

## Contents – Part II

### Full Papers Accepted for Poster Presentation

A Comparative Study for Known Item Visual Search Using Position Color Feature Signatures . . . . .	3
<i>Jakub Lokoč, David Kuboň, and Adam Blažek</i>	
A Novel Affective Visualization System for Videos Based on Acoustic and Visual Features. . . . .	15
<i>Jianwei Niu, Yiming Su, Shasha Mo, and Zeyu Zhu</i>	
A Novel Two-Step Integer-pixel Motion Estimation Algorithm for HEVC Encoding on a GPU. . . . .	28
<i>Keji Chen, Jun Sun, Zongming Guo, and Dachuan Zhao</i>	
A Scalable Video Conferencing System Using Cached Facial Expressions . . .	37
<i>Fang-Yu Shih, Ching-Ling Fan, Pin-Chun Wang, and Cheng-Hsin Hsu</i>	
A Unified Framework for Monocular Video-Based Facial Motion Tracking and Expression Recognition . . . . .	50
<i>Jun Yu</i>	
A Virtual Reality Framework for Multimodal Imagery for Vessels in Polar Regions . . . . .	63
<i>Scott Sorensen, Abhishek Kolagunda, Andrew R. Mahoney, Daniel P. Zitterbart, and Chandra Kambhamettu</i>	
Adaptive and Optimal Combination of Local Features for Image Retrieval . . .	76
<i>Neelanjan Bhowmik, Valérie Gouet-Brunet, Lijun Wei, and Gabriel Bloch</i>	
An Evaluation of Video Browsing on Tablets with the ThumbBrowser . . . . .	89
<i>Marco A. Hudelist and Klaus Schoeffmann</i>	
Binaural Sound Source Distance Reproduction Based on Distance Variation Function and Artificial Reverberation . . . . .	101
<i>Jiawang Xu, Xiaochen Wang, Maosheng Zhang, Cheng Yang, and Ge Gao</i>	
Color-Introduced Frame-to-Model Registration for 3D Reconstruction . . . . .	112
<i>Fei Li, Yunfan Du, and Rujie Liu</i>	
Compressing Visual Descriptors of Image Sequences. . . . .	124
<i>Werner Bailer, Stefanie Wechitsch, and Marcus Thaler</i>	

Deep Convolutional Neural Network for Bidirectional Image-Sentence Mapping . . . . .	136
<i>Tianyuan Yu, Liang Bai, Jinlin Guo, Zheng Yang, and Yuxiang Xie</i>	
Discovering Geographic Regions in the City Using Social Multimedia and Open Data . . . . .	148
<i>Stevan Rudinac, Jan Zahálka, and Marcel Worring</i>	
Discovering User Interests from Social Images . . . . .	160
<i>Jiangchao Yao, Ya Zhang, Ivor Tsang, and Jun Sun</i>	
Effect of Junk Images on Inter-concept Distance Measurement: Positive or Negative? . . . . .	173
<i>Yusuke Nagasawa, Kazuaki Nakamura, Naoko Nitta, and Noboru Babaguchi</i>	
Exploiting Multimodality in Video Hyperlinking to Improve Target Diversity . . . . .	185
<i>Rémi Bois, Vedran Vukotić, Anca-Roxana Simon, Ronan Sicre, Christian Raymond, Pascale Sébillot, and Guillaume Gravier</i>	
Exploring Large Movie Collections: Comparing Visual Berrypicking and Traditional Browsing . . . . .	198
<i>Thomas Low, Christian Hentschel, Sebastian Stober, Harald Sack, and Andreas Nürnberger</i>	
Facial Expression Recognition by Fusing Gabor and Local Binary Pattern Features . . . . .	209
<i>Yuechuan Sun and Jun Yu</i>	
Frame-Independent and Parallel Method for 3D Audio Real-Time Rendering on Mobile Devices. . . . .	221
<i>Yucheng Song, Xiaochen Wang, Cheng Yang, Ge Gao, Wei Chen, and Weiping Tu</i>	
Illumination-Preserving Embroidery Simulation for Non-photorealistic Rendering. . . . .	233
<i>Qiqi Shen, Dele Cui, Yun Sheng, and Guixu Zhang</i>	
Improving the Discriminative Power of Bag of Visual Words Model. . . . .	245
<i>Achref Ouni, Thierry Urruty, and Muriel Visani</i>	
M-SBIR: An Improved Sketch-Based Image Retrieval Method Using Visual Word Mapping . . . . .	257
<i>Jianwei Niu, Jun Ma, Jie Lu, Xuefeng Liu, and Zeyu Zhu</i>	
Movie Recommendation via BLSTM. . . . .	269
<i>Song Tang, Zhiyong Wu, and Kang Chen</i>	

