The 10th International Workshop on Robust Computer Vision (IWRCV 2015)

Saturday, Nov. 21 – Sunday, Nov. 22, 2015

Overseas Exchange Center of Peking University

Beijing, China

The University of Tokyo, KAIST,

Osaka University, Peking University,

Chonnam National University, Kagoshima University

The 10th International Workshop on Robust Computer Vision (IWRCV 2015)

Saturday, Nov. 21 – Sunday, Nov. 22, 2015

Overseas Exchange Center of Peking University

Beijing, China

The University of Tokyo, KAIST,

Osaka University, Peking University,

Chonnam National University, Kagoshima University

Organization

General chairs

- Katsushi Ikeuchi (Microsoft Research Asia) katsuike@microsoft.com
- Hongbin Zha (Peking University, China) zha@cis.pku.edu.cn

Program chairs

- In So Kweon (KAIST, Korea) iskweon@kaist.ac.kr
- Yasushi Yagi (Osaka University, Japan) yagi@am.sanken.osaka-u.ac.jp
- Chil-Woo Lee (Chonnam National University, Korea) leecw@chonnam.ac.kr
- Takeshi Oishi (The University of Tokyo, Japan) oishi@cvl.iis.u-tokyo.ac.jp
- Tong Lin (Peking University, China) lintong@pku.edu.cn

Map of Peking University

Information

Local Weather

Based on the weather forecast, it will snow and rain in Beijing on Nov. 21-22.

Wifi

Wireless connections are IWRCV2015-1N, IWRCV2015-1N5G, IWRCV2015-2N, and IWRCV2015-2N5G with same password “iwrcv2015”. **Caution**: Mobile phones should turn into the silence mode.

Coffee Breaks

Coffee breaks will be held at the Room 2111, Science Building #2, which is very close to and on the left side of the workshop venue (see the following map). Fruits, cakes, and drinks will be also provided.

Lunch Break

The lunch will be provided on Nov. 21, at the Room 2111 (same with coffee breaks) for professors and Room 2133 and Room 2135 for students. See the following map.

Poster Spotlight Session

Please copy the one-minute poster slides (ppt or pdf) to one laptop at the coffee breaks in the morning and the lunch break of Nov. 21, and rename the slides with the corresponding poster number (01-18). Please check if it can open. At the poster spotlight session, please wait at a queue to give the spotlight talk in order to save time.

Poster Session

All posters should stick up on the poster boards with the corresponding number 01-18 at the lunch break of Nov. 21. At the poster session, at least one author should be available to introduce the research work. When the poster session finishes at 17:40, the author should take away the poster because all poster boards will be moved outside soon.

Banquet

Volunteer students should help participants from Japan and Korea find the way to the Lakeview hotel. The banquet will start from 18:00 on Nov. 21.

Program

November 21, Saturday

Opening Remarks (9:00—9:05)

Plenary Talk 1 (9:05—9:50, Chair: Hongbin Zha)

Adaptive visual information processing induced by perceptual learning
Fang Fang, Peking University

Coffee Break (9:50—10:05)

Oral Session I (10:05—11:05, Chairs: Chil-Woo Lee, Jinshi Cui)

Fast Randomized Singular Value Thresholding for Nuclear Norm Minimization
Tae-Hyun Oh, Yasuyuki Matsushita, Yu-Wing Tai, In So Kweon

Adaptive Partial Differential Equation Learning for Visual Saliency Detection
Risheng Liu, Junjie Cao, Zhouchen Lin, Shiguang Shan

Changing Season in a Single Photograph by Unifying Color and Texture Transfer
Fumio Okura, Kenneth vanhoey, Adrien Bousseau, Alexei A. Efros, George Drettakis

Coffee Break (11:05—11:20)

Oral Session II (11:20—12:20, Chairs: Jinshi Cui, Xianghua Ying)

Robust and Accurate Aerial Scanning System
Ryoichi Ishikawa, Bo Zheng, XiangQi Huang, Takeshi Oishi, Katsushi Ikeuchi

Development of Light Field Projector and Its Applications
Hiroshi Kawasaki, Marco Visentini-Scarzanella, Ryo Furukawa, Shinsaku Hiura

Panorama to Cube: A Content-Aware Representation Method
Zeyu Wang, Xiaohan Jin, Fei Xue, Xin He, Renju Li, Hongbin Zha

Lunch Break (12:20—13:00)

Oral Session III (13:00—14:20, Chairs: Hiroshi Kawasaki, Xianghua Ying)

One-Day Outdoor Photometric Stereo via Skylight Estimation
Jiyoung Jung, Joon-Young Lee, In So Kweon

Spatio-Temporal Optimization-Based Motion Inpainting for Video Completion
Menandro Roxas, Takaaki Shiratori, Katsushi Ikeuchi

Complementary Sets of Shutter Sequences for Motion Deblurring
Hae-Gon Jeon, Joon-Young Lee, Yudeog Han, Seon Joo Kim, In So Kweon

Super Resolution of Fisheye Images Captured by On-Vehicle Camera for Visibility Assistance
Teruhisa Takano, Shintaro Ono, Yuki Matsushita, HIroshi Kawasaki, Katsushi Ikeuchi

Coffee Break (14:20—14:40)

Oral Session IV (14:40—16:20, Chairs: Takeshi Oishi, Gang Zeng)

Fine-Grained Object Classification: A General Approach
Shubhra Aich, Chil-Woo Lee

Depth-based Gait Authentication for Practical Sensor Settings
Taro Ikeda, Ikuhisa Mitsugami, and Yasushi Yagi

Video-Based Gait Analysis in Cerebrospinal Fluid Tap Test for Idiopathic Normal Pressure Hydrocephalus Patients
Ruochen Liao, Yasushi Makihara，Daigo Muramatsu，Ikuhisa Mitsugami，Yasushi Yagi，Kenji Yoshiyama，Hiroaki Kazui，Masatoshi Takeda

AttentionNet: Aggregating Weak Directions for Accurate Object Detection
Donggeun Yoo, Sunggyun Park, Joon-Young Lee, Anthony Paek, In So Kweon

Cell Detection for Automatic Drug Susceptibility Testing System
Kazuma Kikuchi, Andrey Grushnikov, Yoshimi Matsumoto, Kunihiko Nishino, Takeo Kanade, Yasushi Yagi

Poster Spotlight Session (16:20—16:40, Chair: Gang Zeng)

01. Robust Registration with Multiple Panoramic Cameras for Mixed Reality on Moving Vehicle
Kousuke Fujimoto, Yasuhide Okamoto, Takeshi Oishi, Katsushi Ikeuchi

02. Accurate Camera Calibration Robust to Defocus using a Smartphone
Hyowon Ha, Yunsu Bok, Kyungdon Joo, Jiyoung Jung, In So Kweon

03. Trajectory-based Stereo Visual Odometry with Statistical Outlier Rejection
Jiyuan Zhang, Rui Gan, Gang Zeng, Falong Shen, Hongbin Zha

04. Individuality-preserving Silhouette Extraction for Gait Recognition
Yasushi Makihara, Takuya Tanoue, Daigo Muramatsu, Yasushi Yagi, Syunsuke Mori, Yuzuko Utsumi, Masakazu Iwamura, Koichi Kise

05. An Efficient Approach for Eigen-joints-based 3D Activity Recognition
Hao Xu, Chilwoo Lee

06. Multispectral Pedestrian Detection: Benchmark Dataset and Baselines
Soonmin Hwang, Jaesik Park, Namil Kim, Yukyung Choi, In So Kweon

07. Occlusion Handling in Gait Recognition via Feature Regeneration
Daigo Muramatsu, Yasushi Makihara, Yasushi Yagi

08. Layered Contextual Model For Face Alignment With Group Sparse Feature
Falong Shen, Jiyuan Zhang, Rui Gan, Jingdong Wang, Gang Zeng, Hongbin Zha

09. Which Gait Feature Is Effective for Impairment Estimation?
Chengju Zhou, Ikuhisa Mitsugami, and Yasushi Yagi

10. Structure from Small Motion for Rolling Shutter Cameras
Sunghoon Im, Hyowon Ha, Gyeongmin Choe, Hae-Gon Jeon, Kyungdon Joo, In So Kweon

11. Moving Target Learning and Following for Collaborative Vision-Equipped Drone
Moju Zhao, Koji Kawasaki, Kei Okada, Masayuki Inaba

12. Gaze Direction Classification and Eye Fixation Analysis of Children Based on Camera Classes
Tiejian Zhang, Songjiang Li, Jinshi Cui, Li Wang, Xia Li, Wen Cui, Hongbin Zha

13. Learning a Deep Convolutional Network for Light-Field Image Super-Resolution
Youngjin Yoon, Hae-Gon Jeon, Donggeun Yoo, Joon-Young Lee, In So Kweon

14. Depth Estimation and Video Completion by Motion Analysis for Outdoor Omni-directional View
Carlos Morales, Shintaro Ono, Yasuhide Okamoto, Menandro Roxas, Takeshi Oishi, Katsushi Ikeuchi

15. Ellipse-Specific Fitting by Relaxing the 3l Constraints Using Semidefinite Programming
Jiangpeng Rong, Sen Yang, Xiang Mei, Xianghua Ying, Shiyao Huang, Hongbin Zha

16. Large-Population Dual-Task Dataset for Physical/Mental Condition Estimation
Ikuhisa Mitsugami, Chengju Zhou, Fumio Okura, Masataka Niwa, Yasushi Yagi

17. Development and Practice for the VR/MR Guided-Tour System
Yoshihiro Sato, Takeshi Oishi, Katsushi Ikeuchi

18. An Automatic System for Drug Susceptibility Testing
Andrey Grushnikov, Kazuma Kikuchi, Takeo Kanade, Yoshimi Matsumoto, Kunihiko Nishino, Yasushi Yagi

Poster Session and Coffee Break (16:40—17:40)

Banquet at the Lakeview Hotel (18:00—20:00)

November 22, Sunday

Plenary Talk 2 (9:00—9:45, Chair: Zhouchen Lin)

Structure Based Image Editing and Synthesis
Ralph R. Martin, Cardiff University, UK

Coffee Break (9:45—10:00)

Oral Session V (10:00—11:20, Chairs: Yuru Pei, Tong Lin)

Cluttered Object Separation through Active Robot Interactions
Krishneel Chaudhary, Chi Wun Au, Xiangyu Chen, Kotaro Nagahama, Hiroaki Yaguchi, Kei Okada, Masayuki Inaba

Reasoning based Vision Cognition Approach for Tomato Harvesting in Agriculture Automation
Xiangyu Chen, Krishneel Chaudhary, Hiroaki Yaguchi, Kei Okada, Masayuki Inaba

Off-Road Depth-Aided Visual Localization Using Mono Camera and LiDAR Sensor
Yufeng Yu, Huijing Zhao, Hongbin Zha

Ego-Centric Traffic Behavior Understanding through Multi-Level Vehicle Trajectory Analysis
Donghao Xu, Huijing Zhao, Hongbin Zha

Coffee Break (11:20—11:35)

Plenary Talk 3 (11:35—12:20, Chair: Huijing Zhao)

Task Modeling for Recognition, Reconstruction and Analysis of Folk Dances
Katsushi Ikeuchi, Microsoft Research Asia

Best Poster Award (12:20—12:25)

Closing Remarks (12:25—12:30)

PLENARY TALK

Adaptive visual information processing induced by perceptual learning
Fang Fang, Peking University

Structure Based Image Editing and Synthesis
Ralph R. Martin, Cardiff University, UK

Task Modeling for Recognition, Reconstruction and Analysis of Folk Dances
Katsushi Ikeuchi, Microsoft Research Asia

Adaptive Visual Information Processing Induced by Perceptual Learning

Fang Fang

Department of Psychology and PKU-IDG/McGovern Institute for Brain Research,

Peking University, Beijing 100871, China.

ffang@pku.edu.cn

Abstract: Visual perceptual learning refers to the phenomenon that training improves human perceptual abilities on sensory feature discrimination and object recognition. Using psychophysics, functional magnetic resonance imaging (MRI) and transcranial magnetic stimulation (TMS), we studied the behavioral characteristics and the neural mechanisms of visual perceptual learning over a long time course. The first part of my talk aims to investigate whether human visual cortical area(s) and or high-level decision-making area(s) could be altered by motion perceptual learning. We found that the long-term neural mechanisms of motion perceptual learning may be implemented by sharpening cortical tuning to trained stimuli at the sensory processing stage, and optimizing the connections between sensory and decision-making areas. The second part aims to evaluate the causal contributions of two visual areas (V3A and MT+) before and after motion perceptual learning. We found a significant transfer of learning from 100% coherence motion to 40% coherence motion. The TMS results indicate that perceptual learning could alter the contributions of specific visual areas in motion discrimination tasks at different coherence levels. In sum, our findings demonstrate that visual perceptual learning not only refines neural representations of trained stimuli within individual visual areas, but also alters cortico-cortico communications and the functional architecture of visual processing network.

Biography: Dr. Fang Fang is Chang Jiang Professor of Psychology, chair of the Department of Psychology, and executive associate director of the IDG/McGovern Institute for Brain Research at Peking University. He obtained a Ph.D. in Cognitive and Biological Psychology at the University of Minnesota in 2006, and was a Postdoctoral Research Associate between 2006 and 2007. His research seeks to understand the neural mechanisms of visual and cognitive processes by combining neuroimaging, psychophysical and computational techniques. Topics under investigation include object and face perception, visual adaptation, cortical plasticity, perceptual grouping, contextual modulation, visual attention and awareness. He currently serves on the editorial board for Current Biology, Experimental Brain Research, Frontiers in Biology, Frontiers in Perception Science and Science China: Life Sciences.

Structure Based Image Editing and Synthesis

Ralph R. Martin

School of Computer Science & Informatics

Cardiff University, UK

martinrr@cardiff.ac.uk, ralph@cs.cf.ac.uk

Abstract: Because of the semantic gap between digital image and human perception, traditional pixel/patch-based image editing techniques often fail to obtain satisfactory results by destroying the original structure of scene. By analyzing images at the structural level, we can find the most suitable region for editing with the most compatible content from image library or the image itself, making the editing process more intelligent. We introduce PatchNets, a compact, hierarchical representation describing structural and appearance characteristics of image regions; in a PatchNet, an image region with coherent appearance is summarized by a graph node associated with a single representative patch, and the geometric relationships among different regions are encoded by labelled graph edges giving contextual information. The hierarchical structure of a PatchNet allows a coarse-to-fine structural description of the image. This PatchNet representation can be used as a basis for interactive library-driven image editing. We also propose a data-driven image synthesis approach to extrapolate an image to a distinctly larger one by graph-based image structure representation. In this application, a novel sub-graph similarity measurement method is proposed to matching the local structures of different images. Images with the same sub-graph can be composite together to synthesize an extrapolated image.

Biography: Ralph R. Martin is currently a Professor at Cardiff University. He obtained his PhD degree in 1983 from Cambridge University. He has published more than 200 papers and 12 books, covering such topics as solid and surface modeling, intelligent sketch input, geometric reasoning, reverse engineering, and various aspects of computer graphics. He is a Fellow of: the Learned Society of Wales, the Institute of Mathematics and its Applications, and the British Computer Society. He is on the editorial boards of Computer Aided Design, Computer Aided Geometric Design, Geometric Models, the International Journal of Shape Modeling, CAD and Applications, and he is associate Editor-in-Chief of Computational Visual Media. He received a Friendship Award of Chinese Government in 2014.

Task Modeling for Recognition, Reconstruction and Analysis of Folk Dances

Katsushi Ikeuchi[1], Yoshihiro Sato[2], Shin'ichro Nakaoka[3], Shunsuke Kudoh[4]
Takahiro Okamoto[2], Hauchin Hu[2]

[1]*Microsoft Research Asia, China*
[2]*Institute of Industrial Science, The University of Tokyo, Japan*
[3]*National Institute of Advance Industrial Science and Technology, Japan*
[4]*The University of Electro-communications, Japan*

katsushi.ikeuchi@outlook.jp, yoshi@cvl.iis.u-tokyo.ac.kr,
s.nakaoka@aist.go.jp, kudoh@is.uec.ac.jp, kaikin@cvl.iis.u-tokyo.ac.jp

Intangible cultural assets such as folk dances and native languages have been disappearing day-by-day. It is important to develop new methods to preserve such assets. Toward this goal, this paper focuses on how to preserve folk dances as performances by humanoid robots. This new method provides not only preservation of such folk dances, but also understanding dance structures. We demonstrate this preservation method on a humanoid by using Japanese and Taiwanese folk dances. We also explain how such demonstrations provide new insights to folk dance, which leads interdisciplinary studies of Taiwanese folk dances.

This research was, in part, supported by Japan Society for the promotion of science, Grants-in-Aid for scientific research under science research A23240026.

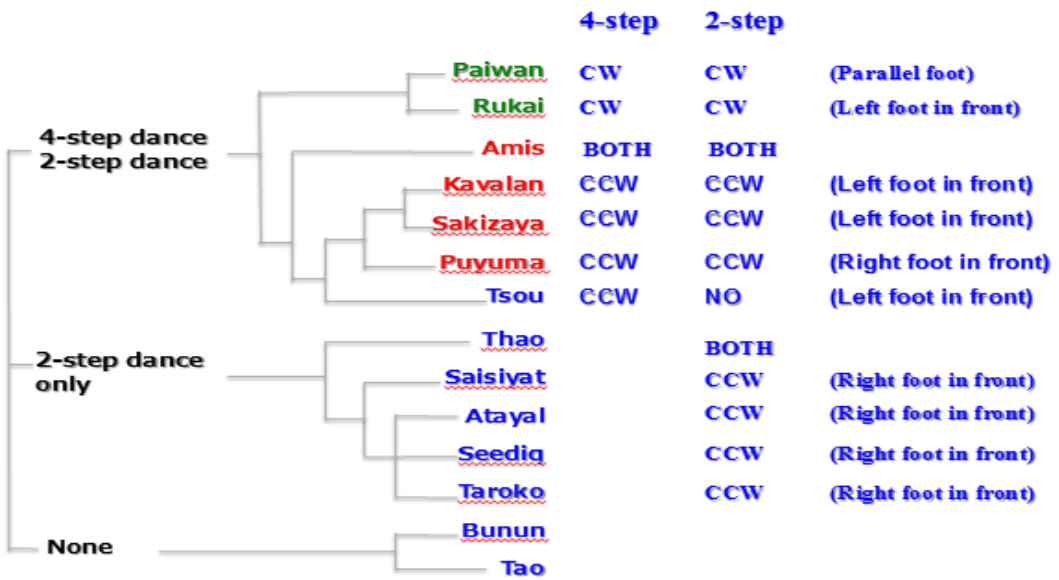


Fig.1. Taiwanese folk dance classification

ORAL SESSION I

Fast Randomized Singular Value Thresholding for Nuclear Norm Minimization
Tae-Hyun Oh, Yasuyuki Matsushita, Yu-Wing Tai, In So Kweon

Adaptive Partial Differential Equation Learning for Visual Saliency Detection
Risheng Liu, Junjie Cao, Zhouchen Lin, Shiguang Shan

Changing Season in a Single Photograph by Unifying Color and Texture Transfer
Fumio Okura, Kenneth vanhoey, Adrien Bousseau, Alexei A. Efros, George Drettakis

Fast Randomized Singular Value Thresholding for Nuclear Norm Minimization

Tae-Hyun Oh[1], Yasuyuki Matsushita[2], Yu-Wing Tai[3], and In So Kweon[1]

[1]*Robotics and Computer Vision Lab, KAIST, Korea*

[2] *Matsushita Lab, Osaka University, Japan*

[3] *SenseTime Group Limited, Hong Kong*

thoo@rcv.kaist.ac.kr, yasumat@ist.osaka-u.ac.jp, yuwing@sensetime.com, iskweon@kaist.ac.kr

Rank minimization problem can be boiled down to either Nuclear Norm Minimization (NNM) or Weighted NNM (WNNM) problem. The problems related to NNM (or WNNM) can be solved iteratively by applying a closed form proximal operator, called Singular Value Thresholding (SVT) (or Weighted SVT), but they suffer from high computational cost to compute a Singular Value Decomposition (SVD) at each iteration. In this paper, we propose an accurate and fast approximation method for SVT, called fast randomized SVT (FRSVT), where we avoid direct computation of SVD. The key idea is to extract an approximate basis for the range of a matrix from its compressed matrix. Given the basis, we compute the partial singular values of the original matrix from a small factored matrix. While the basis approximation is the bottleneck, our method is already several fold faster than thin SVD. By adopting a range propagation technique, we can further avoid one of the bottleneck at each iteration. Our theoretical analysis provides a stepping stone between the approximation bound of SVD and its effect to NNM via SVT. Along with the analysis, our empirical results on both quantitative and qualitative studies show our approximation rarely harms the convergence behavior of the host algorithms. We apply it and validate the efficiency of our method on various vision problems, e.g. subspace clustering, weather artifact removal, simultaneous multi-image alignment and rectification.

