UMN
UNIVERSITAS MULTIMEDIA NUSANTARA

IWAIT 2020

International Workshop on Advanced Image Technology

January 5th - 7th, 2020
Yogyakarta City - **Indonesia**

CONFERENCE PROGRAMME AND ABSTRACT PROCEEDINGS

CONTENTS

Message from Advisory Committee

Masayuki NAKAJIMA (Emeritus Prof)
Uppsala University, Sweden /
Tokyo Institute of Technology, Japan

Greetings from IWAIT Conference Steering Committee Chair

It is my great pleasure to welcome all of you to IWAIT 2020 (International Workshop on Advanced Image Technology) in Indonesia, the 23rd series of IWAIT since the inaugural 1998 conference at Jeju Island, Korea.

Well, IWAIT 2020 is the second time held in Indonesia, currently in Yogyakarta City, the city of culture of Indonesia. The first one was in 2011 in Jakarta Capital City of Indonesia.

I remember that IWAIT 2011 was a very successful conference in IWAIT history, so I strongly hope that this second time conference in Yogyakarta will be an opportunity once again to exchange current information and establish friendship between researchers of the conferences again.

I appreciate the cooperation of all the attendees for the Papers and Posters presentation. Further I would like to express my sincere thanks to the conference Chair, Dr. Winarno and all the International Committee members of IWAIT 2020.

Message from IWAIT General Chair

Dr. Ir. P.M. Winarno,M.Kom.
Universitas Multimedia Nusantara
Jakarta, Indonesia

On behalf of the organizing committee, we are welcoming you to the 2020 International Workshop on Advanced Image Technology. We are very honoured that this International Conference is held in Yogyakarta – Indonesia from 5th to 7th January 2020 at the Santika Premiere Hotel Yogyakarta.

The participants of the IWAIT 2020 conference will have the opportunity to learn more about Yogyakarta City along with its stunning and beautiful culture and panorama. We will visit Borobudur and Prambanan Temples during the consequtive days of the conference.

We hope that all participants will enjoy all the programmes we create and become experiences that will be remembered for long.

And also if you stay at Santika Primiere Hotel of Yogayakarta, you will enjoy the local cuisines with the best recipes.

Happy conference at Yogyakarta, Indonesia...

Message from IWAIT Technical Programme Chair & Co-Chairs

The International Workshop on Advanced Image Technology (IWAIT) has become a well-established and well-known forum for sharing and discussing the latest research development in the broad field of image/ video and multimedia technologies.

IWAIT 2020, which features 2 keynote speakers, 2 invited speakers, and 132 papers, is organized in Yogyakarta City by Universitas Multimedia Nusantara, Indonesia.

We warmly welcome all keynote and invited speakers, authors and delegates to the Cultural City of Yogyakarta. This is a good opportunity not only for research exchange, but also for picturesque scenery and delicious local food.

More importantly, as an IWAIT signature, you will meet old friends, make new friends, and build up everlasting friendship.

We express our sincere appreciation to all the authors for their excellent works, the reviewers for their timely reviews, the keynote and invited speakers for their inspiring talks, and organizing committee members for their hard work in ensuring a smooth technical programme.

Hisaki Nate (Tokyo Polytechnic University, Japan)
Jae-Gon Kim (Korea Aerospace University, Republic of Korea)
Wen-Nung Lie (National Chung Cheng University, Taiwan)
Kemao Qian (Nanyang Technological University, Singapore)
Phooi Yee Lau (Universiti Tunku Abdul Rahman, Malaysia)
Lu Yu (Zhejiang University, China)
Alethea Suryadibrata (Universitas Multimedia Nusantara, Indonesia)

IWAIT 2020 COMMITTEE

ADVISORY COMMITTEE

Masayuki Nakajima (Uppsala University, Sweden)

Byeungwoo Jeon (Sungkyunkwan University, Republic of Korea)

Masayuki Tanimoto (Nagoya University, Japan)

Zen-Chung Shih (National Chiao Tung University, Taiwan)

Hock-Soon Seah (Nanyang Technological University, Singapore)

Kosin Chamnongthai (King Mongkut's University of Technology Thonburi, Thailand)

Hiroshi Fujita (Gifu University, Japan)

GENERAL CHAIR

P.M. Winarno (Universitas Multimedia Nusantara, Indonesia)

GENERAL CO-CHAIRS

Nobuhiko Mukai (Tokyo City University, Japan)

Yung-Lyul Lee (Sejong University, Republic of Korea)