Multimodal Video-to-Video Linking: Turning to the Crowd for Insight and Evaluation . . . . .	280
<i>Maria Eskevich, Martha Larson, Robin Aly, Serwah Sabetghadam, Gareth J.F. Jones, Roeland Ordelman, and Benoit Huet</i>	
Online User Modeling for Interactive Streaming Image Classification . . . . .	293
<i>Jiagao Hu, Zhengxing Sun, Bo Li, Kewei Yang, and Dongyang Li</i>	
Recognizing Emotions Based on Human Actions in Videos . . . . .	306
<i>Guolong Wang, Zheng Qin, and Kaiping Xu</i>	
Rocchio-Based Relevance Feedback in Video Event Retrieval . . . . .	318
<i>G.L.J. Pingen, M.H.T. de Boer, and R.B.N. Aly</i>	
Scale-Relation Feature for Moving Cast Shadow Detection. . . . .	331
<i>Chih-Wei Lin</i>	
Smart Loudspeaker Arrays for Self-Coordination and User Tracking . . . . .	343
<i>Jungju Jee and Jung-Woo Choi</i>	
Spatial Verification via Compact Words for Mobile Instance Search . . . . .	356
<i>Bo Wang, Jie Shao, Chengkun He, Gang Hu, and Xing Xu</i>	
Stochastic Decorrelation Constraint Regularized Auto-Encoder for Visual Recognition. . . . .	368
<i>Fengling Mao, Wei Xiong, Bo Du, and Lefei Zhang</i>	
The Perceptual Lossless Quantization of Spatial Parameter for 3D Audio Signals . . . . .	381
<i>Gang Li, Xiaochen Wang, Li Gao, Ruimin Hu, and Dengshi Li</i>	
Unsupervised Multiple Object Cosegmentation via Ensemble MIML Learning . . . . .	393
<i>Weichen Yang, Zhengxing Sun, Bo Li, Jiagao Hu, and Kewei Yang</i>	
Using Object Detection, NLP, and Knowledge Bases to Understand the Message of Images . . . . .	405
<i>Lydia Weiland, Ioana Hulpus, Simone Paolo Ponzetto, and Laura Dietz</i>	
Video Search via Ranking Network with Very Few Query Exemplars . . . . .	419
<i>De Cheng, Lu Jiang, Yihong Gong, Nanning Zheng, and Alexander G. Hauptmann</i>	

## Demonstrations

A Demo for Image-Based Personality Test . . . . .	433
<i>Huaiwen Zhang, Jiaming Zhang, Jitao Sang, and Changsheng Xu</i>	

A Web-Based Service for Disturbing Image Detection . . . . .	438
<i>Markos Zampoglou, Symeon Papadopoulos, Yiannis Kompatsiaris, and Jochen Spangenberg</i>	
An Annotation System for Egocentric Image Media . . . . .	442
<i>Aaron Duane, Jiang Zhou, Suzanne Little, Cathal Gurrin, and Alan F. Smeaton</i>	
DeepStyleCam: A Real-Time Style Transfer App on iOS . . . . .	446
<i>Ryosuke Tanno, Shin Matsuo, Wataru Shimoda, and Keiji Yanai</i>	
V-Head: Face Detection and Alignment for Facial Augmented Reality Applications . . . . .	450
<i>Zhiwei Wang and Xin Yang</i>	
<b>Video Browser Showdown</b>	
Collaborative Feature Maps for Interactive Video Search . . . . .	457
<i>Klaus Schoeffmann, Manfred Jürgen Primus, Bernd Muenzer, Stefan Petscharnig, Christof Karisch, Qing Xu, and Wolfgang Huerst</i>	
Concept-Based Interactive Search System. . . . .	463
<i>Yi-Jie Lu, Phuong Anh Nguyen, Hao Zhang, and Chong-Wah Ngo</i>	
Enhanced Retrieval and Browsing in the IMOTION System. . . . .	469
<i>Luca Rossetto, Ivan Giangreco, Claudiu Tănase, Heiko Schuldt, Stéphane Dupont, and Omar Seddati</i>	
Semantic Extraction and Object Proposal for Video Search . . . . .	475
<i>Vinh-Tiep Nguyen, Thanh Duc Ngo, Duy-Dinh Le, Minh-Triet Tran, Duc Anh Duong, and Shin'ichi Satoh</i>	
Storyboard-Based Video Browsing Using Color and Concept Indices . . . . .	480
<i>Wolfgang Hürst, Algernon Ip Vai Ching, Klaus Schoeffmann, and Manfred J. Primus</i>	
VERGE in VBS 2017 . . . . .	486
<i>Anastasia Mourtzidou, Theodoros Mironidis, Fotini Markatopoulou, Stelios Andreadis, Ilias Gialampoukidis, Damianos Galanopoulos, Anastasia Ioannidou, Stefanos Vrochidis, Vasileios Mezaris, Ioannis Kompatsiaris, and Ioannis Patras</i>	
Video Hunter at VBS 2017 . . . . .	493
<i>Adam Blažek, Jakub Lokoč, and David Kuboň</i>	
<b>Author Index</b> . . . . .	499