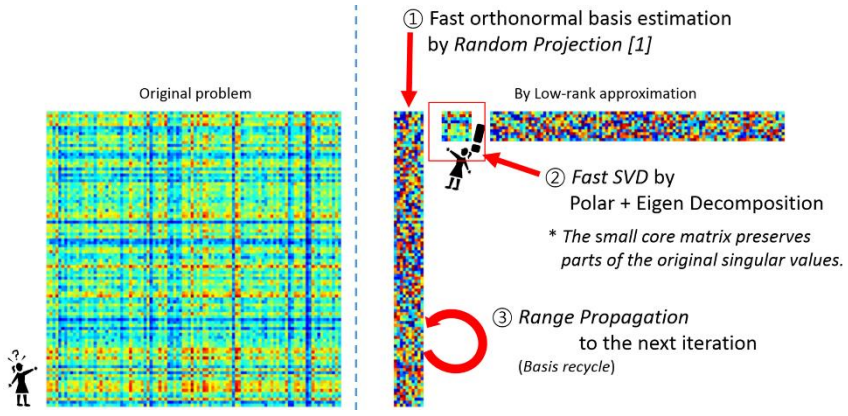


Figure 1: An overview of the fast randomized singular value thresholding for nuclear norm minimization.

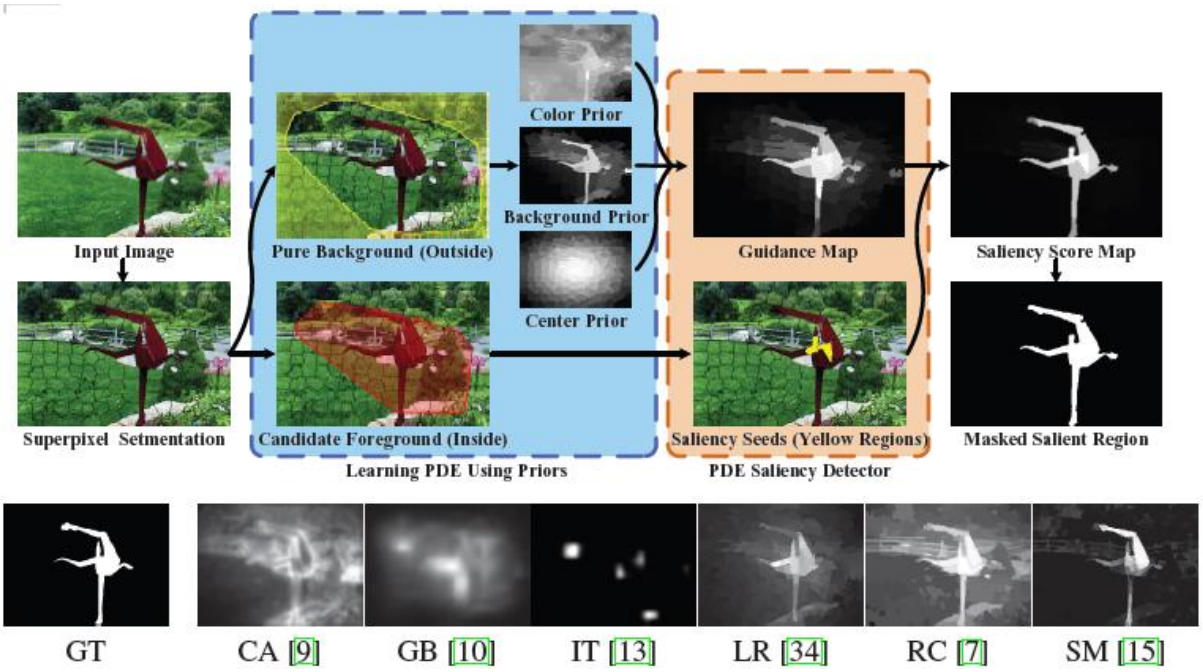
This research was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (No. 2010-0028680).

Adaptive Partial Differential Equation Learning for Visual Saliency Detection

Risheng Liu[y], Junjie Cao[y], Zhouchen Lin[] and Shiguang Shan[z]

[y]Dalian University of Technology
[]Key Lab. of Machine Perception (MOE), Peking University (zlin@pku.edu.cn)
[z]Key Lab. of Intelligent Information Processing of Chinese Academy of Sciences (CAS)

Partial Differential Equations (PDEs) have been successful in solving many low-level vision tasks. However, it is a challenging task to directly utilize PDEs for visual saliency detection due to the difficulty in incorporating human perception and high-level priors to a PDE system. Instead of designing PDEs with fixed formulation and boundary condition, this paper proposes a novel framework for adaptively learning a PDE system from an image for visual saliency detection. We assume that the saliency of image elements can be carried out from the relevances to the saliency seeds (i.e., the most representative salient elements). In this view, a general Linear Elliptic System with Dirichlet boundary (LESD) is introduced to model the diffusion from seeds to other relevant points. For a given image, we first learn a guidance map to fuse human prior knowledge to the diffusion system. Then by optimizing a discrete submodular function constrained with this LESD and a uniform matroid, the saliency seeds (i.e., boundary conditions) can be learnt for this image, thus achieving an optimal PDE system to model the evolution of visual saliency. Experimental results on various challenging image sets show the superiority of our proposed learning-based PDEs for visual saliency detection.



This research was supported by NSF China.

Changing Season in a Single Photograph by Unifying Color and Texture Transfer

Fumio Okura[1], Kenneth vanhoey[2], Adrien Bousseau[2], Alexei A. Efros[3], and George Drettakis[2]

[1]*Department of Intelligent Media, ISIR, Osaka University, Japan*

[2]*GRAPHDECO, INRIA, France*

[3]*Visual Computing Lab, University of California, Berkeley, USA*

okura@am.sanken.osaka-u.ac.jp

Recent color transfer methods use local information to learn the transformation from a source to an exemplar image, and then transfer this appearance change to a target image. These solutions achieve very successful results for general mood changes, e.g., changing the appearance of an image from “sunny” to “overcast”. However, such methods have a hard time creating new image content, such as leaves on a bare tree. Texture transfer, on the other hand, can synthesize such content but tends to destroy image structure.

We propose the first algorithm that unifies color and texture transfer, outperforming both by leveraging their respective strengths. A key novelty in our approach resides in teasing apart appearance changes that can be modeled simply as changes in color versus those that require new image content to be generated. Our method starts with an analysis phase which evaluates the success of color transfer by comparing the exemplar with the source. This analysis then drives a selective, iterative texture transfer algorithm that simultaneously predicts the success of color transfer on the target and synthesizes new content where needed. We demonstrate our unified algorithm by transferring large temporal changes between photographs, such as change of season --- e.g., leaves on bare trees or piles of snow on a street --- and flooding.

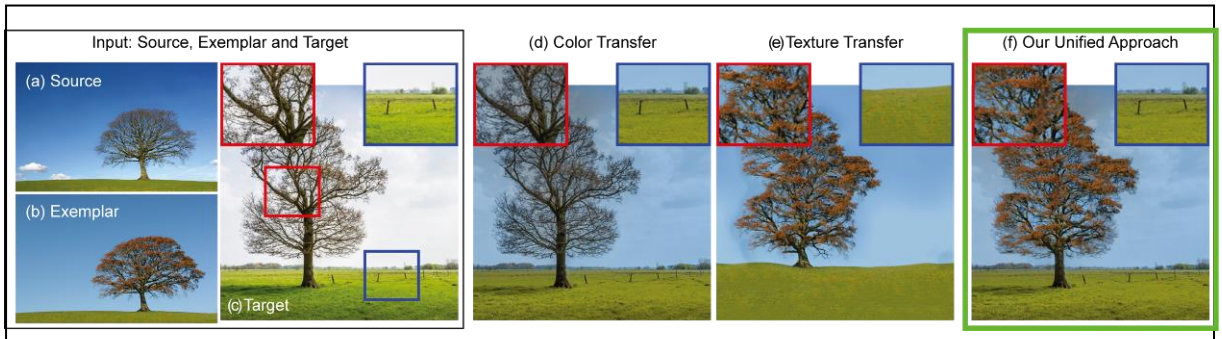


Fig. 1. Given *source*, *exemplar* and *target* images (left box), local color transfer does not change the texture on the tree in the target; while texture transfer destroys target structure such as the poles in the field and the bushes at the horizon. We unify the two approaches by predicting where color transfer is sufficient and where texture transfer is needed, which allows our algorithm to synthesize leaves on the tree while preserving other details of the target.

For details, please refer our CGF paper:

F. Okura, K. Vanhoey, A. Bousseau, A. A. Efros, and G. Drettakis. Unifying Color and Texture Transfer for Predictive Appearance Manipulation. *Computer Graphics Forum (Proc. EGSR)*, 34: 53–63, 2015.

ORAL SESSION II

Robust and Accurate Aerial Scanning System
Ryoichi Ishikawa, Bo Zheng, XiangQi Huang, Takeshi Oishi, Katsushi Ikeuchi

Development of Light Field Projector and Its Applications
Hiroshi Kawasaki, Marco Visentini-Scarzanella, Ryo Furukawa, Shinsaku Hiura

Panorama to Cube: A Content-Aware Representation Method
Zeyu Wang, Xiaohan Jin, Fei Xue, Xin He, Renju Li, Hongbin Zha

Robust and Accurate Aerial Scanning System

Ryoichi Ishikawa, Bo Zheng, XiangQi Huang, Takeshi Oishi and Katsushi Ikeuchi

Computer Vision Lab, The University of Tokyo, Japan

{zheng, huang, ishikawa, oishi, ki}@cvl.iis.u-tokyo.ac.jp

3D digital archiving of large structures like cultural heritage assets is useful technique, however lacks of data exist by occlusion in scans from the ground. We try to complement these missing data by scanning from high position with hanging omini-directional laser range sensor on balloon. Previously we proposed two rectification methods of distortion witch caused by sensor motion during scan [1, 2]. The method proposed in [1] is to register scan lines of aerial scan to ground data with smoothness constraint. In [2], sensor fusion method of laser range sensor and omini-directional camera is proposed. For achieving more robust and accurate results, in this paper, we present combination of these two methods. We first rectify distorted data by using method proposed in [2] and extract parameters of sensor motion. Second, we process the method proposed in [1] and obtain more accurate results.

In experiments, we demonstrate that our aerial scanning system achieves dramatically good robustness and accuracy for 3D digital archiving applications.

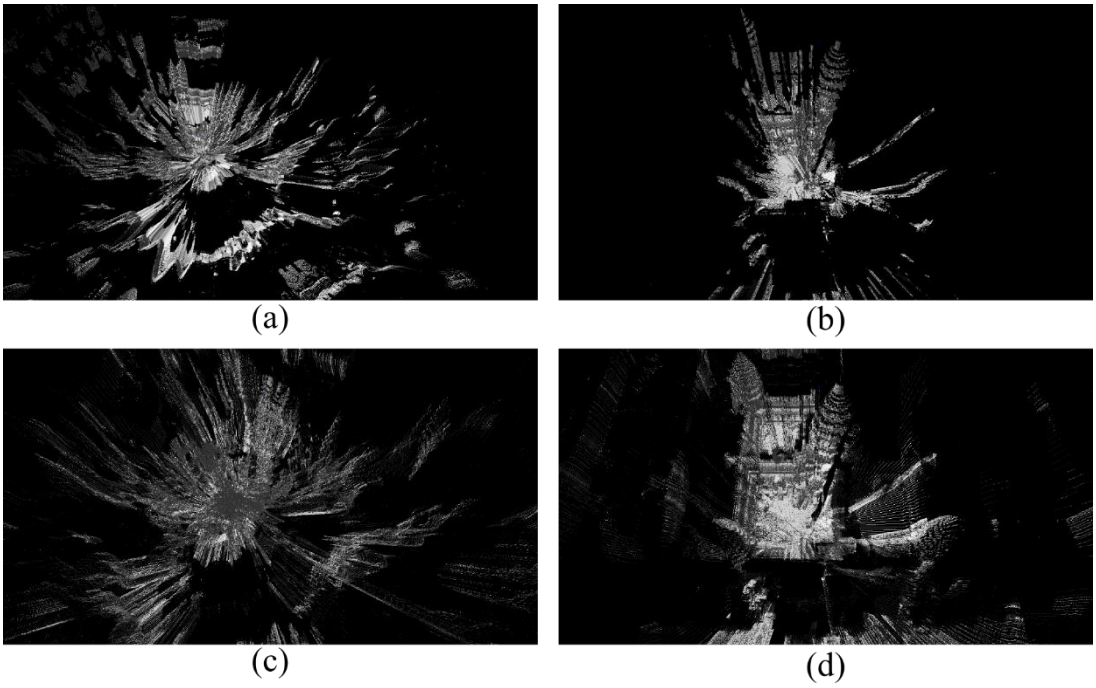


Fig. 1: 3D image of rectified aerial scans. (a) highly-distorted raw data. (b) Aerial scan and omini-directional camera method [2]. (c) Aerial scan and ground data method [1]. (d) Fusion method [1]+[2]

This work is partly supported by JSPS KAKENHI Grant Number 24254005 and 25257303 and Next-generation Energies for Tohoku Recovery (NET), MEXT, Japan.

[1] R. Ishikawa, B. Zheng, T. Oishi and K. Ikeuchi, “Rectification of aerial 3d laser scans via line-based registration to ground model” IPSJ Tran. on Computer Vision and Applications, pp.89-93, Jul, 2015.

[2] B. Zheng, X. Huang, R. Ishikawa, T. Oishi and K. Ikeuchi, “A New Flying Range Sensor: Aerial Scan in Omini-directions”, International Conference on 3D Vision (3DV), Oct, 2015.

Development of Light Field Projector and Its Applications

Hiroshi Kawasaki[1], Marco Visentini-Scarzanella[1], Ryo Furukawa[2], and Shinsaku Hiura[2]

[1]*Department of Information Systems and Biomedical Engineering,*

Kagoshima University, Japan

[2]*Department of Intelligent Systems, Hiroshima City University, Japan*

Recently, researches using projector are widely conducted in various areas, such as user interface, 3D reconstruction and so on. One of the severe problems of a projector is its narrow depth range of in-focus zone. For a camera, which is optically same as the projector, has the same problem, but solved by facilitating the light field theory using micro-polygon lens or multiple cameras. If a light field projector can be developed, such problem for a projector will be solved. In this talk, two types of a light field projector will be introduced.

The first system consists of video projector attached with a slit aperture in front of the lens, which realize depth-dependent and non-blurry pattern in wide depth range. Using the projector, 3D shapes can be reconstructed with wider depth range then a usual projector.

The second system consists of multiple video projectors, which project different patterns on the same screen at multiple depths. Those patterns are created by decomposing the target images for each projector. With the system, different images appear simultaneously at different depths.

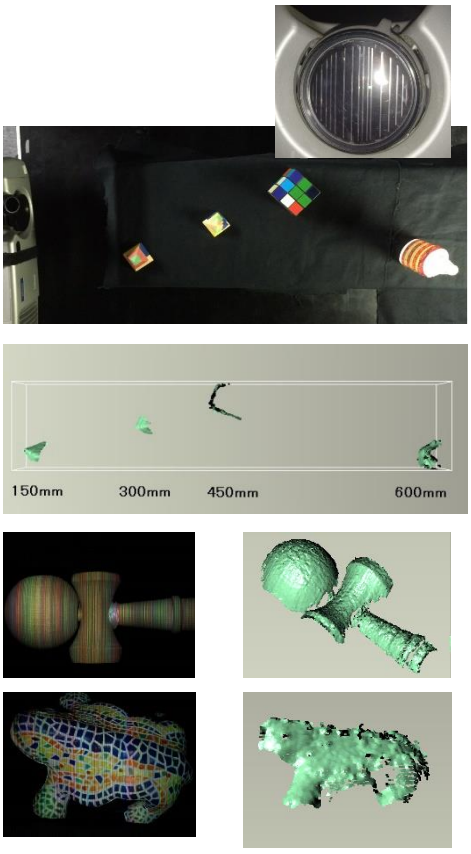


Figure 1. Wide depth range 3D capturing system.

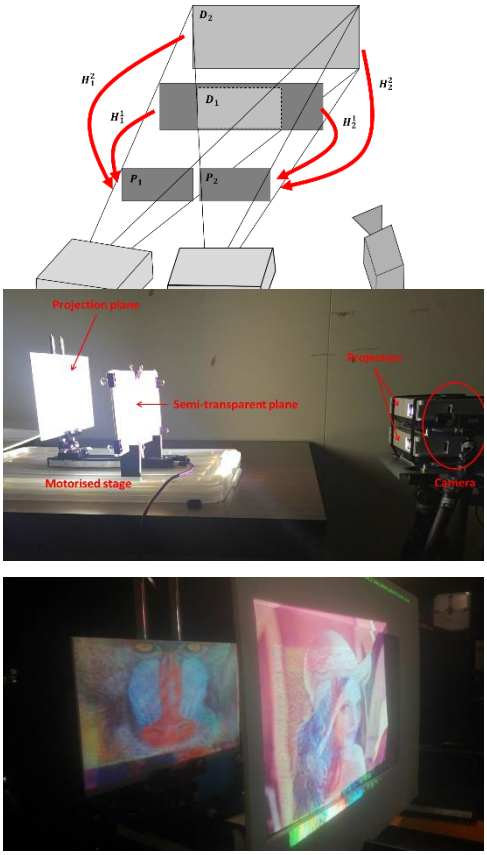


Figure 2. Multiple image projection at different depth.