Pau-Choo Chung (National Cheng Kung University, Taiwan)

Feng Lin (Nanyang Technological University, Singapore)

LOCAL ARRANGEMENT CHAIR

Rangga Winantyo (Universitas Multimedia Nusantara, Indonesia)

PROGRAMME CHAIRS

Hisaki Nate (Tokyo Polytechnic University, Japan)

Jae-Gon Kim (Korea Aerospace University, Republic of Korea)

Wen-Nung Lie (National Chung Cheng University, Taiwan)

Kemao Qian (Nanyang Technological University, Singapore)

Phooi Yee Lau (Universiti Tunku Abdul Rahman, Malaysia)

Lu Yu (Zhejiang University, China)

Alethea Suryadibrata (Universitas Multimedia Nusantara, Indonesia)

PUBLICATION CHAIR

M. Shobri (Universitas Multimedia Nusantara, Indonesia)

FINANCE CHAIR

Andrey Andoko (Universitas Multimedia Nusantara, Indonesia)

DECORATION & PERFORMANCE ART

Clemens Felix Setiyawan (Universitas Multimedia Nusantara, Indonesia)

Erwin Alfian (Universitas Multimedia Nusantara, Indonesia)

Itta Ernawati (Universitas Multimedia Nusantara, Indonesia)

SECRETARY

Andy Firmansyah (Universitas Multimedia Nusantara, Indonesia)

Dyah Ayu (Universitas Multimedia Nusantara, Indonesia)

DOCUMENTARY

Agus Syawal (Universitas Multimedia Nusantara, Indonesia)

Rizky Fitria (Universitas Multimedia Nusantara, Indonesia)

PROGRAM AND SESSIONS

Sunday, 5th January 2020

09.00-10.45	IWAIT2020 Registration Lobby : Hotel Santika Premiere
10.45-11.00	Welcome coffee break
11.00-16.00	Borobudur Temple Tour (Group I)
18.00-21.00	IWAIT2020 Welcome Dinner Bale Raos
21.00-21.30	Back to Santika Hotel

Monday, 6th January 2020

08.00-09.00	Opening Ceremony Room : Ballroom **Speeches from** **1. General Chair of IWAIT2020 : PM Winarno** **2. Guest of Honor**
09.00-10.20	Keynote Session Keynote 1 **Bayu Aria** Keynote 2 **Chiteuk Ahn, ETRI**
10.20-10.30	Coffee Break
10.30-12.00	**Papers Presentation : Parallel Session I**

Session 1A	Best Paper Session I
Room	Ballroom
Session Chair(s)	Jeon Byeongwoo, Wen-Nung Lie

11	**Fast Seam Carving for Video Images** Yuta Muraki, Kengo Kitamura, Koji Nishio, Takayuki Kanaya and Ken-Ichi Kobori
17	**Human Performance of Face Recognition Inferred from Characteristics of Observing Eye Movement Patterns Learned by Hidden Markov Model** Shun Oue, Ryoko Yamada and Shigeru Akamatsu
28	**GAN based Single-image Reflection Removal using Depth of Field Guidance** Yoonsik Choe and Miran Heo
90	**Handwriting Feature Extraction Method for Writer Verification Independent of Charactor Type by using AdaBN and AdaIN**

Kimiya Murase, Shunsuke Nakatsuka, Mariko Hosoe and Kunihito Kato

93 **Automated Classification Method of Lung Tumor Type using Cytological Image and Clinical Record**
Ayumi Yamada, Atsushi Teramoto, Yuka Kiriyama, Tetsuya Tsukamoto, Kazuyoshi Imaizumi, Masato Hoshi, Kuniaki Saito and Hiroshi Fujita

95 **Elastic Net with Adaptive Weight for Image Denoising**
Jun Xiao, Rui Zhao and Kin-Man Lam

114 **Deep Neural Network for Joint Light Field Deblurring and Super-Resolution**
Jonathan Samuel Lumentut and In Kyu Park

115 **Effective Binarization for Historically Degraded As-built Drawing Maps Using Convolutional Neural Networks**
Kuo-Liang Chung, De-Wei Hsieh and Chi-Huang Liao

Session 1B	Image Processing I
Room	Sida Mukti at Mezanin Floor
Session Chair(s)	Kim Jae Gon

5 **Sub-window Median-like Filter in Constant Time**
Daiki Isono, Xiaohua Zhang, Wenbo Jiang and Yuelan Xin

