the recognition accuracy increases 30% after applying our improvement.

In the second experiment, in order to obtain a detailed analysis of the behavior of our three improvements, we perform different comparisons on the dataset of [11]. As described in section II.B, our method has three improvements: adaptive patch, normalized gradient orientation, and hand pyramid structure. We observe the performance of the method in the following cases: **Case 1**: Apply only adaptive patch; **Case 2** Combine both the adaptive patch and hand pyramid structure; **Case 3**: Combine all improvements.

Based on the obtained result shown in Tab. VI, we can see that the adaptive patch improvement makes a great difference. The performance increases 6% after applying the adaptive patch instead of the uniform patch in the original method. With this dataset, the hand pyramid structure and normalized gradient orientation have a minor contribution. Tab. V provides the recognition accuracy obtained for 21 hand posture classes in three testing cases. From this result, one time again, the adaptive patch improvement shows that it has an important impact on the recognition accuracy. This improvement makes the recognition accuracies of 15 over 21 classes increases. The spatial hand posture is relatively sensitive. Its robustness depends on the characteristics of the hand posture.

TABLE VI: Effects of our improvements in three cases: Case 1: Apply only adaptive patch; Case 2: Combine both the adaptive patch and hand pyramid structure; Case 3: Combine all improvements

Method	[11]	Case 1	Case 2	Case 3
Accuracy (%)	**84.4**	**90.6**	91.0	91.2

Concerning computation time, our method takes averagely 0.3s per image when working with 50×100 image using Matlab 8 (R2013a), Window 64-bit Operating System with processor Intel(R) Core(TM) i5-2520M.

IV. CONCLUSIONS AND FUTURE WORKS

We presented in this paper a new representation of hand posture using kernel methods. This representation is invariant to rotation, robust to scale change and suitable for specific hand structures. The experiment results show that the adaptive patch brought a significant improvement in recognition accuracy (from 84.4% to 90.6%). Rotation invariance and hand structure suitability properties increased the performance slightly (from 90.6% to 91.2%). The reason was that three datasets are not enough appropriate to demonstrate these properties. In the future, we will evaluate our method with a more challenging dataset (images with rotation and changes in scale) such as the one used in [6]. We also plan to apply the proposed method to human-robot interaction using hand gestures. The principles underlying our method could also be extended to other types of objects.

REFERENCES

[1] S. S. Rautaray and A. Agrawal, "Vision based hand gesture recognition for human computer interaction: a survey," *AIR*, Nov. 2012.

[2] M. M. Hasan and P. K. Mishra, "Hand Gesture Modeling and Recognition using Geometric Features : A Review," *CJIPCV*, vol. 3, no. 1, 2012.

[3] V. I. Pavlovic, R. Sharma, and T. S. Huang, "Visual interpretation of hand gestures for human-computer interaction: a review," *TPAMI*, vol. 19, no. 7, pp. 677–695, Jul. 1997.

[4] A. Chaudhary, J. L. Raheja, K. Das, and S. Raheja, "Intelligent Approaches to interact with Machines using Hand Gesture Recognition in Natural way: A Survey," *IJCSES*, vol. 2, no. 1, pp. 122–133, Feb. 2011.

[5] E. Stergiopoulou and N. Papamarkos, "Hand gesture recognition using a neural network shape fitting technique," *EAAI*, vol. 22, no. 8, pp. 1141–1158, 2009.

[6] K. Hu and L. Yin, "Multi-scale topological features for hand posture representation and analysis," in *ICCV*. IEEE, 2013, pp. 1928–1935.

[7] J. Triesch and C. Von Der Malsburg, "Robust classification of hand postures against complex backgrounds," in *FG*, 1996, pp. 170 – 175.

[8] Y.-T. Li and J. P. Wachs, "HEGM: A hierarchical elastic graph matching for hand gesture recognition," *PR*, no. 765, 2014.

[9] N. H. Dardas and N. D. Georganas, "Real-Time Hand Gesture Detection and Recognition Using Bag-of-Features and Support Vector Machine Techniques," *TIM*, vol. 60, no. 11, pp. 3592–3607, Nov. 2011.

[10] P. K. Pisharady, P. Vadakkepat, and A. P. Loh, "Attention Based Detection and Recognition of Hand Postures Against Complex Backgrounds," *IJCV*, Aug. 2012.

[11] V.-T. Nguyen, T.-L. Le, T.-H. Tran, R. Mullot, and V. Courboulay, "Hand posture recognition using kernel descriptor," *Procedia Computer Science*, vol. 39, no. 0, pp. 154 – 157, 2014, iHCI.

[12] L. Bo, X. Ren, and D. Fox, "Kernel descriptors for visual recognition," in *NIPS*, 2010, pp. 244–252.

[13] D. G. Lowe, "Object recognition from local scale-invariant features," in *ICCV*, vol. 2, 1999, pp. 1150–1157 vol.2.

[14] S. Maji, A. C. Berg, and J. Malik, "Efficient classification for additive kernel svms," *TPAMI*, vol. 35, no. 1, pp. 66–77, 2013.

[15] A. Vedaldi and A. Zisserman, "Efficient additive kernels via explicit feature maps." *TPAMI*, vol. 34, no. 3, pp. 480–492, Mar. 2012.

[16] L. Bo and C. Sminchisescu, "Efficient match kernel between sets of features for visual recognition," in *NIPS*, 2009, pp. 135–143.

[17] V.-t. Nguyen, T.-l. Le, T.-t.-h. Tran, R. Mullot, and V. Courboulay, "A method for hand detection based on Internal Haar-like features and Cascaded AdaBoost Classifier," in *ICCE*, 2012, pp. 608–613.