Panorama to Cube: A Content-Aware Representation Method

Zeyu Wang, Xiaohan Jin, Fei Xue, Xin He, Renju Li, Hongbin Zha
Key Laboratory of Machine Perception, Peking University, China
{1200012927, 1200012629, 1200018415, hex}@pku.edu.cn
{lirenju, zha}@cis.pku.edu.cn

As panoramas provide a brand new viewpoint for the public, relevant cameras and software such as RICOH Theta, Microsoft Photosynth are embracing more and more users. However, the display methods for panoramas remain monotonous. In this paper, we propose a novel representation method called Content-Aware Cube Unwrapping using the effective and interactive techniques of orientational rectification, image modification and energy estimation. Thus, a number of fascinating applications will come into reality. For instance, six surfaces of a Rubik's cube can be automatically rendered from a vertically oriented panorama, without cutting any person or significant object apart. Moreover, seam carving and inserting are applied to each surface to enhance the key content and to make the scenery more consistent.

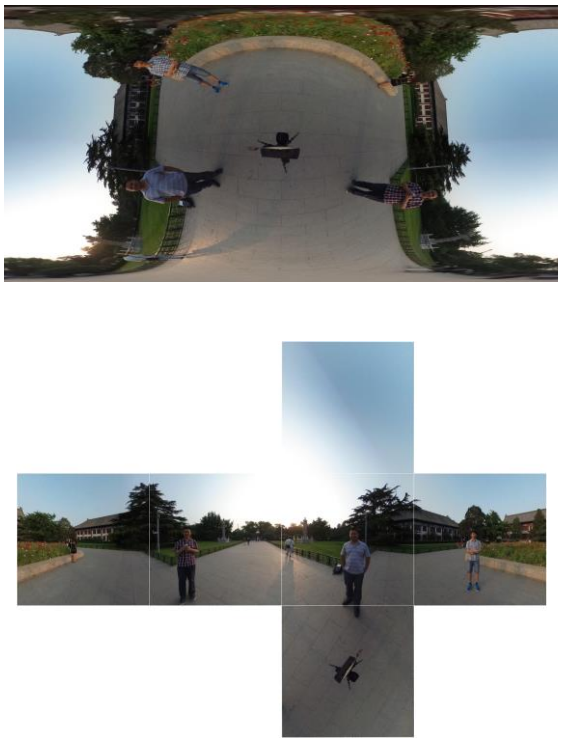
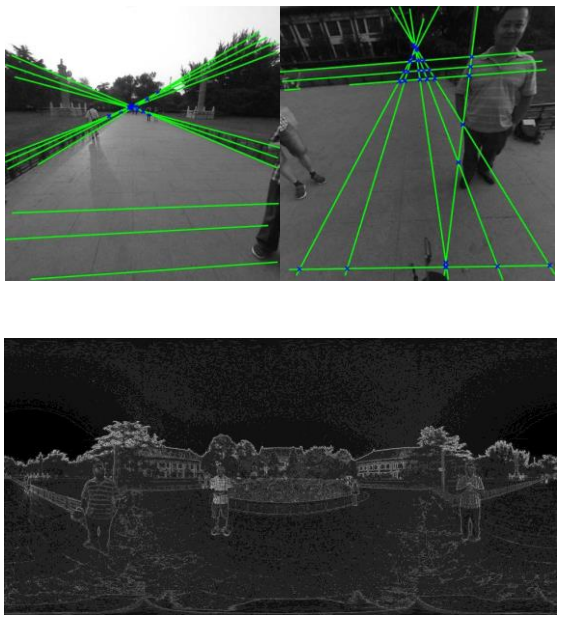


Fig. 2. A novel representation method for panoramas using Content-Aware Cube Unwrapping.



$$\tan\theta = \frac{1}{\cos\phi}$$

Fig. 2. Vanishing point estimation, energy function synthesis, and spherical panorama unwrapping according to geometric constraints.

This research was supported by the Undergraduate Research Training Program, Peking University.

ORAL SESSION III

One-Day Outdoor Photometric Stereo via Skylight Estimation
Jiyoung Jung, Joon-Young Lee, In So Kweon

Spatio-Temporal Optimization-Based Motion Inpainting for Video Completion
Menandro Roxas, Takaaki Shiratori, Katsushi Ikeuchi

Complementary Sets of Shutter Sequences for Motion Deblurring
Hae-Gon Jeon, Joon-Young Lee, Yudeog Han, Seon Joo Kim, In So Kweon

Super Resolution of Fisheye Images Captured by On-Vehicle Camera for Visibility Assistance
Teruhisa Takano, Shintaro Ono, Yuki Matsushita, HIroshi Kawasaki, Katsushi Ikeuchi

One-day outdoor photometric stereo via skylight estimation

Jiyoung Jung, Joon-Young Lee, and In So Kweon
Robotics and Computer Vision Lab, KAIST, Korea
{*jyjung, jylee*}*@rcv.kaist.ac.kr, iskweon@kaist.ac.kr*

Traditional photometric stereo methods require three or more input images under different distant point light sources of which the directions are nonplanar on the unit sphere. An outdoor environment has been regarded as full of unknowns and complexities compared to a controllable lab environment. The appearance of an open field changes drastically depending on its weather condition and time of day, but at the same time, it does have a general appearance. While the appearance of a room can easily be influenced by which kind of light we turn on, an outdoor field on a clear day presents a relatively predictable scene.

We present an outdoor photometric stereo method based on the motivation that the outdoor illumination which is mainly contributed by the sun and clear sky can be generally modeled. We process geo-tagged, timestamped images captured from a static camera in a single day to estimate the surface normal of the scene. We simulate a sky hemisphere for each image according to its GPS and timestamp, and parameterize the obtained sky hemisphere into a quadratic skylight and a Gaussian sunlight distribution. Unlike previous works which usually model outdoor illumination as a sum of constant ambient light and a distant point light, our method models natural illumination according to a popular sky model and thus provides sufficient constraints for shape reconstruction from one day images. We generate pixel profiles of uniformly sampled unit vectors for the corresponding time of captures and evaluate them using correlation with the actual pixel profiles. The estimated surface normal is refined by MRF optimization. We have tested our method to recover objects and scenes of various sizes in real-world outdoor daylight and compared with the previous methods [1, 2].

(a) Input image (b) Albedo estimation (c) Normal estimation (d) 3D reconstruction

Figure 1: An example of one-day outdoor photometric stereo result.

[1] A.Abrams, C.Hawley, and R.Pless, “Heliometric stereo: shape from sun position”, ECCV 2012.

[2] A.Abrams, K.Miskell, and R.Pless, “The episolar constraint: monocular shape from shadow correspondence”, CVPR 2013.

This research was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (No. 2010-0028680).

Spatio-Temporal Optimization-Based Motion Inpainting for Video Completion

Menandro Roxas[1], Takaaki Shiratori[2], and Katsushi Ikeuchi[2]

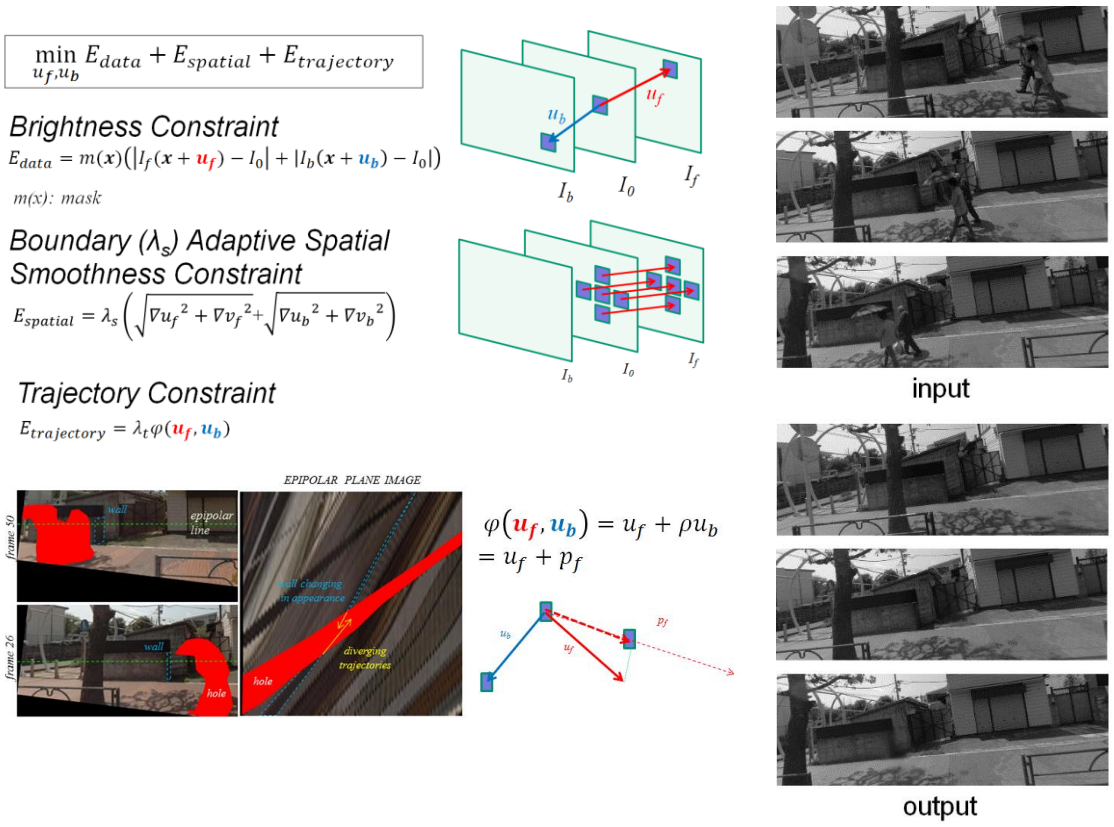
[1]*Computer Vision Laboratory, The University of Tokyo, Japan*

[2]*Microsoft Research Asia, China*

roxas@cvl.iis.u-tokyo.ac.jp, takaakis@microsoft.com, katsuike@microsoft.com

In this work, we propose a method to complete damaged videos using motion inpainting (MI) and color propagation (CP). We first constraint the interpolated motion of the regions to be completed, called holes, to be spatio-temporally smooth. To achieve this, we simultaneously solve and interpolate the motion of the known regions and the hole by minimizing an optical flow estimation function with the spatio-temporal smoothness constraints.

The solved optical flow is used to propagate the color of the known pixels to the hole using bicubic warping. We embed this two-step (MI+CP) method in an iterative optimization framework where we use the newly inpainted color to further improve the optical flow estimation. This is done by introducing a spatially varying mask function that is dependent on the frame distance of the source of the inpainted color. We also accurately impose the temporal smoothness constraint by solving a trajectory prior based on the camera's egomotion. We propose a fast estimation of the translation parameters through point correspondence among only three frames.



Complementary Sets of Shutter Sequences and its applications to motion deblurring and privacy-protecting video surveillance

Hae-Gon Jeon, Joon-Young Lee, Yudeog Han, and In So Kweon

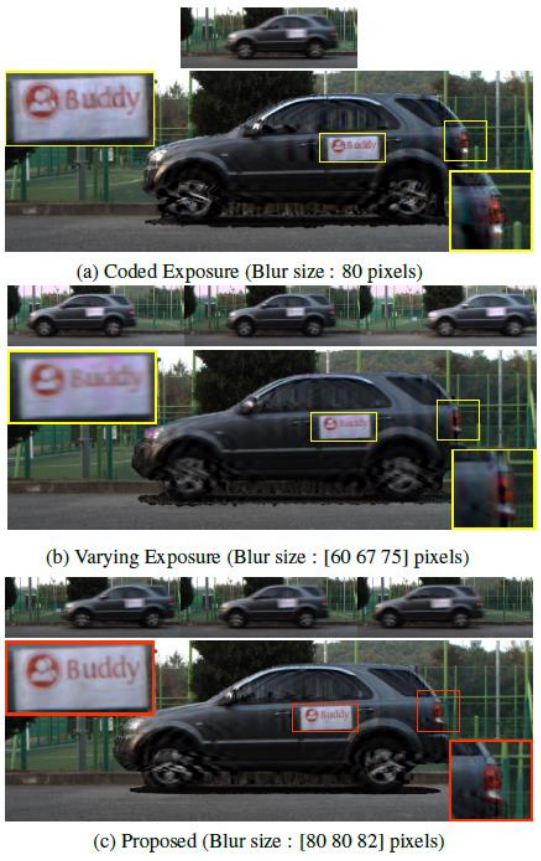
Robotics and Computer Vision Lab, KAIST, Korea

hgjeon@rcv.kaist.ac.kr, jylee@rcv.kaist.ac.kr, ydhan@rcv.kaist.ac.kr, iskweon@kaist.ac.kr

In this paper, we present a novel multi-image motion deblurring method utilizing the coded exposure technique. The key idea of our work is to capture video frames with a set of complementary fluttering patterns to preserve spatial frequency details.

We introduce an algorithm for generating a complementary set of binary sequences based on the modern communication theory and implement the coded exposure video system with an off-the-shelf machine vision camera.

The effectiveness of our method is demonstrated on various challenging examples with quantitative and qualitative comparisons to other computational image capturing methods used for image deblurring.



(a) Coded Exposure (Blur size : 80 pixels)

(b) Varying Exposure (Blur size : [60 67 75] pixels)

(c) Proposed (Blur size : [80 80 82] pixels)

Fig. 3. Outdoor deblurring result

(a) Random sequences set

(b) MURA

(b) Proposed

Fig. 4. Privacy-protecting video surveillance

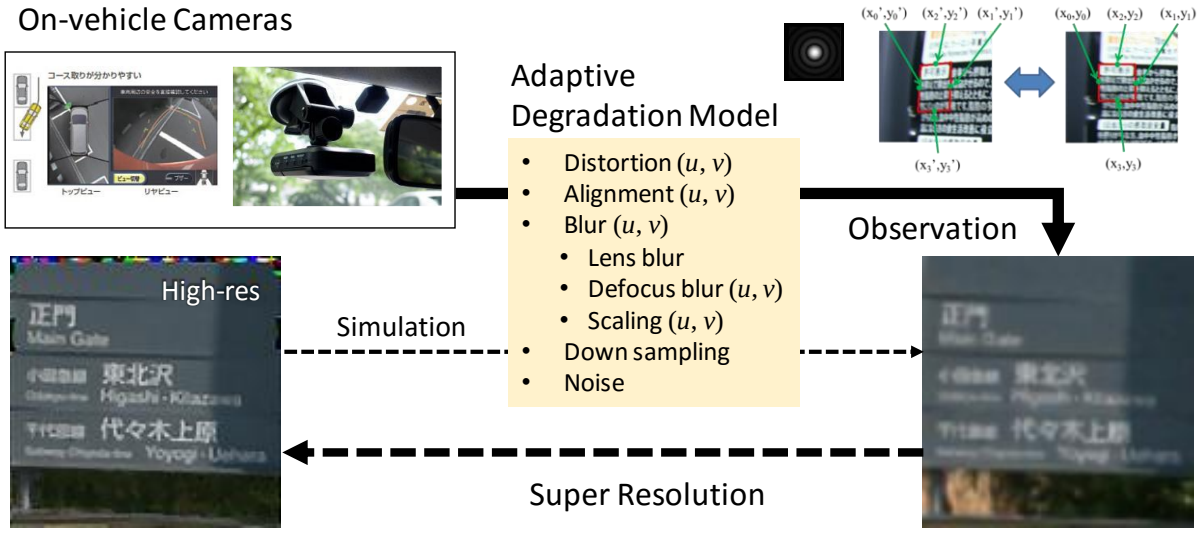
This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIP) (No.2010- 0028680). Hae-Gon Jeon was partially supported by Global PH.D Fellowship Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education (NRF-2015H1A2A1034617).

Super Resolution of Fisheye Images Captured by On-Vehicle Camera for Visibility Assistance

Teruhisa Takano[1], Shintaro Ono[1], Yuki Matsushita[2],
Hiroshi Kawasaki[2], and Katsushi Ikeuchi[3,1]
[1]*Institute of Industrial Science, The University of Tokyo, Japan*
[2]*Graduate School of Science and Engineering, Kagoshima University, Japan*
[3]*Microsoft Research Asia, China*
{teruhisa-takano,onoshin}@cvl.iis.u-tokyo.ac.jp,
{sc108046, kawasaki}@ibe.kagoshima-u.ac.jp, katsuike@microsoft.com

Fisheye cameras have become widely used for on-vehicle camera, especially of general users, to assist their visibility in driving and parking, or to record the surrounding traffic situation. When such a huge amount video become available online, various applications will become possible, although the resolution and image quality are relatively low compared to the recent digital cameras for ordinary use.

In this work, we present a method to restore the high resolution images from an on-vehicle fisheye camera with reconstruction-based super resolution. Since the fisheye lens give non-uniform characteristic within the sight unlike the ordinary perspective cameras, we designed the adaptive image degradation model for capturing fisheye images, which is caused by lens blur, defocus blur, and image scaling, depending on the position within the sight, so that we can simulate the image degradation process more properly. The experimental results showed the effectiveness compared with the result using uniformal (non-adaptive) blur model.



ORAL SESSION IV

Fine-Grained Object Classification: A General Approach
Shubhra Aich, Chil-Woo Lee

Depth-based Gait Authentication for Practical Sensor Settings
Taro Ikeda, Ikuhisa Mitsugami, and Yasushi Yagi

Video-Based Gait Analysis in Cerebrospinal Fluid Tap Test for Idiopathic Normal Pressure Hydrocephalus Patients
Ruochen Liao, Yasushi Makihara，Daigo Muramatsu，Ikuhisa Mitsugami，Yasushi Yagi，Kenji Yoshiyama，Hiroaki Kazui，Masatoshi Takeda

AttentionNet: Aggregating Weak Directions for Accurate Object Detection
Donggeun Yoo, Sunggyun Park, Joon-Young Lee, Anthony Paek, In So Kweon

Cell Detection for Automatic Drug Susceptibility Testing System
Kazuma Kikuchi, Andrey Grushnikov, Yoshimi Matsumoto, Kunihiko Nishino, Takeo Kanade, Yasushi Yagi

Fine-Grained Object Classification: A General Approach

Shubhra Aich, and Chil-Woo Lee

Chonnam National University, Korea Rep.

s.aich.72@gmail.com, leecw@jnu.ac.kr

We deal with the classification problem of visually similar objects which is also known as fine-grained recognition. We consider both rigid and non-rigid types of objects. We investigate the classification performance of different combinations of bag-of-visual words models to find out a generalized set of visual words for different types of fine-grained classification. We use different linear combination of the color models for different color, shape and texture feature extraction. The weights of the combination of the color models are determined from the training set using inter-class separability measures. We combine the feature sets using multi-class multiple learning algorithm. We evaluate the models on two datasets; in the non-rigid, deformable object category, Oxford 102 class flower dataset is chosen and 17 class make and model recognition car dataset is selected in the rigid category. The recognition performance on both dataset is found to be satisfactory. Figure 1 below shows the block diagram of our workflow.