14 **A Method of Automatic Cage Generation by Variational Remeshing Method**
Takayuki Kanaya, Naoyuki Awano, Yuta Muraki and Kenichi Kobori

23 **Noise reduction in direct multi-material decomposition for dual-energy CT**
Haenghwa Lee, Dong-Hoon Lee, Pil-Hyun Jeon and Hee-Joung Kim

46 **An approach to metric rectification using sequential estimation of spatial adjacent planes appeared in single view image**
Haruka Kurose and Akio Kimura

58 **Texture Reconstruction Based on Underlying Pattern Modification that Reflects User's Intention**
Mami Nagoya, Tomoaki Kimura and Hiroyuki Tsuji

97 **Temporal and Contextual Aggregation for Road Making Semantic Segmentation**
Hiroaki Aizawa, Yuto Ura and Kunihito Kato

Session 1C	Animation and Virtual Reality I
Room	Griya Kresna at Mezanin Floor
Session Chair(s)	Seah Hock-Soon

20 **Automatic Classification of Manga Characters using Density-Based Clustering**
Hideaki Yanagisawa, Kengo Kyogoku, Jain Ravi and Hiroshi Watanabe

38 **Articulation awareness with 3D tongue using VR system**
Sayoko Takano and M.J. Hirayama

39 **Development of An Interface Simulating Swimming Motion for Virtual Space Walk-Through**
Kenta Uchida, Keito Tanizaki, Tomoaki Moriya, Tokiichiro Takahashi

45 **A Study of AR Advertising to Estimate the Volume of Invisible Contents in a Packaged Product**
Natusumi Motoura and Mie Sato

54 **A MR-based Visualization System of IoT Security**
Yuma Akasaka and Mitsunori Makino

59 **A Tuna dismantling education system in Virtual Reality**
Hikaru Yokoyama and Mitsunori Makino

Session 1D	Artificial Intelligence and Interdisciplinary Research I
Room	Sekar Jagad at 2nd Floor
Session Chair(s)	Lee Yung-Lyul

	86	**Visual Analysis of Fish Feeding Intensity for Smart Feeding in Aquaculture Using Deep Learning** Jui-yuan Su, Pei-Hua Zhang, Sin-Yi Cai, Shyi-Chyi Cheng and Chin-Chun Chang
	91	**Fast and Effective Object-Aware Domain Enhancement and Adaptation for Semantic Segmentation** Kuo-Liang Chung, Ya-Yun Cheng, Arie Tando and Don-Kai Chiang
	104	**Radiomic feature-based prediction model of lung cancer recurrence in NSCLC patients** Soomin Lee, Julip Jung, Helen Hong and Bongseok Kim
	106	**Super-resolution Image Generation for Improvement of Orbital Thin Bone Segmentation** Hee Rim Yun, Min Jin Lee, Helen Hong and Kyu Won Shim
	107	**Identifying Blurry Car License Number Plate Using Machine Learning** Erika Sawada, Ayaka Fujima, Seiichi Gohshi, Naiwala P. Chandrasiri
12.00-13.00	Lunch	
13.00-14.30	**Papers Presentation : Parallel Session II**	

Session 2A	Best Paper Session II
Room	Ballroom
Session Chair(s)	Masayuki Tanimoto, Kosin Chamnongthai, Nobuhiko Mukai

22	**Fast Total-Variation-based JPEG Artifact Removal via the Accelerated ADMM** Chihiro Tsutake, Ryoya Yamada, Ryoga Yamazaki and Toshiyuki Yoshida
24	**Transform selection for video coding** Nam-Uk Kim and Yung-Lyul Lee
33	**Automated detection of fundic gland polyps and hyperplastic polyps from endoscopic images using SSD**

Session 2B	Poster
Room	Foyer
Session Chair(s)	Masayuki Tanimoto, Kosin Chamnongthai, Nobuhiko Mukai

2 **Photo Spot Recommendation for Theme Park Visitors Using Collage Images**
Hayato Tsukushi, Tomohiro Fukuhara, Koichi Yamada, Hironobu Abe and Hidetaka Masuda

3 **System for Searching Illustrations of Anime Characters Focusing on Degrees of Character Attributes**
Yuta Koyama, Tomohiro Fukuhara, Koichi Yamada, Hironobu Abe and Hidetaka Masuda

4 **Wappen: Annotation System using Scene Matching with Multiple Terminals**
Kouhei Maeda, Tomohiro Fukuhara, Koichi Yamada, Hironobu Abe and Hidetaka Masuda