# Contents – Part I

## Full Papers Accepted for Oral Presentation

3D Sound Field Reproduction at Non Central Point for NHK 22.2 System. . .	3
<i>Song Wang, Ruimin Hu, Shihong Chen, Xiaochen Wang, Yuhong Yang, Weiping Tu, and Bo Peng</i>	
A Comparison of Approaches for Automated Text Extraction from Scholarly Figures . . . . .	15
<i>Falk Böschén and Ansgar Scherp</i>	
A Convolutional Neural Network Approach for Post-Processing in HEVC Intra Coding . . . . .	28
<i>Yuanying Dai, Dong Liu, and Feng Wu</i>	
A Framework of Privacy-Preserving Image Recognition for Image-Based Information Services . . . . .	40
<i>Kojiro Fujii, Kazuaki Nakamura, Naoko Nitta, and Noboru Babaguchi</i>	
A Real-Time 3D Visual Singing Synthesis: From Appearance to Internal Articulators . . . . .	53
<i>Jun Yu</i>	
A Structural Coupled-Layer Tracking Method Based on Correlation Filters . . .	65
<i>Sheng Chen, Bin Liu, and Chang Wen Chen</i>	
Augmented Telemedicine Platform for Real-Time Remote Medical Consultation . . . . .	77
<i>David Anton, Gregorij Kurillo, Allen Y. Yang, and Ruzena Bajcsy</i>	
Color Consistency for Photo Collections Without Gamut Problems . . . . .	90
<i>Qi-Chong Tian and Laurent D. Cohen</i>	
Comparison of Fine-Tuning and Extension Strategies for Deep Convolutional Neural Networks . . . . .	102
<i>Nikiforos Pittaras, Foteini Markatopoulou, Vasileios Mezaris, and Ioannis Patras</i>	
Describing Geographical Characteristics with Social Images . . . . .	115
<i>Huangjie Zheng, Jiangchao Yao, and Ya Zhang</i>	
Fine-Grained Image Recognition from Click-Through Logs Using Deep Siamese Network . . . . .	127
<i>Wu Feng and Dong Liu</i>	

Fully Convolutional Network with Superpixel Parsing for Fashion Web Image Segmentation . . . . .	139
<i>Lixuan Yang, Helena Rodriguez, Michel Crucianu, and Marin Ferecatu</i>	
Graph-Based Multimodal Music Mood Classification in Discriminative Latent Space. . . . .	152
<i>Feng Su and Hao Xue</i>	
Joint Face Detection and Initialization for Face Alignment . . . . .	164
<i>Zhiwei Wang and Xin Yang</i>	
Large-Scale Product Classification via Spatial Attention Based CNN Learning and Multi-class Regression . . . . .	176
<i>Shanshan Ai, Caiyan Jia, and Zhineng Chen</i>	
Learning Features Robust to Image Variations with Siamese Networks for Facial Expression Recognition . . . . .	189
<i>Wissam J. Baddar, Dae Hoe Kim, and Yong Man Ro</i>	
M3LH: Multi-modal Multi-label Hashing for Large Scale Data Search. . . . .	201
<i>Guan-Qun Yang, Xin-Shun Xu, Shanqing Guo, and Xiao-Lin Wang</i>	
Model-Based 3D Scene Reconstruction Using a Moving RGB-D Camera . . .	214
<i>Shyi-Chyi Cheng, Jui-Yuan Su, Jing-Min Chen, and Jun-Wei Hsieh</i>	
Modeling User Performance for Moving Target Selection with a Delayed Mouse . . . . .	226
<i>Mark Claypool, Ragnhild Eg, and Kjetil Raaen</i>	
Multi-attribute Based Fire Detection in Diverse Surveillance Videos . . . . .	238
<i>Shuangqun Li, Wu Liu, Huadong Ma, and Huiyuan Fu</i>	
Near-Duplicate Video Retrieval by Aggregating Intermediate CNN Layers. . .	251
<i>Giorgos Kordopatis-Zilos, Symeon Papadopoulos, Ioannis Patras, and Yiannis Kompatsiaris</i>	
No-Reference Image Quality Assessment Based on Internal Generative Mechanism. . . . .	264
<i>Xinchun Qian, Wengang Zhou, and Houqiang Li</i>	
On the Exploration of Convolutional Fusion Networks for Visual Recognition . . . . .	277
<i>Yu Liu, Yanming Guo, and Michael S. Lew</i>	
Phase Fourier Reconstruction for Anomaly Detection on Metal Surface Using Salient Irregularity . . . . .	290
<i>Tzu-Yi Hung, Sriram Vaikundam, Vidhya Natarajan, and Liang-Tien Chia</i>	