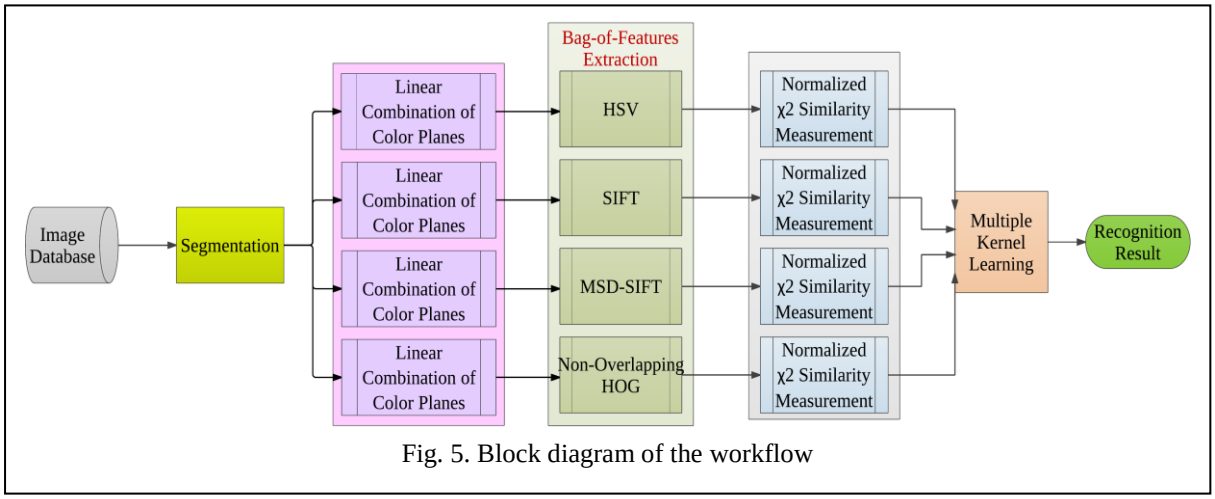


Fig. 5. Block diagram of the workflow

This research is financially supported by the Ministry of Education, Science and Technology (MEST) and National Research Foundation of Korea (NRF) through the Human Resource Training Project for Regional Innovation.

Depth-based Gait Authentication for Practical Sensor Settings

Taro Ikeda[1], Ikuhisa Mitsugami[1], and Yasushi Yagi[1]

[1]*Osaka University, 8-1 Mihogaoka, Ibaraki-shi, Osaka, 567-0047 Japan*

ikeda@am.sanken.osaka-u.ac.jp, mitsugami@am.sanken.osaka-u.ac.jp, yagi@am.sanken.osaka-u.ac.jp

Gait, way of walking, is regarded as one of biometrics as with fingerprint, vein, and iris. Gait has an advantage that it can be obtained without any contact to devices. Considering this great property of gait, we would like to apply gait authentication for automatic security door, visitor logging, and global people tracking, for example. In such cases, we have to locate a sensor quite near to a walk way. However, no one investigates the performance of gait authentication considering such a situation. In this paper, therefore, we investigate performances of silhouette-based and depth-based gait authentication considering practical sensor settings where sensors are located in an environments afterwards such as an entrance/exit of buildings shown in Fig. 1. Sensors have to be located quite near to people in such situations. To realize fair comparison between different sensors and methods, we construct full-body volume of walking people by a multi-camera environment shown in Fig. 2 so as to reconstruct virtual silhouette and depth images at arbitrary sensor positions. In addition, we also investigate performances when we have to authenticate between frontal and rear views. Experimental results confirm that the depth-based methods outperform the silhouette-based ones in the realistic situations. We also confirm that by introducing Depth-based Gait Feature, we can authenticate between the frontal and rear views.

Fig. 1. Entrance of building

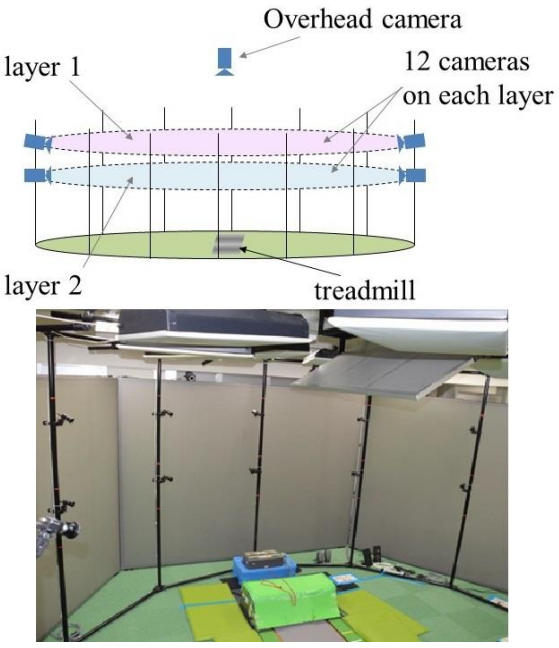


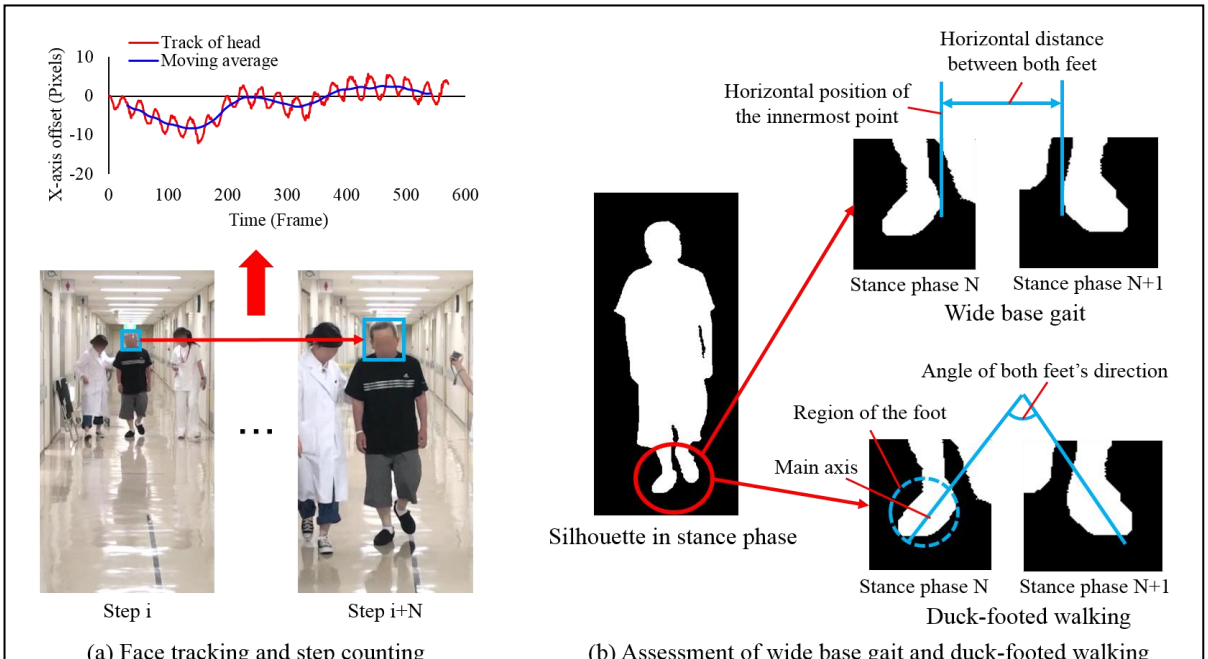
Fig.2. Multi-camera environment

Video-Based Gait Analysis in Cerebrospinal Fluid Tap Test for Idiopathic Normal Pressure Hydrocephalus Patients

Ruochen Liao[1], Yasushi Makihara[1] , Daigo Muramatsu[1] , Ikuhisa Mitsugami[1] , Yasushi Yagi[1] ,
Kenji Yoshiyama[2] , Hiroaki Kazui[2] , Masatoshi Takeda[2]
[1] *ISIR, Osaka Univ., Japan*
[2] *Graduate School of Medicine, Osaka Univ., Japan*
{liao, makihara, muramatsu, mitsugami, yagi}@am.sanken.osaka-u.ac.jp
{yosiyama, kazui, mtakeda}@ psy.med.osaka-u.ac.jp

Recently, idiopathic normal pressure hydrocephalus (iNPH) draw a lot of attention as a treatable dementia. iNPH develops in elderly population, causes gait disturbance and urinary incontinence in addition to dementia, and the gait is frequently used in diagnosis. In clinical use, patients' gait usually evaluated by visual inspection or time required of walking tests. However, the visual inspection strongly dependent on the subjectivity of the medical doctors, and it is difficult to determine the result by only a single quantitative item, the time required. It is therefore necessary to develop a method to assess the features of iNPH patients' gait disturbance quantitatively and from multiple aspects.

In this study, we proposed a video-based assessment method for evaluating the gait of iNPH patients. We assessed 4 features: lateral sway, petit-pas gait, wide base gait and duck-footed walking, as the symptoms of gait disturbance caused by iNPH, and compared the result with the diagnosis of the medical doctors toward effective prediction of the surgical treatment.



(a) Face tracking and step counting (b) Assessment of wide base gait and duck-footed walking

Fig. 6. Gait assessment for iNPH diagnosis

AttentionNet: Aggregating Weak Directions for Accurate Object Detection

Donggeun Yoo[1], Sunggyun Park[2], Joon-Young Lee[1], Anthony S. Paek[3] and In So Kweon[1]

[1]*Robotics and Computer Vision Lab, KAIST, Korea.*

[2]*Risk Analysis and Management Lab, KAIST, Korea.*

[3]*Lunit Inc., Korea.*

dgyoo@rcv.kaist.ac.kr, sunggyun@kaist.ac.kr, jylee@rcv.kaist.ac.kr, apaek@lunit.io, iskweon@kaist.ac.kr

We present a novel detection method using a deep convolutional neural network (CNN), named AttentionNet. We cast an object detection problem as an iterative classification problem, which is the most suitable form of a CNN. AttentionNet provides quantized weak directions pointing a target object and the ensemble of iterative predictions from AttentionNet converges to an accurate object boundary box. Since AttentionNet is a unified network for object detection, it detects objects without any separated models from the object proposal to the post bounding-box regression. We evaluate AttentionNet by a human detection task and achieve the state-of-the-art performance of 65% (AP) on PASCALVOC 2007/2012 with an 8-layered architecture only.

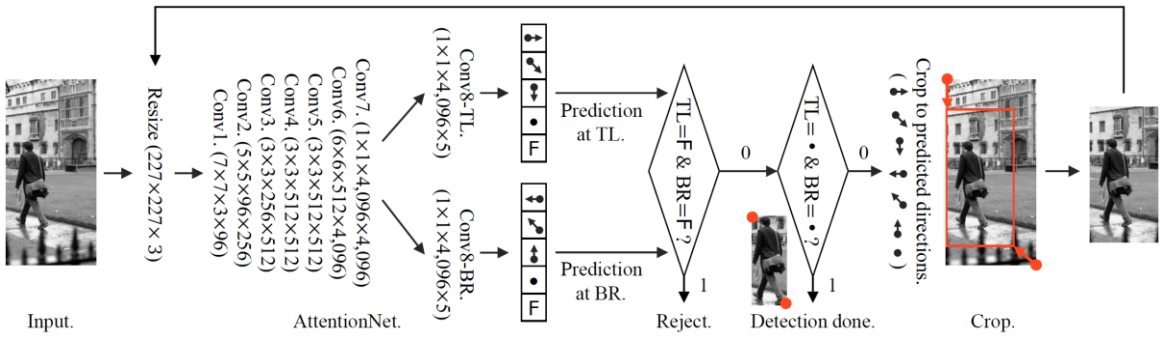


Fig. 7. A pipeline of our detection framework.

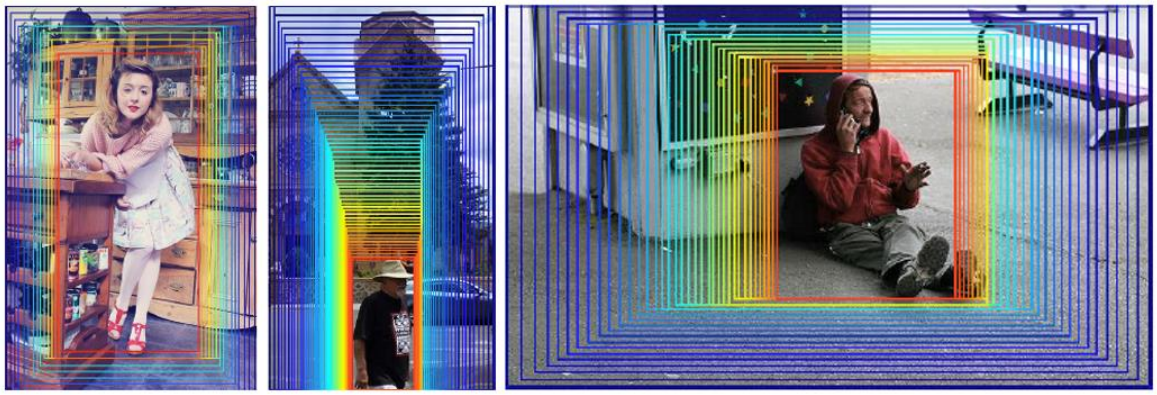


Fig. 2. Real detection examples of our detection framework.

This work was supported by the Technology Innovation Program (No. 10048320), funded by Korea government (MOTIE).

Cell Detection for Automatic Drug Susceptibility Testing System

Kazuma Kikuchi[1], Andrey Grushnikov[1], Yoshimi Matsumoto[1], Kunihiko Nishino[1],
Takeo Kanade[1]and Yasushi Yagi[1]
[1] *Institute of Scientific and Industrial Research, Osaka University*
kikuchi@am.sanken.osaka-u.ac.jp andrey@am.sanken.osaka-u.ac.jp

Drug Susceptibility Testing (DST) determines whether a drug has an effect on a bacteria strain or not. In recent years a number of bacteria strains resistant to modern antibiotics has increased significantly, thus effective and rapid DST method is highly desirable. Drug Susceptibility Testing Microfluidic device (DSTM) was designed to resolve this problem. This device consists of multiple channels printed in polymer on a glass cover. Examiner injects bacteria strain and a drug into the DSTM then observes drug's affect via visual changes relevant to drug susceptibility. However the main problem of this approach is that the result is not reliable, since this process is performed manually and is affected by examiner's subjective feeling.

To make testing with DSTM more accurate and objective, we developed an automatic testing system. This system has several components: channel localization, cell detection and feature extraction. Cell detection is the most important part of it. Wide variety of bacteria sizes, low quality of the captured images, overlapping of elongated cells make this task difficult.

Our main focus is separation of the overlapped cells. It is performed in 4 steps. First, connected components representing cell candidates are extracted and labeled from the binary image received from preprocessed input image. Next, for each component intersections are detected, then it is modeled as cell graph. Finally, segments are merged according to smoothness constraints maximizing length [Fig. 1].

Evaluation of the developed algorithm is hard since ground truth is not easily acquired.
Instead of the evaluation, we confirmed the validity with bioscience expert's check and the system received a valuation.

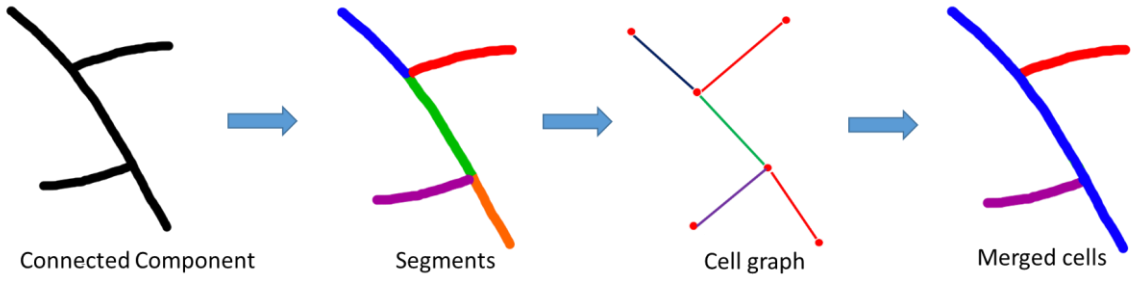


Fig. 8. Overview of cell detection algorithm

POSTER SPOTLIGHT SESSION

01. Robust Registration with Multiple Panoramic Cameras for Mixed Reality on Moving Vehicle
Kousuke Fujimoto, Yasuhide Okamoto, Takeshi Oishi, Katsushi Ikeuchi

02. Accurate Camera Calibration Robust to Defocus using a Smartphone
Hyowon Ha, Yunsu Bok, Kyungdon Joo, Jiyoung Jung, In So Kweon

03. Trajectory-based Stereo Visual Odometry with Statistical Outlier Rejection
Jiyuan Zhang, Rui Gan, Gang Zeng, Falong Shen, Hongbin Zha

04. Individuality-preserving Silhouette Extraction for Gait Recognition
Yasushi Makihara, Takuya Tanoue, Daigo Muramatsu, Yasushi Yagi, Syunsuke Mori, Yuzuko Utsumi, Masakazu Iwamura, Koichi Kise

05. An Efficient Approach for Eigen-joints-based 3D Activity Recognition
Hao Xu, Chilwoo Lee

06. Multispectral Pedestrian Detection: Benchmark Dataset and Baselines
Soonmin Hwang, Jaesik Park, Namil Kim, Yukyung Choi, In So Kweon

07. Occlusion Handling in Gait Recognition via Feature Regeneration
Daigo Muramatsu, Yasushi Makihara, Yasushi Yagi

08. Layered Contextual Model For Face Alignment With Group Sparse Feature
Falong Shen, Jiyuan Zhang, Rui Gan, Jingdong Wang, Gang Zeng, Hongbin Zha

09. Which Gait Feature Is Effective for Impairment Estimation?
Chengju Zhou, Ikuhisa Mitsugami, and Yasushi Yagi

10. Structure from Small Motion for Rolling Shutter Cameras
Sunghoon Im, Hyowon Ha, Gyeongmin Choe, Hae-Gon Jeon, Kyungdon Joo, In So Kweon

11. Moving Target Learning and Following for Collaborative Vision-Equipped Drone
Moju Zhao, Koji Kawasaki, Kei Okada, Masayuki Inaba

12. Gaze Direction Classification and Eye Fixation Analysis of Children Based on Camera Classes
Tiejian Zhang, Songjiang Li, Jinshi Cui, Li Wang, Xia Li, Wen Cui, Hongbin Zha

13. Learning a Deep Convolutional Network for Light-Field Image Super-Resolution
Youngjin Yoon, Hae-Gon Jeon, Donggeun Yoo, Joon-Young Lee, In So Kweon

14. Depth Estimation and Video Completion by Motion Analysis for Outdoor Omni-directional View
Carlos Morales, Shintaro Ono, Yasuhide Okamoto, Menandro Roxas, Takeshi Oishi, Katsushi Ikeuchi

15. Ellipse-Specific Fitting by Relaxing the 3l Constraints Using Semidefinite Programming
Jiangpeng Rong, Sen Yang, Xiang Mei, Xianghua Ying, Shiyao Huang, Hongbin Zha

16. Large-Population Dual-Task Dataset for Physical/Mental Condition Estimation
Ikuhisa Mitsugami, Chengju Zhou, Fumio Okura, Masataka Niwa, Yasushi Yagi

17. Development and Practice for the VR/MR Guided-Tour System
Yoshihiro Sato, Takeshi Oishi, Katsushi Ikeuchi

18. An Automatic System for Drug Susceptibility Testing
Andrey Grushnikov, Kazuma Kikuchi, Takeo Kanade, Yoshimi Matsumoto, Kunihiko Nishino, Yasushi Yagi

Robust Registration with Multiple Panoramic Cameras for Mixed Reality on Moving Vehicle

Kosuke Fujimoto[1], Yasuhide Okamoto[2], Takeshi Oishi, and Katsushi Ikeuchi[1]

[1]*Computer Vision Lab, The University of Tokyo, Japan*

{fujimoto, okamoto, oishi, ki} @cvl.iis.u-tokyo.ac.jp

GPS and gyro sensors are often used for estimating pose and position of a moving vehicle for outdoor mixed reality applications. However, gyro sensors cause error accumulation on each frame, and GPS cannot get accurate locations.

Therefore we estimate those parameters by extracting correspondences between current images captured by panoramic cameras and rendered panoramic images of the pre-measured 3D model around the vehicle for mixed reality. We can get pose and position globally by utilizing pre-measured 3D model.

In addition, we demonstrate more robust registration by using multiple panoramic cameras installed on both sides of the vehicle.

Figure 1. Mixed Reality bus system with 2 panoramic cameras

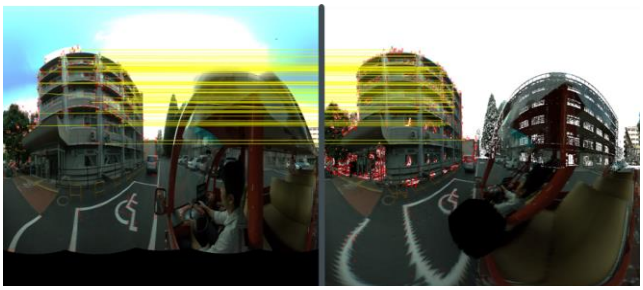


Figure 2. Matching between input image and rendered image from pre-measured 3D model

Figure 3. Average position estimation error between the methods which use only left camera, right camera and the one which combine both cameras (our method)

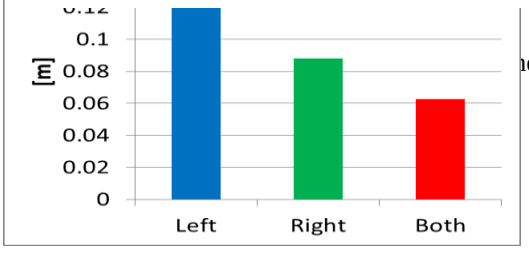
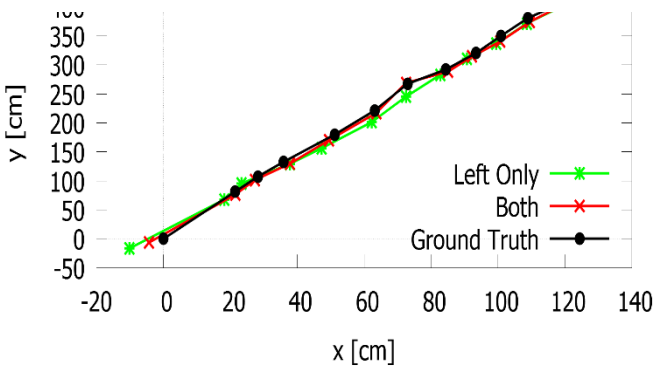


Figure 4. Estimated Paths between the method which uses only left camera, the method which combines both cameras (our method) and ground truth



Accurate Camera Calibration Robust to Defocus using a Smartphone

Hyowon Ha, Yunsu Bok, Kyungdon Joo, Jiyoung Jung, and In So Kweon
Robotics and Computer Vision Lab, KAIST, Korea
{hwha, ysbok, kdjoo, jyjung}@rcv.kaist.ac.kr, iskweon@kaist.ac.kr

We propose a novel camera calibration method for defocused images using a smartphone under the assumption that the defocus blur is modeled as a convolution of a sharp image with a Gaussian point spread function (PSF). In contrast to existing calibration approaches which require well-focused images, the proposed method achieves accurate camera calibration with severely defocused images. This robustness to defocus is due to the proposed set of unidirectional binary patterns, which simplifies 2D Gaussian deconvolution to a 1D Gaussian deconvolution problem with multiple observations. By capturing the set of patterns consecutively displayed on a smartphone, we formulate the feature extraction as a deconvolution problem to estimate feature point locations in sub-pixel accuracy and the blur kernel in each location. We also compensate the error in camera parameters due to refraction of the glass panel of the display device. We evaluate the performance of the proposed method on synthetic and real data. Even under severe defocus, our method shows accurate camera calibration result.

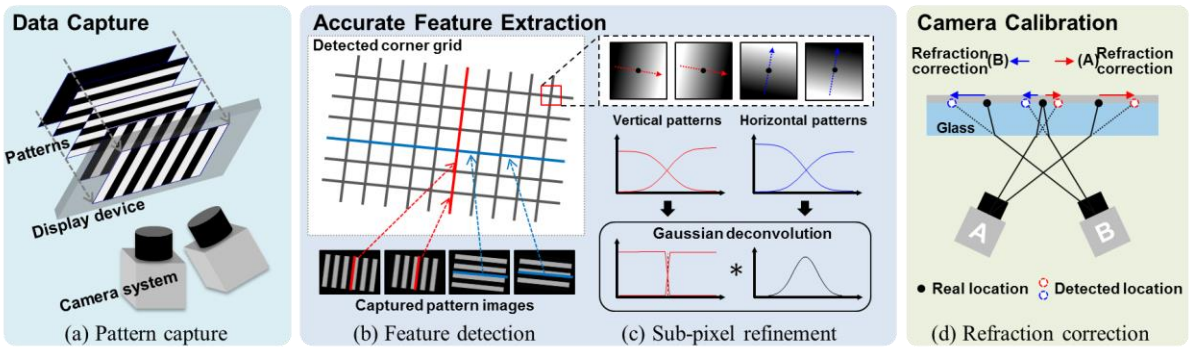


Fig. 1. The overall procedure of the proposed camera calibration using a display device.

Camera setup	Without refraction correction					With refraction correction				
	Harris corner	Geiger et al.	Placht et al.	Zero crossing	Proposed	Harris corner	Geiger et al.	Placht et al.	Zero crossing	Proposed
Left	1.1432	0.6447	0.6429	0.2125	0.1159	1.2457	0.6354	0.6341	0.1999	0.0896
Right	1.2343	0.7405	0.7401	0.231	0.1293	1.3758	0.7304	0.7300	0.2184	0.1079
Stereo	1.2178	0.7728	0.7735	0.3646	0.3406	1.1565	0.7014	0.7024	0.2272	0.1158

Table 1. Mean reprojection errors (pixel) from the calibration of a stereo camera system focused at infinity.

This research is supported by the National Research Foundation, Korea, under the NRF-ANR joint research programme (No. 2011-0031920).

Trajectory-based Stereo Visual Odometry with Statistical Outlier Rejection

Jiyuan Zhang, Rui Gan, Gang Zeng, Falong Shen, and Hongbin Zha
Key Laboratory of Machine Perception, Peking University, China
zhangjiyuan@water.pku.edu.cn, raygan@pku.edu.cn, g.zeng@pku.edu.cn, shenfalong@pku.edu.cn, zha@cis.pku.edu.cn

Visual odometry (VO) is a basic problem in computer vision, solving the ego-motion of camera only by captured images. It has been evaluated in benchmarks and lots of real world applications, e.g. robotics and autonomous driving. For the lack of global map, it suffers from drifting of accumulated error and irruptive error caused by other moving objects.

We present a stereo visual odometry algorithm with trajectorical information accumulated over time and consistency among multiple trajectories. The objective function considers transfer error of all previously observed points to reduce drifting, and can be efficiently approximated and optimized. We also exploit the linear system in non-linear optimization to evaluate the influence of each point for outlier rejection. Experiments with real world dataset show that both drifting and irruptive error are reduced by combining trajectorical information of multiple motions.

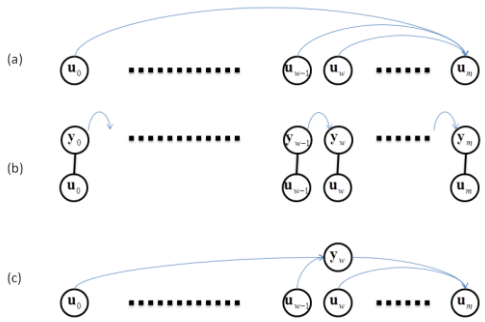
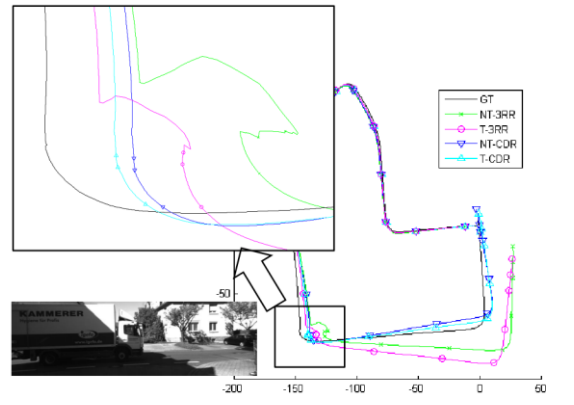


Fig. 9. (a): the original objective function over the whole trajectory, measured by transforming all former points to frame m. (b): the Markov chain in Kalman filter. Each observation has connection with its corresponding state. Only the last state is used in final estimation. (c): proposed method with inserted state at selected frame, which contains all former points. Thus fewer terms are included while whole trajectory is considered.



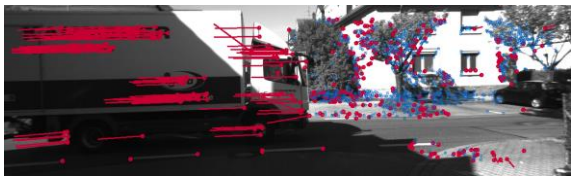


Fig. 10. Top: example of recovered path. Bottom: Outliers are effectively rejected by our method.

This work is supported by National Natural Science Foundation of China (NSFC) 61375022 and 61403005, and Microsoft Research Asia (MSRA) Re-search Grants.

Individuality-preserving Silhouette Extraction for Gait Recognition

Yasushi Makihara[1], Takuya Tanoue[1], Daigo Muramatsu[1], Yasushi Yagi[1],
Syunsuke Mori[2], Yuzuko Utsumi[2], Masakazu Iwamura[2], and Koichi Kise[2]
[1]*ISIR, Osaka Univ., Japan*
[2] *Graduate School of Engineering, Osaka Prefecture Univ., Japan*
{makihara, tanoue, muramatsu, yagi}@am.sanken.osaka-u.ac.jp, mori_s@m.cs.osakafu-u.ac.jp, {yuzuko, masa, kise}@ cs.osakafu-u.ac.jp

We propose a method of individuality-preserving silhouette extraction for gait recognition using standard gait models (SGMs) composed of clean silhouette sequences of a variety of training subjects as a shape prior. We firstly match the multiple SGMs to a background subtraction sequence of a test subject by dynamic programming and select the training subject whose SGM fit the test sequence the best. We then formulate our silhouette extraction problem in a well-established graph-cut segmentation framework while considering a balance between the observed test sequence and the matched SGM. More specifically, we define an energy function to be minimized by the following three terms: (1) a data term derived from the observed test sequence, (2) a smoothness term derived from spatio-temporally adjacent edges, and (3) a shape-prior term derived from the matched SGM. We demonstrate that the proposed method successfully extracts individuality-preserved silhouettes and improved gait recognition accuracy through experiments using 56 subjects.

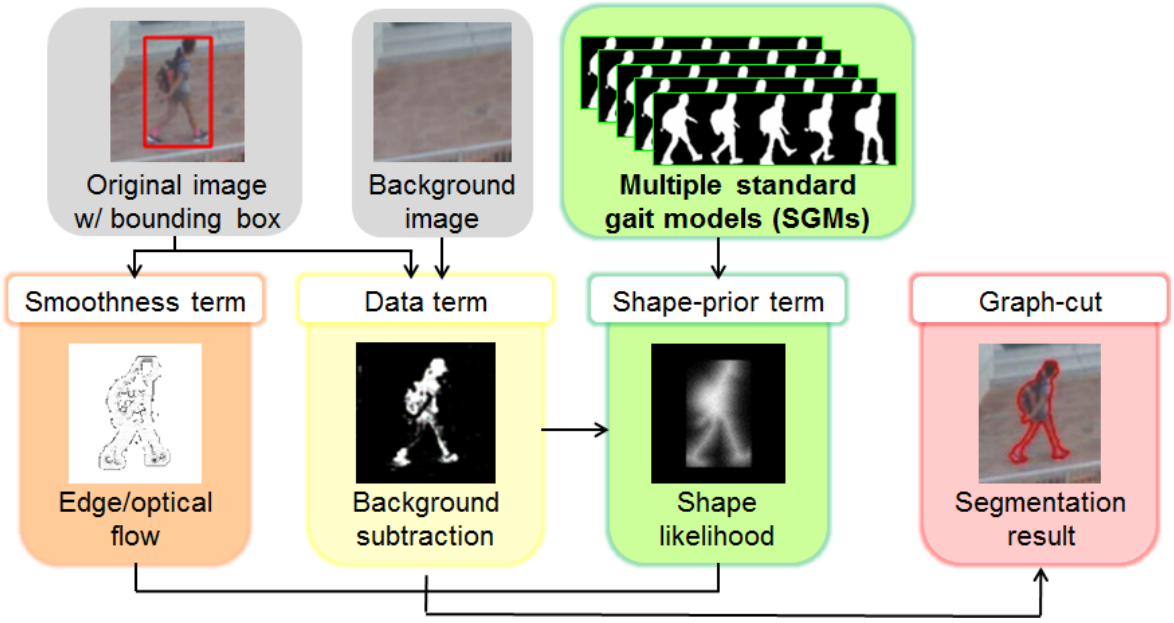


Fig. 11. Overview

This research was partly supported by a JSPS Grant-in-Aid for Scientific Research (A) 15H01693, ``R&D Program for Implementation of Anti-Crime and Anti-Terrorism Technologies for a Safe and Secure Society".

An Efficient Approach for Eigen-joints-based

3D Activity Recognition

Hao Xu and Chilwoo Lee
Department of Electronics and Computer Engineering, Chonnam National University, Korea
hs.xuhao@gmail.com, leecw@chonnam.ac.kr

Human activities can be represented by the movement locus of skeleton joints. In this paper, we present a variable approach for activity recognition by using 3D skeleton data obtained with a Kinect sensor. Primarily, in the preprocessing step, we use the simplified dynamic time wrapping method as the chronological order and Euclidean geometry as the spatial distance to select the probable candidates from the overall possible activities. In the processing step, there are three procedures in general. For feature abstraction, we define a modified activity feature descriptor using the interrelation of correlated joints in a frame for each activity. As to feature processing, we employ normalization to avoid non-uniformity in coordinates, and then Principal Component Analysis (PCA) is applied to deduce redundancy and decrease the dimensionality. As the result Eigen-joints for each activity are obtained. Finally with regard to classification, we classify the joints into multiple actions using Naïve-Bayes-Nearest-Neighbor (NBNN). The experimental result on benchmark dataset shows that the respective accuracy behaves well compared to state-of-the-arts. Besides, we implement a real-time activity recognition system to validate our method.

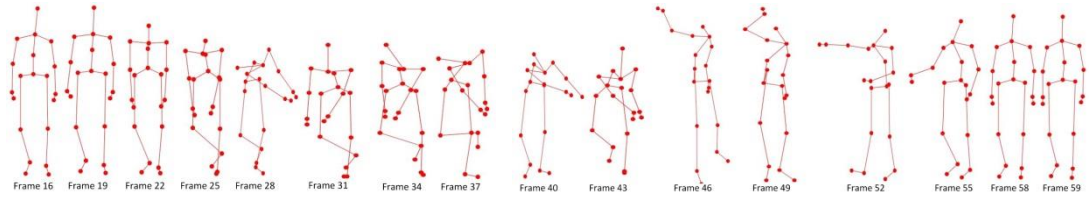


Fig.1. Skeleton joints of activity

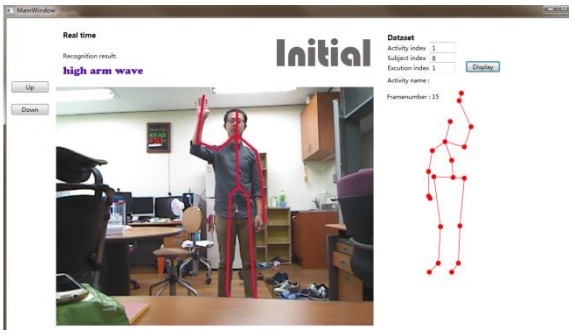


Fig.2. Real-time system

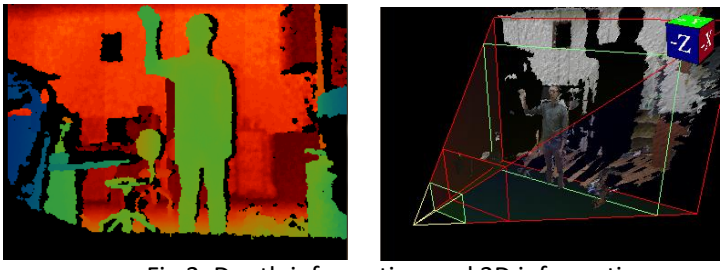


Fig.3. Depth information and 3D information

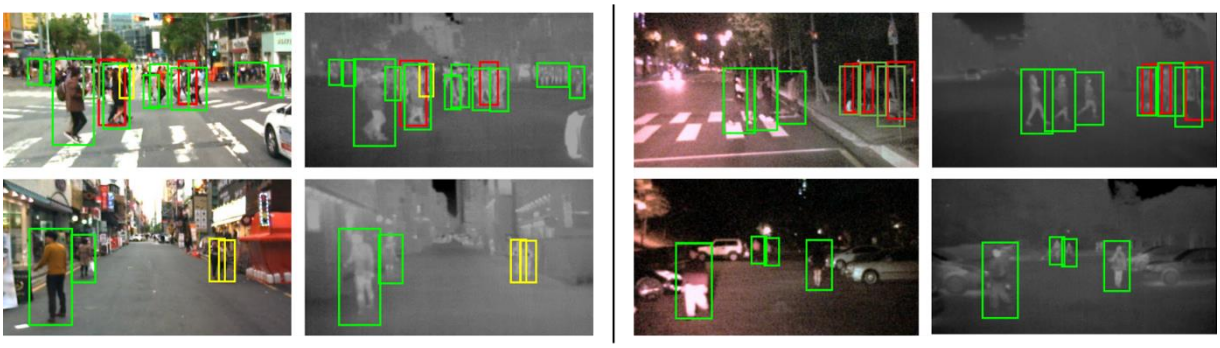
Multispectral Pedestrian Detection: Benchmark Dataset and Baseline

Soonmin Hwang, Jaesik Park, Namil Kim, Yukyung Choi and In So Kweon

Robotics and Computer Vision Lab, KAIST, Korea

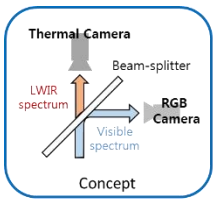
{smhwang, jspark, nikim, ykchoi}@rcv.kaist.ac.kr, iskweon@kaist.ac.kr

With the increasing interest in pedestrian detection, pedestrian datasets have also been the subject of research in the past decades. However, most existing datasets focus on a color channel, while a thermal channel is helpful for detection even in a dark environment. With this in mind, we propose a multispectral pedestrian dataset which provides well aligned color-thermal image pairs, captured by beam splitter-based special hardware. The color-thermal dataset is as large as previous color-based datasets and provides dense annotations including temporal correspondences. With this dataset, we introduce multispectral ACF, which is an extension of aggregated channel features (ACF) to simultaneously handle color-thermal image pairs. Multispectral ACF reduces the average miss rate of ACF by 15%, and achieves another breakthrough in the pedestrian detection task.