ReMagicMirror: Action Learning Using Human Reenactment with the Mirror Metaphor . . . . .	303
<i>Fabian Lorenzo Dayrit, Ryosuke Kimura, Yuta Nakashima, Ambrosio Blanco, Hiroshi Kawasaki, Katsushi Ikeuchi, Tomokazu Sato, and Naokazu Yokoya</i>	
Robust Image Classification via Low-Rank Double Dictionary Learning . . . .	316
<i>Yi Rong, Shengwu Xiong, and Yongsheng Gao</i>	
Robust Scene Text Detection for Multi-script Languages Using Deep Learning . . . . .	329
<i>Ruo-Ze Liu, Xin Sun, Hailiang Xu, Palaiahnakote Shivakumara, Feng Su, Tong Lu, and Ruoyu Yang</i>	
Robust Visual Tracking Based on Multi-channel Compressive Features . . . .	341
<i>Jianqiang Xu and Yao Lu</i>	
Single Image Super-Resolution with a Parameter Economic Residual-Like Convolutional Neural Network . . . . .	353
<i>Ze Yang, Kai Zhang, Yudong Liang, and Jinjun Wang</i>	
Spatio-Temporal VLAD Encoding for Human Action Recognition in Videos . . . . .	365
<i>Ionut C. Duta, Bogdan Ionescu, Kiyoharu Aizawa, and Nicu Sebe</i>	
Structure-Aware Image Resizing for Chinese Characters . . . . .	379
<i>Chengdong Liu, Zhouhui Lian, Yingmin Tang, and Jianguo Xiao</i>	
Supervised Class Graph Preserving Hashing for Image Retrieval and Classification . . . . .	391
<i>Lu Feng, Xin-Shun Xu, Shanqing Guo, and Xiao-Lin Wang</i>	
Visual Robotic Object Grasping Through Combining RGB-D Data and 3D Meshes. . . . .	404
<i>Yiyang Zhou, Wenhai Wang, Wenjie Guan, Yirui Wu, Heng Lai, Tong Lu, and Min Cai</i>	
What Convnets Make for Image Captioning? . . . . .	416
<i>Yu Liu, Yanming Guo, and Michael S. Lew</i>	
What are Good Design Gestures? –Towards User- and Machine-friendly Interface– . . . . .	429
<i>Ryo Kawahata, Atsushi Shimada, and Rin-ichiro Taniguchi</i>	

**SS1: Social Media Retrieval and Recommendation**

Collaborative Dictionary Learning and Soft Assignment for Sparse Coding  
of Image Features . . . . . 443  
*Jie Liu, Sheng Tang, and Yu Li*

LingoSent — A Platform for Linguistic Aware Sentiment Analysis for  
Social Media Messages . . . . . 452  
*Yuting Su and Huijing Wang*

Multi-Task Multi-modal Semantic Hashing for Web Image Retrieval with  
Limited Supervision . . . . . 465  
*Liang Xie, Lei Zhu, and Zhiyong Cheng*

Object-Based Aggregation of Deep Features for Image Retrieval. . . . . 478  
*Yu Bao and Haojie Li*

Uyghur Language Text Detection in Complex Background Images Using  
Enhanced MSERs . . . . . 490  
*Shun Liu, Hongtao Xie, Chuan Zhou, and Zhendong Mao*

**SS2: Modeling Multimedia Behaviors**

CELoF: WiFi Dwell Time Estimation in Free Environment . . . . . 503  
*Chen Yan, Peng Wang, Haitian Pang, Lifeng Sun, and Shiqiang Yang*

Demographic Attribute Inference from Social Multimedia Behaviors:  
A Cross-OSN Approach. . . . . 515  
*Liancheng Xiang, Jitao Sang, and Changsheng Xu*