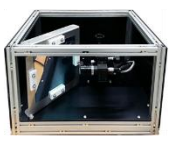


Examples of our multispectral datasets.

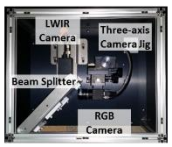
Thermal Color Blend color and thermal for visualization



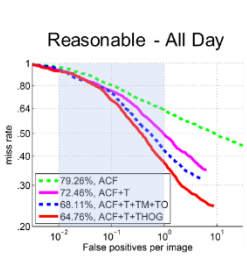
Concept

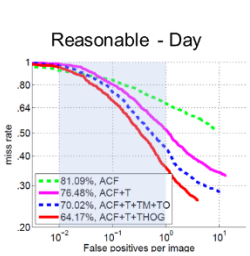


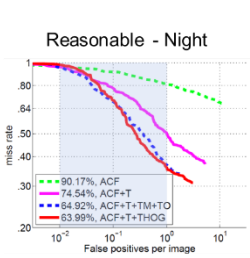
Frontal view



Top view







	ACF†	E-ACF	Δ
Reasonable-A	79.26	**64.76**	+14.50
Reasonable-D	81.09	**64.17**	+16.92
Reasonable-N	90.17	**63.99**	+26.18
Near-A	67.72	**50.09**	+17.63
Medium-A	84.08	**70.67**	+13.41
Partial Occ-A	91.59	**78.96**	+12.63

This research was supported by the MOTIE (The Ministry of Trade, Industry and Energy), Korea.

Occlusion Handling in Gait Recognition via Feature Regeneration

Daigo Muramatsu[1], Yasushi Makihara[1], and Yasushi Yagi[1]

[1] *Department of. Intelligent Media, ISIR, Osaka University, Japan*

{muramatsu, makihara, yagi}@am.sanken.osaka-u.ac.jp

Gait feature has potential to recognize subject in CCTV footages thanks to robustness against spatial resolution of the footages. In the CCTV footage, partial body-regions of subjects are, however, often occluded and un-observable, and therefore, an occlusion handling approach in gait recognition is necessary. The most popular approach for recognition from partially observed data is utilizing only the data from common observable region. This approach, however, cannot work in the cases where the matching pair has no common observable region. We therefore, propose an approach to enable recognition even from the pair with no common observable region. In the proposed approach, we consider to regenerate an entire gait feature from a partial gait feature and match the regenerated entire gait features for recognition. We focus on a subspace-based method and realize the regeneration. We evaluated the proposed approach against two different datasets. In the best case, the proposed approach achieves recognition accuracy with EER of 16.2% from such a pair.

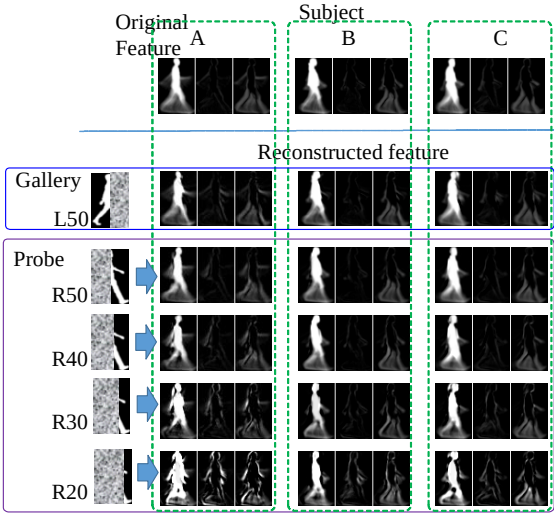


Fig. 1. Regenerated entire gait features from partial gait features

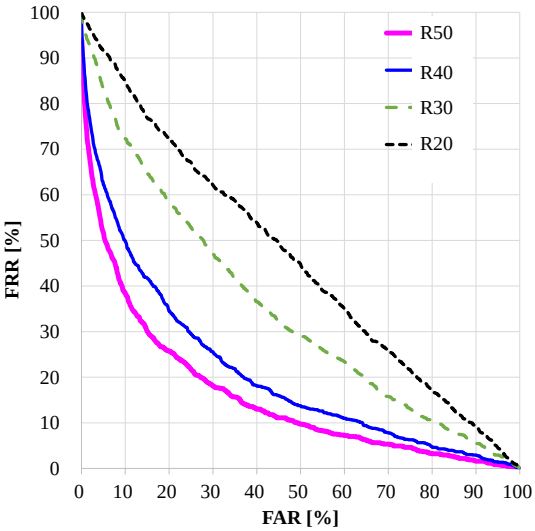


Fig. 2. ROC curves (Gallery mask: L50)

This research was supported by JSPS KAKENHI Grant Number 15K12037 and the JST CREST "Behavior Understanding based on Intention-Gait Model" project.

Layered Contextual Model For Face Alignment With Group Sparse Feature

Falong Shen[1], Jiyuan Zhang[1], Rui Gan[1], Jingdong Wang[1], Gang Zeng[1], Hongbin Zha[1]

[1]Key Laboratory on Machine Perception, Peking University

[2]Microsoft Research Asia

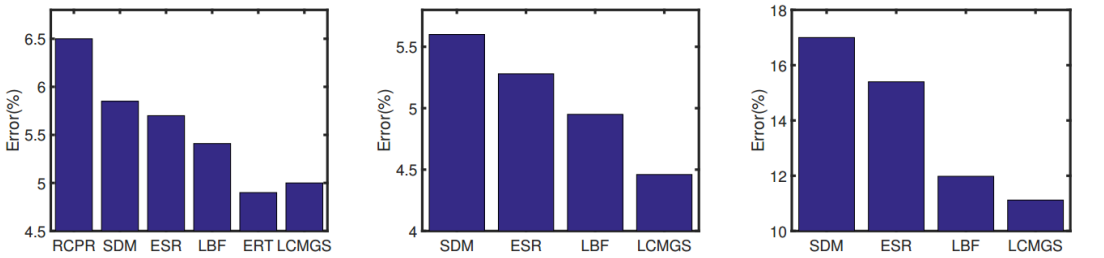
shenfalong@pku.edu.cn

In this paper we present a layered contextual model with group sparse feature (LCMGS) for face alignment. Layered contextual model has gained increasing amount of interests on face alignment these years. In each layer, features from all facial points are usually put together to form a feature pool for the learned function to capture rich contextual information. Previous methods usually feed the global feature to a linear regressor or a random fern. However, feature selection is very important in machine learning algorithm. Rather than choosing features for each landmark by hand, we propose a group sparse regression method to choose useful features for each landmark in each layer. The objective of group sparse regression is optimized under the accelerated proximal gradient (APG) framework. Experiments on Helen (194 points) and 300-W (68 points) benchmark datasets show that our model outperforms the state-of-the-art.

Our proposed layered contextual model with group sparse feature for one of the x-y coordinates of the $\lfloor (i+1)/2 \rfloor$-th landmark is reached through optimizing the following objective function for the (t + 1)-th layer,

$$\min_{\mathbf{w}} \frac{1}{2}||\Delta p_i - \mathbf{\Phi w}||_2^2 + \quad \frac{\mu}{2}||\mathbf{w}_1||_2^2 + \lambda||\mathbf{w}_2||_{2,1} + \frac{\alpha}{2}||\mathbf{w}_3||_2^2,$$

where $\Delta p_i = p^*{}_i - p^t{}_i$, $p^*{}_i$ is the ground truth pose and pt i is the input pose to this layer, all are N × 1 vectors where N is the number of training samples.



(a)Helen(194 points) (b)300-W common subset(68 points) (c) 300-W challenging subset(68 points)

Figure 1. Error on Helen dataset, 300-W common subset and 300-W challenging subset, normalized by inter-pupil distance. Results of RCPR and ERT originate from the corresponding author's paper. These methods are among the best works on face alignment. It shows LCMGS outperforms most of them by a large margin. The errors of LCMGS on this 3 test set are **5.00**%, **4.46**%, **11.12**%.

This work is supported by National Natural Science Foundation of China (NSFC) 61375022 and 61403005, and Microsoft Research Asia (MSRA) Research Grants.

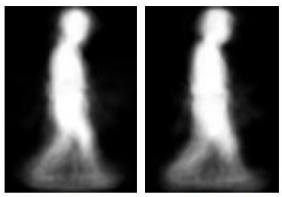
Which Gait Feature Is Effective for Impairment Estimation?

Chengju Zhou, Ikuhisa Mitsugami, and Yasushi Yagi
Yagi Laboratory, ISIR, Osaka University, Japan
{zhou, mitsugami, yagi}@am.sanken.osaka-u.ac.jp

People who have some impairments, such as a person whose leg is stiff, and a cataract patient has a weak sight view, show different walking styles comparing with people with no impairment. Indeed we can usually distinguish the differences quite easily just by observing their ways of walking. If we can realize a system that can automatically detect such impaired people from their walking styles, it could be very useful in many applications such as diagnosis of Parkinson's disease, rehabilitation of injured people, and monitoring physical condition of elderly.

To realize such a system that can automatically estimate impairment from gait observation, we have to consider which features should be extracted from the gait. Gait Energy Image (GEI) is often used since it show high performance for personal authentication. GEI, however, does not preserve temporal information of gait, i.e., a duration time of a walking period, and phase fluctuation, which may be effective for the impairment estimation but have not been GEI. We prepare two kinds of impairments (leg impairment and visual impairment) and normal walking. Experiment results confirm GEI is the most reasonable feature for impairment estimation from the viewpoint of accuracy and robustness.

normal vs. leg impairment	
normal	74.4%±2.2%
leg impairment	73.8%±2.3%



normal vs. visual impairment	
normal	82.0%±2.5%
visual impairment	82.0%±2.3%

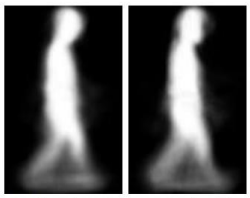


Fig.1. Classification accuracy of two impaired walking by GEI

Frame Order	1st	3rd	5th	7th	9th	11th	13th	15th	17th
Average accuracy Normal vs. visual impairment	79.9% ±4.1%	76.8% ±4.5%	80.2% ± 4.1%	76.2% ±4.1%	80.8% ±5.1%	77.6% ±4.3%	76.9% ±3.8%	72.4% ±4.7%	71.5% ± 4.9%
Average accuracy Normal vs. leg impairment	74.3% ±4.7%	73.5% ±5.4%	72.4% ± 5.1%	67.6% ±5.1%	74.8% ±6.2%	69.1% ±6.8%	63.1% ±5.3%	63.9% ±4.5%	70.1% ± 6.4%

Fig.2. Average classification accuracy of each frame of single GSV

This work was partly supported by the JST CREST, Japan.

Structure from Small Motion for Rolling Shutter Cameras

Sunghoon Im, Hyowon Ha, Gyeongmin Choe, Hae-Gon Jeon,
Kyungdon Joo and In So Kweon
Robotics and Computer Vision Lab, KAIST, Korea
{shim, hwha, gmchoe, hgjeon, kdjoo}@rcv.kaist.ac.kr, iskweon77@kaist.ac.kr

We present a practical 3D reconstruction method to obtain a high-quality dense depth map from narrow-baseline image sequences captured by commercial digital cameras, such as DSLRs or mobile phones. Depth estimation from small motion has gained interest as a means of various photographic editing, but important limitations present themselves in the form of depth uncertainty due to a narrow baseline and rolling shutter. To address these problems, we introduce a novel 3D reconstruction method from narrow-baseline image sequences that effectively handles the effects of a rolling shutter that occur from most of commercial digital cameras. Additionally, we present a depth propagation method to fill in the holes associated with the unknown pixels based on our novel geometric guidance model. Both qualitative and quantitative experimental results show that our new algorithm consistently generates better 3D depth maps than those by the state-of-the-art method.

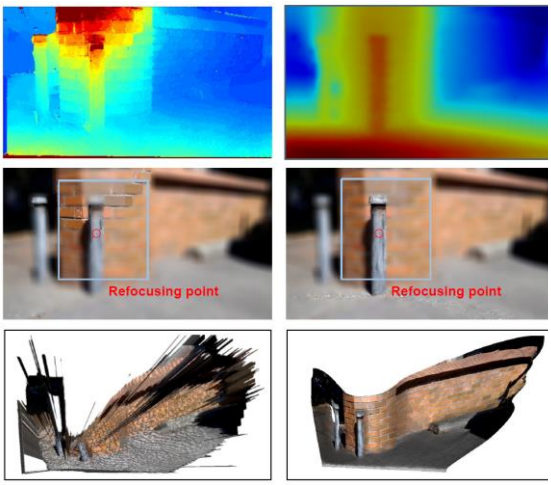


Fig. 13. (L) Yu and Gallup, (R) Proposed method

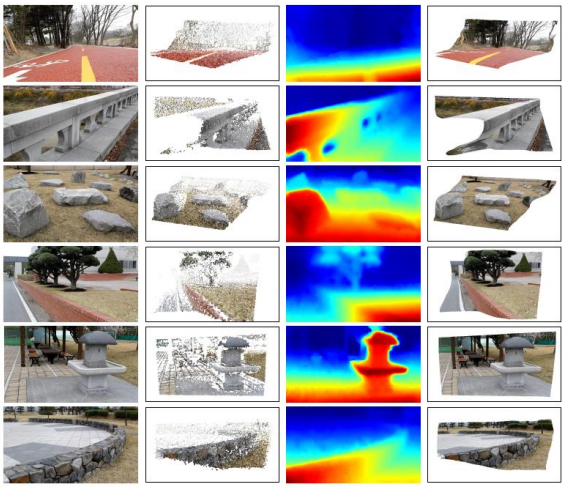


Fig. 12. Results from Proposed method

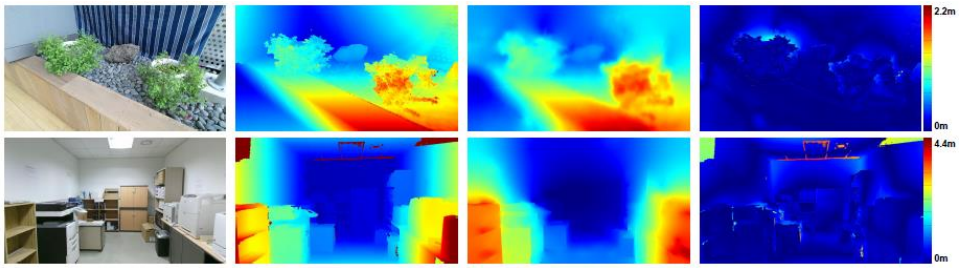


Fig. 3. Result Comparison between Kinect2 and proposed method

This work was supported by the National Research Foundation of Korea (NRF) grant funded by Korea government(MSIP) (No.2010- 0028680), and partially supported by the Study on Imaging Systems for the next generation cameras funded by the Samsung Electronics Co., Ltd (DMC RD center) (IO130806-00717-02). Hae-Gon Jeon was partially supported by Global PH.D Fellowship Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education (NRF-2015H1A2A1034617).

Moving Target Learning and Following for Collaborative Vision-Equipped Drone

Moju Zhao[1], Koji Kawasaki[1], Kei Okada[1] and Masayuki Inaba[1]

[1]*JSK Lab, The University of Tokyo, Japan*

{chou,kawasaki,k-okada,inaba}@jsk.t.u-tokyo.ac.jp

We present the approach for drone to learn and follow a moving target based on the onboard vision processing. As drone is expected to follow a moving target such as human or other types of robots while operating collaborative tasks, the purpose of our work is to build a vision-equipped drone platform to learn and follow the moving target.

As shown in Fig1, in the quad-rotor type drone platform, there are onboard sensors and controllers to achieve the standalone flight. The base micro-controller with IMU unit operates attitude control, while velocity/position control is achieved by processing the optical flow and sonar sensors on the processor[1]. With this drone platform, user teaches which object to follow based on the vision information, and visual tracking starts subsequently. As shown in Fig2, the CamShift algorithm is used to choose and track the certain object in our work. With the certain method of the visual tracking, we then achieve the robust following ability of drone by controlling the velocity of horizontal movement, altitude and yaw angle, based on the basic rule that tracking target should be always at the center of the image view of the onboard camera. In the experiment, we illustrate the ability of the drone to recognize and follow a single read ball.

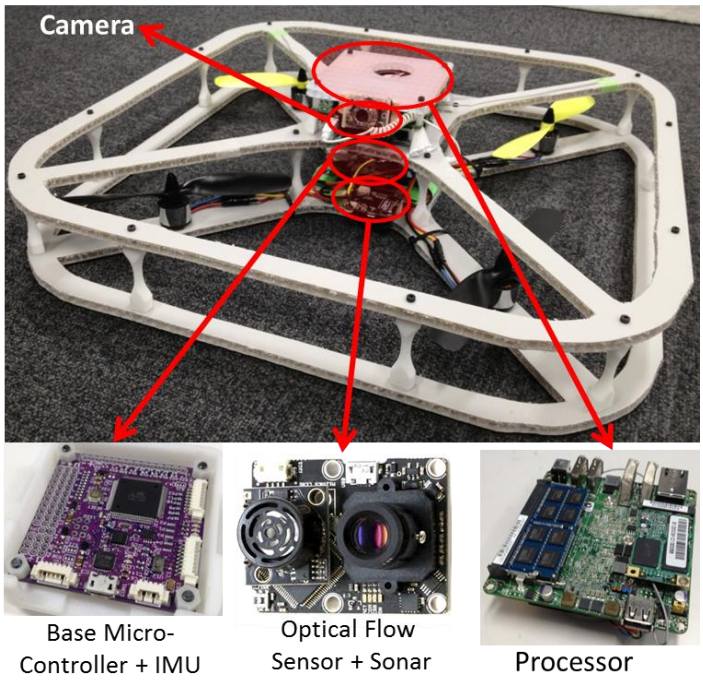


Fig. 1. Vision-Equipped Drone Platform

Fig. 2. Object following based on the

[1] Moju Zhao, Koji Kawasaki, Yohei Kakiuchi, Kei Okada, Masayuki Inaba. “Simultaneous Environment Modeling and Deployment of Network by Dropping Wireless Modules Based on Radio Field Intensity

Measurement Using an Micro Aerial Robot". Journal of the Robotics Society of Japan, Vol. 32, No. 7, pp. 643–650, 2014.

Gaze Direction Classification and Eye Fixation Analysis of Children Based on Camera Classes

Tiejian Zhang, Songjiang Li, Jinshi Cui, Li Wang, Xia Li, Wen Cui, and Hongbin Zha

Key Labotory of Machine Perception, Peking University, China

zhangtiejian@pku.edu.cn, lisongjiang@pku.edu.cn, cjs@cis.pku.edu.cn

Deficits in eye contact have been a hallmark of many behavioral disorders at early age like autism, which are cited widely as a diagnostic feature; however, current diagnosis methods are primarily based on long-term clinical observation, which are not sufficient enough for widespread use. Thus we present a way of gaze direction classification using camera glasses, which can be applied to automatic eye fixation analysis of children and further diagnosis.

In this paper, we extract multilevel Histograms of Oriented Gradients(HOG) of eye images, and combine it with facial landmarks which are highly relevant to head pose, then we apply this discriminative feature to Support Vector Machine(SVM) classifier to predict gaze direction. The test result on Columbia Gaze Dataset shows that our approach provides both accuracy of gaze estimation and robustness to head pose. Also, the analysis of more than 4000 first person video frames of naturalistic interactive scenes of multiple children demonstrates remarkable agreement between manual annotation and our approach. These results indicates that our approach is effective and efficient to automatical eye fixation analysis of children.

Head Pose	Accuracy
-15°	0.83(vertical)
0°	0.95(vertical)
15°	0.93(vertical)
-15°	0.79(horizontal)
0°	0.90(horizontal)
15°	0.89(horizontal)

Fig. 1 Results on Columbia Dataset

Child	Accuracy
1	0.9674
2	0.8591
3	0.9359
4	0.8854
5	0.8376
6	0.9256

Fig. 2 Results on First-Person Videos

This research was supported by the MOE (The Ministry of Education), China.

Learning a Deep Convolutional Network for Light-Field Image Super-Resolution

Youngjin Yoon, Hae-Gon Jeon, Donggeun Yoo, Joon-Young Lee, In so Kweon

Robotics and Computer Vision Lab, KAIST, Korea

yjoon@rcv.kaist.ac.kr, hgjeon@ rcv.kaist.ac.kr, hgjeon@ rcv.kaist.ac.kr, jylee@rcv.kaist.ac.kr

iskweon@kaist.ac.kr

Commercial Light-Field cameras provide spatial and angular information, but its limited resolution becomes an important problem in practical use. In this paper, we present a novel method for Light-Field image super-resolution (SR) via a deep convolutional neural network. Rather than the conventional optimization framework, we adopt a datadriven learning method to simultaneously up-sample the angular resolution as well as the spatial resolution of a Light-Field image. We first augment the spatial resolution of each sub-aperture image to enhance details by a spatial SR network. Then, novel views between the sub-aperture images are generated by an angular super-resolution network.

These networks are trained independently but finally finetuned via end-to-end training. The proposed method shows the state-of-the-art performance on HCI synthetic dataset, and is further evaluated by challenging real-world applications including refocusing and depth map estimation.

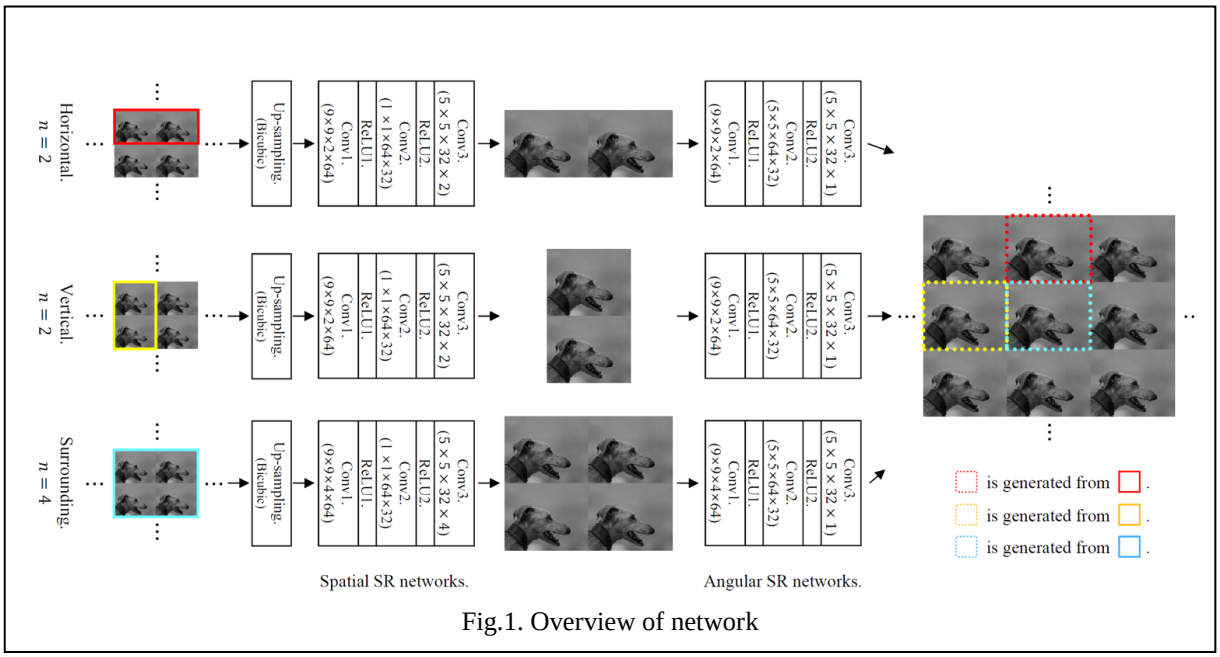


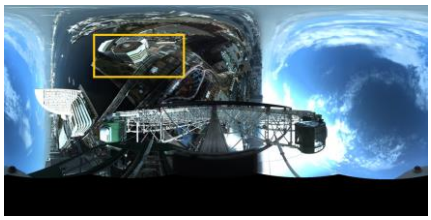
Fig.1. Overview of network

Depth Estimation and Video Completion by Motion Analysis for Outdoor Omni-directional View

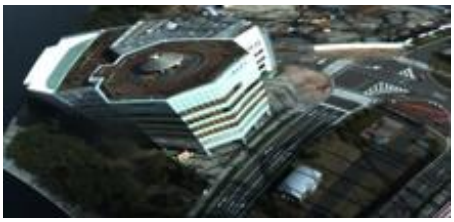
Carlos Morales[1], Shintaro Ono[1], Yasuhide Okamoto[1],
Menandro Roxas[1], Takeshi Oishi[1], and Katsushi Ikeuchi[2]
[1] *Computer Vision Lab, The University of Tokyo, Japan*
[2]*Microsoft Research Asia, China*
{carlos, onoshin, okamoto, roxas, oishi}@cvl.iis.u-tokyo.ac.jp, katsuike@microsoft.com

Depth estimation of relatively static scenes is an essential constituent in outdoor video completion applications with large space-time holes. In those applications the 3D geometry of the scene is needed since common optical-flow approaches are not enough to fill in the missing video portions that propagate along many frames. Using omni-directional view is suitable in such cases since it can provide information about the whole surrounding scene at every frame. However, generating consistent video outputs is still challenging.

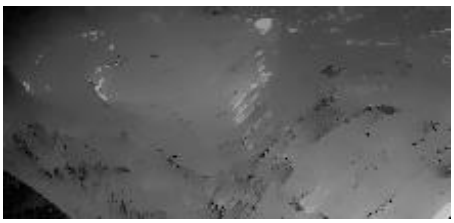
In this paper, we propose an approach for video completion using motion analysis. First the pixel motion along multiple omni-directional image frames is modeled based on the camera's motion (rotation and translation). Then the scene's depth is estimated from the motion of pixels constrained by the camera motion. Finally the estimated depth is used to propagate known color information of one key frame into the hole of the target frame that needs to be filled-in.



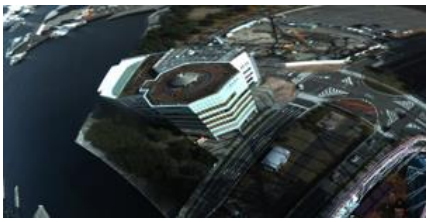
(a) Key frame with portion (yellow)



(b) Portion at the key frame

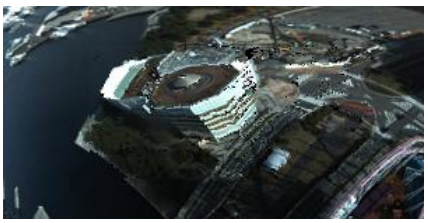


(c) Estimated depth map of the portion



(a) Ground truth at target frame

(b) Ground truth with artificial hole



(c) Our synthesized result

This work is partially supported by NET Project of MEXT, Japan.

Ellipse-Specific Fitting by Relaxing the 3l Constraints Using Semidefinite Programming

Jiangpeng Rong, Sen Yang, Xiang Mei, Xianghua Ying, Shiyao Huang, and Hongbin Zha

Key Laboratory of Machine Perception (Ministry of Education)

Center for Information Science, Peking University, P.R. China

{rjp, magicyang, meix}@pku.edu.cn, xhying@cis.pku.edu.cn, h41@pku.edu.cn, zha@cis.pku.edu.cn

The Fitting of geometric primitives to a given cloud of points is an essential task in computer vision and pattern recognition. One of the most commonly used primitives is the ellipse which, usually being the perspective projection of a circle, has been proved to be very important in numerous applications.

This paper presents a new efficient method to increase accuracy and robustness of ellipse fitting, by utilizing the 3L algorithm and semidefinite programming (SDP). The novelty lies on the combination of relaxed geometric distance constraints and semidefinite programming framework. Due to the relaxed 3L constraints, the proposed approach provides high robustness in the presence of noise. The accuracy of final solution is prominently increased even if the data suffer from strong occlusions or noises. The proposed method represents significant advantages in both accuracy and robustness. Experimental results and comparisons with state-of-the-art fitting methods demonstrate the improvements in ellipse fitting.

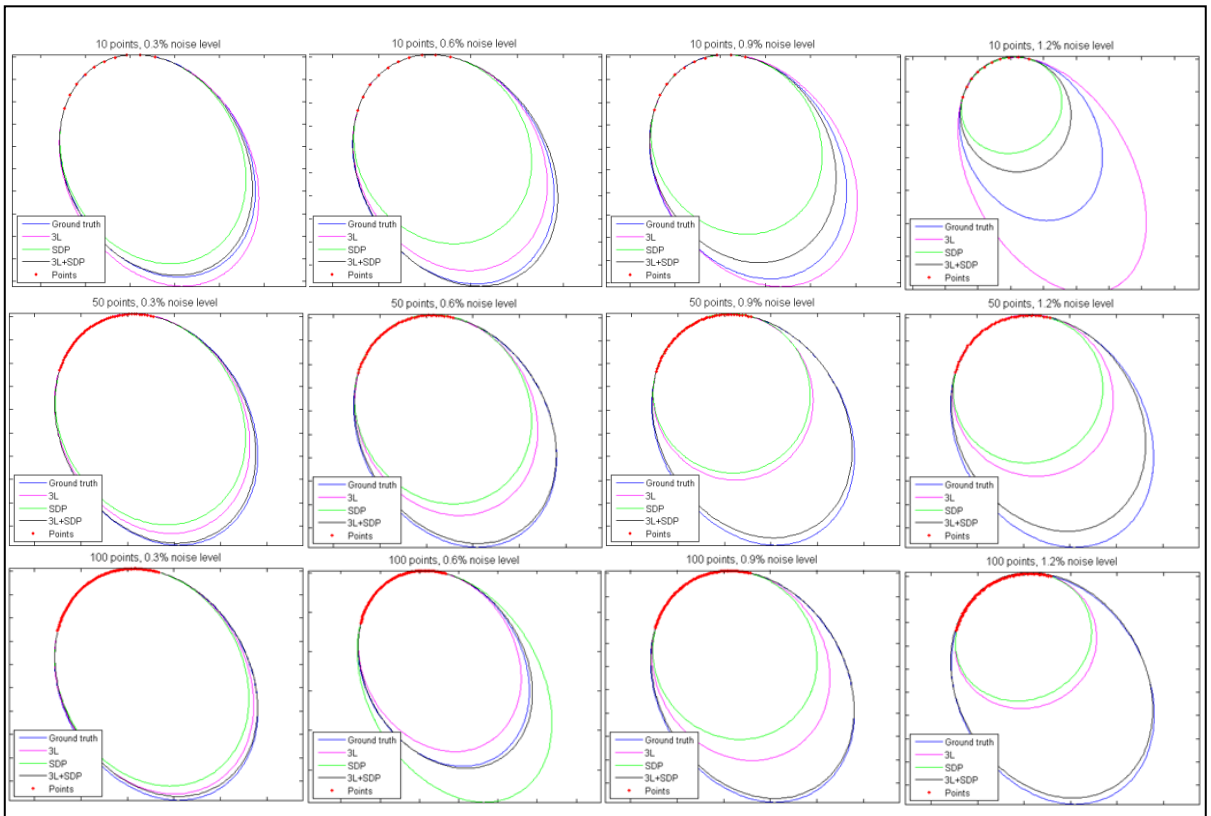


Fig.1 Comparisons between algorithms of ellipse fitting with respect to different noise levels (0.3%, 0.6%, 0.9%, and 1.2%) and different numbers of fitted points (10, 50, and 100).

Large-Population Dual-Task Dataset for Physical/Mental Condition Estimation

Ikuhisa Mitsugami, Chengju Zhou, Fumio Okura, Masataka Niwa, Yasushi Yagi,
ISIR, Osaka Univ., Japan
{mitsugami, zhou, okura, niwa, yagi}@am.sanken.osaka-u.ac.jp,

Dual-task (to simultaneously perform two different tasks, e.g., walking and arithmetic) is reported to be effective for diagnosis and rehabilitation of dementia. However, quantitative relation among performances of these two tasks and degree of dementia is not studies well. Motivated by this fact, we have developed a novel system for collecting dual-task performance observations. This system is located in Miraikan (The National Museum of Emerging Science and Innovation, Japan), which is a very popular museum and gets huge number of visitors every day. As the system is designed so as that every participant (from a child to elderly person) can enjoy it, we have obtained more than 30 thousands participants. In this presentation, we briefly introduce this system and show some primary statistical results.

Fig. 1 Miraikan

Fig. 2 Proposed system locatein in Miraikan

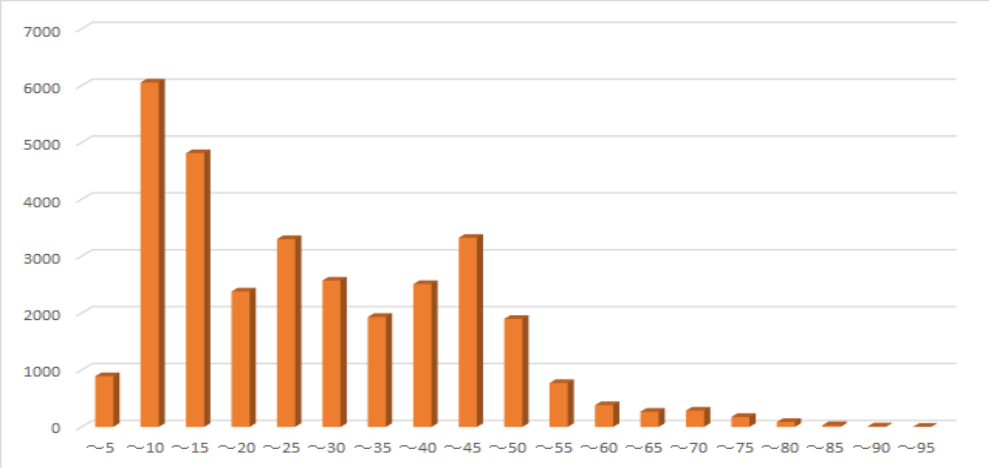


Fig. 3 Age distribution of participants

Development and Practice for the VR/MR Guided-Tour System

Yoshirhi Sato, Takeshi Oishi, and Katsuhi Ikeuchi

Institute of Industrial Science, The University of Tokyo, Japan

yoshi@cvl.iis.u-tokyo.ac.jp, oishi@cvl.iis.u-tokyo.ac.jp, ki@cvl.iis.u-tokyo.ac.jp

Nowadays, taking advantage of information science and technology for tourism is promoted by the local governments, and the feasibility of Virtual/Mixed Reality(VR/MR) technology applied tourism is well studied in the state-of-the-art work.

In this work, we propose a new portable VR/MR system to support the large−area contents reproducing the historical scenes. We applied this system into a guide tour in Asuka−kyo for the tourists and volunteers. Our system consists of three elements:

1） Portable VR/MR Terminal using HMD and Tablet−PC with the backbone system for MR(Fig.1),

2） Simple user interface software(Fig.2),

3） Historical restored objects for VR/MR Contents(Fig.3).

Our system is evaluated by questionnaires from the tourists and volunteers in a long term experience. As result, we obtained valuable comments and useful information.

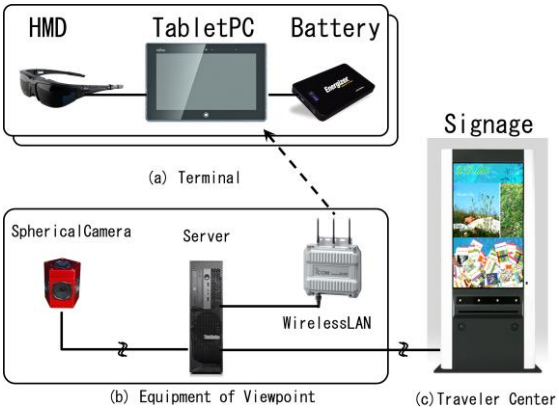


Fig. 16. System Overview

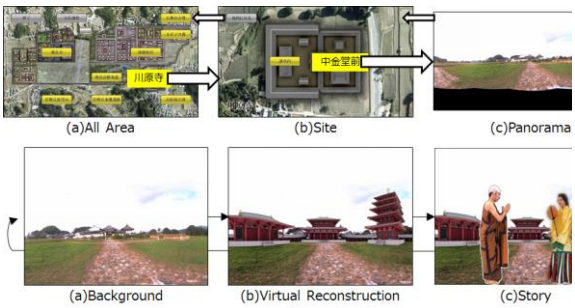


Fig. 15. Simple Interface

Fig. 3. Historical restored objects

An Automatic System for Drug Susceptibility Testing

Andrey Grushnikov[1], Kazuma Kikuchi[1], Takeo Kanade[1], Yoshimi Matsumoto[1],
Kunihiko Nishino[1] and Yasushi Yagi[1]
[1]*Institute of Scientific and Industrial Research, Osaka University*
andrey@am.sanken.osaka-u.ac.jp, kikuchi@am.sanken.osaka-u.ac.jp

The demand for a quick and robust method for drug susceptibility testing (DST) has risen, since a number of discovered bacteria strains resistant to modern antibiotics increased significantly. Traditional methods for DST rely on manual processing and observation of samples, thus are extremely slow.