Understanding Performance of Edge Prefetching . . . . . 527  
*Zhengyuan Pang, Lifeng Sun, Zhi Wang, Yuan Xie, and Shiqiang Yang*

User Identification by Observing Interactions with GUIs . . . . . 540  
*Zaher Hinbarji, Rami Albatal, and Cathal Gurrin*

Utilizing Locality-Sensitive Hash Learning for Cross-Media Retrieval . . . . . 550  
*Jia Yuhua, Bai Liang, Wang Peng, Guo Jinlin, Xie Yuxiang,  
and Yu Tianyuan*

**SS3: Multimedia Computing for Intelligent Life**

A Sensor-Based Official Basketball Referee Signals Recognition System  
Using Deep Belief Networks . . . . . 565  
*Chung-Wei Yeh, Tse-Yu Pan, and Min-Chun Hu*



Compact CNN Based Video Representation for Efficient Video Copy Detection . . . . .	576
<i>Ling Wang, Yu Bao, Haojie Li, Xin Fan, and Zhongxuan Luo</i>	
Cross-Modal Recipe Retrieval: How to Cook this Dish? . . . . .	588
<i>Jingjing Chen, Lei Pang, and Chong-Wah Ngo</i>	
Deep Learning Based Intelligent Basketball Arena with Energy Image. . . . .	601
<i>Wu Liu, Jiangyu Liu, Xiaoyan Gu, Kun Liu, Xiaowei Dai, and Huadong Ma</i>	
Efficient Multi-scale Plane Extraction Based RGBD Video Segmentation. . . .	614
<i>Hong Liu, Jun Wang, Xiangdong Wang, and Yueliang Qian</i>	
Human Pose Tracking Using Online Latent Structured Support Vector Machine. . . . .	626
<i>Kai-Lung Hua, Irawati Nurmala Sari, and Mei-Chen Yeh</i>	
Micro-Expression Recognition by Aggregating Local Spatio-Temporal Patterns . . . . .	638
<i>Shiyu Zhang, Bailan Feng, Zhineng Chen, and Xiangsheng Huang</i>	
egoPortray: Visual Exploration of Mobile Communication Signature from Egocentric Network Perspective . . . . .	649
<i>Qing Wang, Jiansu Pu, Yuanfang Guo, Zheng Hu, and Hui Tian</i>	
i-Stylist: Finding the Right Dress Through Your Social Networks . . . . .	662
<i>Jordi Sanchez-Riera, Jun-Ming Lin, Kai-Lung Hua, Wen-Huang Cheng, and Arvin Wen Tsui</i>	

#### **SS4: Multimedia and Multimodal Interaction for Health and Basic Care Applications**

Boredom Recognition Based on Users' Spontaneous Behaviors in Multiparty Human-Robot Interactions . . . . .	677
<i>Yasuhiro Shibasaki, Kotaro Funakoshi, and Koichi Shinoda</i>	
Classification of sMRI for AD Diagnosis with Convolutional Neuronal Networks: A Pilot 2-D+ $\epsilon$ Study on ADNI . . . . .	690
<i>Karim Aderghal, Manuel Boissenin, Jenny Benois-Pineau, Gwenaëlle Catheline, and Karim Afdel</i>	
Deep Learning for Shot Classification in Gynecologic Surgery Videos. . . . .	702
<i>Stefan Petscharnig and Klaus Schöffmann</i>	

Description Logics and Rules for Multimodal Situational Awareness  
in Healthcare . . . . . 714  
*Georgios Meditskos, Stefanos Vrochidis, and Ioannis Kompatsiaris*

Speech Synchronized Tongue Animation by Combining Physiology  
Modeling and X-ray Image Fitting . . . . . 726  
*Jun Yu*

Erratum to: ReMagicMirror: Action Learning  
Using Human Reenactment with the Mirror Metaphor . . . . . E1  
*Fabian Lorenzo Dayrit, Ryosuke Kimura, Yuta Nakashima,  
Ambrosio Blanco, Hiroshi Kawasaki, Katsushi Ikeuchi, Tomokazu Sato,  
and Naokazu Yokoya*

**Author Index . . . . . 739**

MultiMedia Modeling

23rd International Conference, MMM 2017, Reykjavik,  
Iceland, January 4-6, 2017, Proceedings, Part II

Amsaleg, L.; Guðmundsson, G.P.; Gurrin, C.; Jónsson,  
B.P.; Satoh, S. (Eds.)

2017, XXIV, 502 p. 234 illus., Softcover

ISBN: 978-3-319-51813-8