A recent development of a microfluidic device for susceptibility testing (DSTM) allows test and examine multiple bacteria samples at once. The DSTM consists of several sets of microfluidic channels printed in polymer on a glass cover. Bacteria strain and a drug with predetermined concentration are injected in each sample channel. After several hours of growth, samples are observed and visual changes in cell morphology are studied to estimate resistance to a drug. This approach reduces the time needed to perform DST, however does not exempt from manual estimation of strain resistance to a drug.

The system we created automates the process of determining the degree of drug effect on each bacteria sample. An input for our system is a microscopy image of DSTM with 4 sample channels, one of which is a control one – no drug applied. The system locates each channel, performs cell detection and calculates various features: total number of cells, their length, width and others. The extracted characteristics are used for determining if the particular strain is resistant or sensitive to a drug with the pre-trained SVM.

For the proposed system we developed an application and tested it on a dataset, which contained 80 images of DSTM for 5 different drugs. The SVM output for strain sensitivity was compared with the ground truth data obtained via the minimum inhibitorary concentration (MIC) method. The tests showed efficiency of the created system.

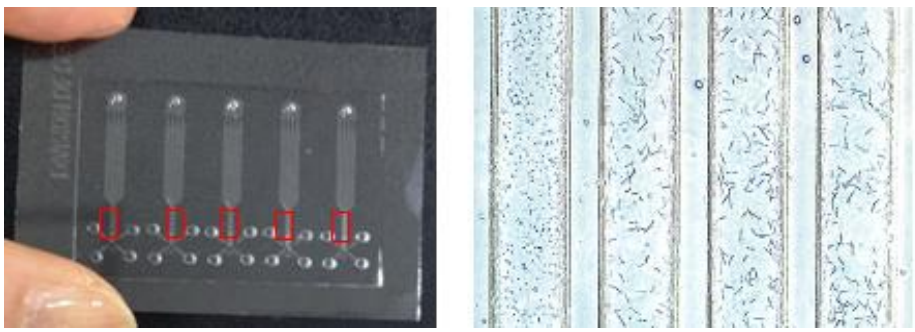


Fig. 1. DSTM and an input image with 4 sample channels. The first channel is the control one. Visible changes (reduction of total cell number and increase of length) can be observed in other 3 test samples.

ORAL SESSION V

Cluttered Object Separation through Active Robot Interactions
Krishneel Chaudhary, Chi Wun Au, Xiangyu Chen, Kotaro Nagahama, Hiroaki Yaguchi, Kei Okada, Masayuki Inaba

Reasoning based Vision Cognition Approach for Tomato Harvesting in Agriculture Automation
Xiangyu Chen, Krishneel Chaudhary, Hiroaki Yaguchi, Kei Okada, Masayuki Inaba

Off-Road Depth-Aided Visual Localization Using Mono Camera and LiDAR Sensor
Yufeng Yu, Huijing Zhao, Hongbin Zha

Ego-Centric Traffic Behavior Understanding through Multi-Level Vehicle Trajectory Analysis
Donghao Xu, Huijing Zhao, Hongbin Zha

Cluttered Object Separation through Active Robot Interactions

Krishneel Chaudhary[1], Chi Wun Au[1], Xiangyu Chen[1], Kotaro Nagahama[1],
Hiroaki Yaguchi[1], Kei Okada[1] and Masayuki Inaba[1]
[1]*JSK Lab, The University of Tokyo, Japan*
{krishneel, au, xychen, nagahama, h-yaguchi, k-okada, inaba}@jsk.t.u-tokyo.ac.jp

Robots operating in a domestic human centered environment, it should be able to adapt independently to the frequently changing structure of the environment in order to operate efficiently. The structure of the environment corresponds to the objects that are constantly moved by humans. Such objects of interest are mostly identified through visual recognition which requires that a pre-segmented model of the object is provided to the robot in offline. On the other hand if a robot is able to learn autonomously without human effort, it should be able to adapt to the changing environment with minimal or no human assistance. We show a method of separating a cluttered scene of unknown objects through robot interactions and manipulations. The aim of our method is to enable a domestic robot to build knowledge through fusion of vision and active interactions such that the robot can learn previously unseen objects. Our approach consists of two step of initial scene labeling which includes estimating a likelihood object candidates and using robot interactions to create high level object hypothesis. The approach is model-free and is able to operate on arbitrary objects.

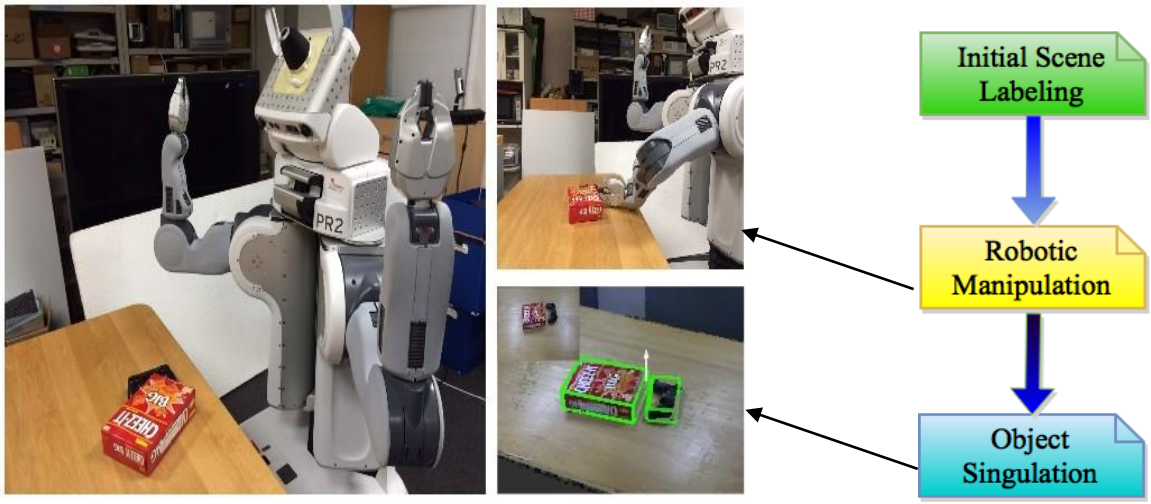


Fig. 1. Overview of the object segmentation (separation) method using robotic manipulation.

Scene	1	2	3	4	5	6	7	8
# of Objects	3	2	3	4	5	5	6	5
# of Pushes	4	3	3	5	5	5	6	5
# of Fail Pushes	0	1	0	2	1	1	2	0
# Failure	0	0	0	0	0	1	1	0
# of Singulation	3	2	3	3	4	4	4	4

Fig. 2. Illustration of preliminary segmentation results. The segmented object is shown using the 3D bounding box. The labels of the objects shown in color coding is merged after the object is segmented (separated).

Reasoning based Vision Cognition Approach for Tomato Harvesting in Agriculture Automation

Xiangyu Chen[1], Krishneel Chaudhary[1], Hiroaki Yaguchi[1], Kei Okada[1] and Masayuki Inaba[1]

[1]*JSK Lab, The University of Tokyo, Japan*

{xychen,krishneel,h-yaguchi,k-okada,inaba}@jsk.t.u-tokyo.ac.jp

We present a vision cognition framework for tomato harvesting based on vision geometrical and physical reasoning[1]. Inspired from the natural human harvesting behaviour, the purpose of our work is managing a humanoid robot to harvest tomatoes autonomously or with minimal human efforts.

Our idea is based on the observation of human harvesting behavior. First we equip our robot with similar picking tools to follow the same picking processes for specific crop. In the vision processing, we modeled crops, namely, tomatoes in one branch and then estimating the pedicel direction of each crop in that branch. Through pointcloud model segmentation, the primitive shape model of each crop can be obtained and we consider a simple fact that crops in one branch should remain stable with respect to gravity and interaction forces from neighboring crops in the branch, which human subconsciously considered. According to this assumption, a probabilistic model is created and the picking orders in the branch are assigned under the evaluated geometrical structure. In the experiments, hundreds of harvest tests were performed of tomato branches with respect to the successful harvesting rate.

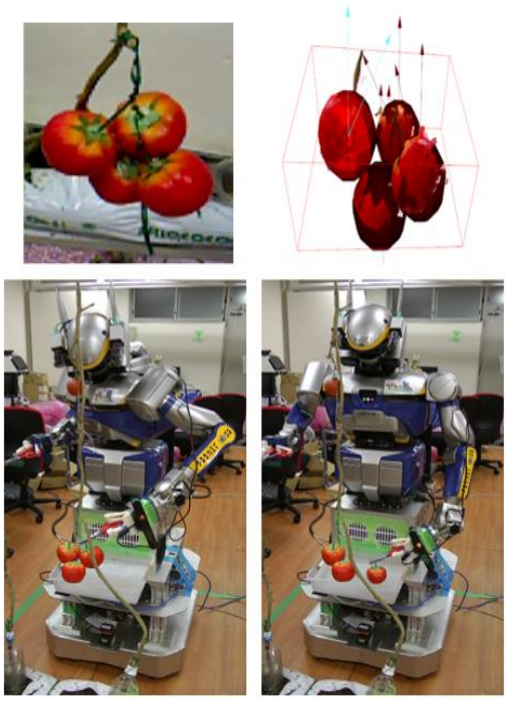


Fig. 1. Estimation and Harvesting

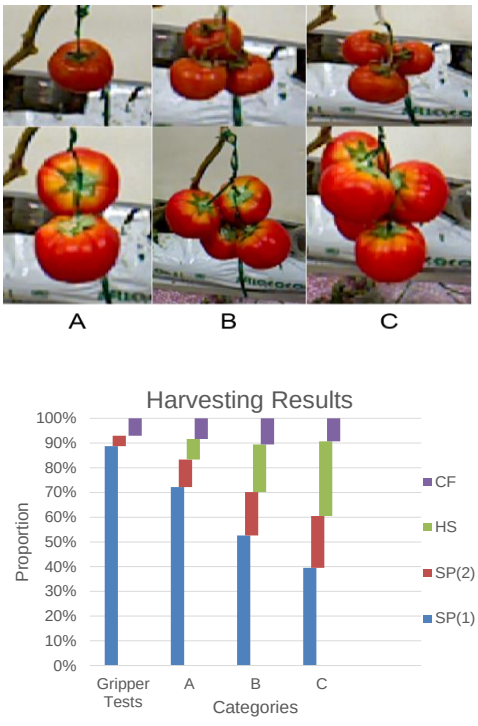


Fig. 17. Harvesting Results

[1] This work was published in IROS2015: *Xiangyu Chen, Chaudhary Krishneel, Tanaka Yoshimaru, Nagahama Kotaro, Yaguchi Hiroaki, Kei Okada, Masayuki Inaba: "Reasoning-Based Vision Recognition for Agricultural Humanoid Robot Toward Tomato Harvesting", in Proceedings of The 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp.6487-6494, 2015.*

Off-road depth-aided visual localization using mono camera and LiDAR sensor

Yufeng Yu, Huijing Zhao, Hongbin Zha

Key Laboratory of Machine Perception, Peking University, Beijing

yuyufeng@pku.edu.cn, zhaohj@cis.pku.edu.cn, zha@cis.pku.edu.cn,

Precise localization is an essential issue for autonomous driving applications, where GPS based systems are challenged to meet such requirement. Visual odometry or visual localization makes good performance in some cases. The video based approach requires RGB images and associated depth information, which is normally provided by RGB-D camera or stereo matching. However, in outdoor off-road cases, depth information can not be provided by RGB-D camera, and that from stereo matching is limited by the baseline.

In this paper, we propose a depth-aided visual localization method using mono camera and LiDAR sensor. The LiDAR on a rolling motor provides 3D point cloud, which is for depth information. Then the features on RGB image are classified into two parts. The core of this localization method is a bundle adjustment that refine the motion estimation as well as 3D point cloud. The confidence of depth information is calculated by the point cloud for robustness, which removes occlusion or miss matching. Finally, 3D point cloud matching is used for large scale localization refinement.

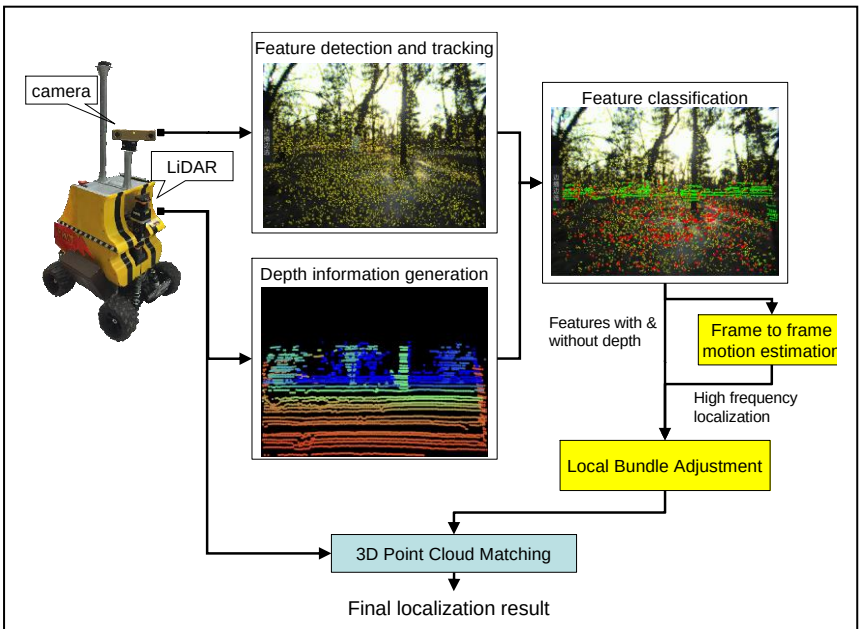


Fig. 18. System Overview. A 2D nodding laser scanner and a mono camera is used for this approach. Other devices are for validation.

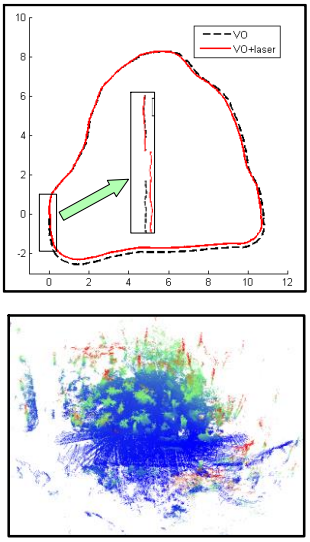


Fig. 2. Localization result and 3D point cloud

Ego-Centric Traffic Behavior Understanding through Multi-Level Vehicle Trajectory Analysis

Donghao Xu[1],Huijing Zhao[1], and Hongbin Zha[1]

[1]*Key Lab of Machine Perception (MOE), School of EECS, Peking University, China*

xudonghao@pku.edu.cn, zhaohj@cis.pku.edu.cn, zha@cis.pku.edu.cn

Understanding driving behavior of human drivers benefits the safety and comfort of driving assistance systems. A great deal of related information lies in the relative motion between the driving vehicles. In a set of trajectories which record the motion of environmental vehicles relative to the ego-vehicle, people can easily identify behaviors such as overtaking, lane change with overtaking, car following and so on. However, it's not straightforward to make computers understand the behaviors, because there is a gap between the trajectory data stored in the form of a series of 2-dimensional points and the semantic concept hold by humans.

To fill in the gap, we propose a multi-level trajectory analysis approach to perform ego-centric traffic behavior mining and understanding in an unsupervised way (Fig. 1). First, we adopt Sticky HDP-HMM to transform each trajectory into a series of discrete label. Then, we iteratively perform pairwise frequent subsequence finding on the trajectories' representation of the input level to output the behavior and trajectories' representation of the higher level. Experimental results show that various patterns of the traffic behavior can be discovered by the proposed approach (Fig. 2).

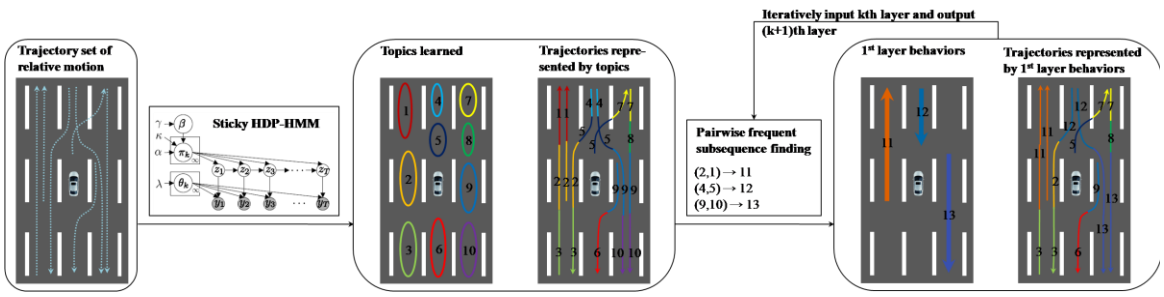


Fig. 1. The flow chart of the proposed multi-level trajectory analysis approach.

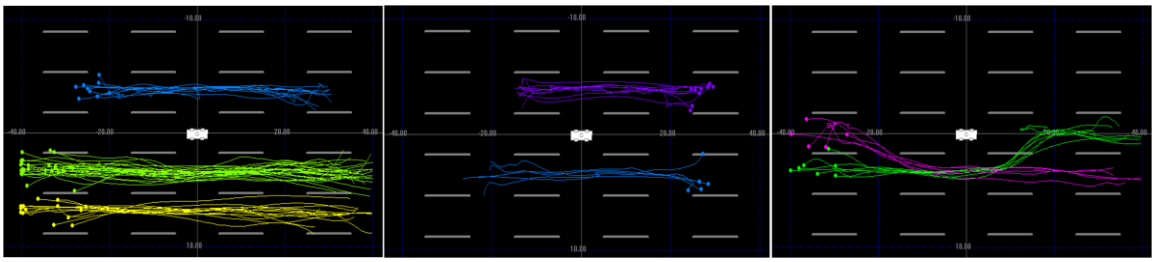


Fig. 2. The 4th layer's behaviors obtained by our approach. Left: the ego-vehicle overtake the surrounding vehicle. Middle: the surrounding vehicle overtake the ego-vehicle. Right: (purple) after overtaken by the ego-vehicle, the surrounding vehicle change its driving lane, and (green) the ego-vehicle change its driving lane to overtake the surrounding vehicle.