15 **Warping-based motion compensaion for triangular patches**
Keisuke Kamiya and Toshiyuki Yoshida

37 **A Study on Thermal Image Generation Based on Deep Learning and Abnormal Temperature Detection**
Ziyun Zhang and Makoto Hasegawa

44 **A Method for Enhancing Playground Equipment Experience using VR Technology with Smartphones**
Yoichi Shinchi and Shinji Mizuno

61 **Detection and Motion Analysis of Knee Joint in 3-D Point Cloud Data Measured by Depth Camera**
Wataru Morozumi, Norio Tagawa, Kan Okubo and Hiromichi Fujie

70 **Efficient Bin Allocation for Chroma Intra Mode Coding**
Jeeyoon Park and Byeungwoo Jeon

73 **Proposal of a Rescue Operation Support System based on 3D Reconstruction, GPS, and Digital Pen**
Wei-Chung Chen, Shuuichi Ihara and Makoto Hasegawa

84 **Pose Estimation of Excavators**
Hinako Nakamura, Yumeno Tsukada, Toru Tamaki, Bisser Raytchev and Kazufumi Kaneda

96 **Operator overloading for cv::UMat converted to equivalent function calls at compile time**
Yusuke Kameda, Yoshihiro Maeda and Takayuki Hamamoto

111 **PSF Optimization for Motion Deblurring using Temporally Coded Light**
Jaelin Lee, Gyoheak Oh and Byeungwoo Jeon

Session 2C	Special Session Medical Imaging I
Room	Sida Mukti at Mezanin Floor
Session Chair(s)	Hiroshi Fujita

123 **Breast Ultrasound Computer-aided Diagnosis Using Deep Learning**
Ruey-Feng Chang

120 **Multimodality breast mass classification using CNN-based similarity estimation**
Chisako Muramatsu, Mikinao Oiwa, Takako Morita, Tomonori Kawasaki and Hirohi Fujita

121 **Automated Retinopathy Detection Based on Convolutional Neural Network on Retinal Images**
Yuji Hatanaka, Phooi Yee Lau, Paulo Correia, Paulo Fonseca, Aida Campos

128 **A Novel Tracking Scheme for Norway Lobster and Knee Replacement Surgery Phase Recognition with Wearable Camera**
Syoji Kobashi, Shoichi Nishio, Belayat Hossain

79 **Super Resolution for 8K Endoscope**
Seiichi Gohshi, Kenkichi Tanioka, Hiromasa Yamashita

14.20-14.30 Coffee Break

14.30-16.00 **Papers Presentation : Parallel Session III**

Session 3A	Best Paper Session III
Room	Ballroom
Session Chair(s)	Masayuki Nakajima, Lam Kin Man

13 **Detecting Image Frames which Contain a Moving Object from a Severely Distorted Video Stream Using Dynamic Mode Decomposition**
Daisuke Uchida, Atsushi Yamashita and Hajime Asama

30 **Environment Understanding During Walking via Modality Conversion from Visual to Haptic Information**
Sinichiro Morita, Yoshihiro Kanehara, Kazuho Kito, Harumi Kobayashi, Naoki Mukawa and Akio Nakamura

32 **ElectroMagnetic Guitar: Chord Playing Support System on Guitar by Electromagnets**
Nozomu Yoshida and Katsutsugu Matsuyama

56 **Improvement of Accuracy of Wide-surrounded Multi-projection in Indoor Spaces**
Kentaro Miura and Naoki Hashimoto

67 **Moving Obstacle Tracking and Estimation on Crosswalk for Blind-People Navigation System**
Kongpob Maksap and Kosin Chamnongthai

74 **Privacy-Preserving Machine Learning Using EtC Images**
Ayana Kawamura, Yuma Kinoshita and Hitoshi Kiya

82 **Integrated and Scalable Augmented-Reality Multiplayer Robotic Platform**
Budianto Tandianus, Hock Soon Seah, Li Wang and Chee Kwang Quah

99 **Indonesian Culture Recognition Portal based on Crowd Sourcing Contents**
Winarno Darmoyuwono, Rudy Pramono and Wella Wella

Session 3B	Optimization
Room	Sida Mukti at Mezanin Floor
Session Chair(s)	Qian Kemao, Haekwang Kim

52 **Examination of Group Head Angle Acceleration Analysis Method for Learning Evaluation in Outdoor Education**
Kazunari Haruta, Takayuki Nakata, Ryo Sakurai, Kazuhiko W. Nakamura, and Kousaku Hashimoto

60 **Pitching Form Evaluation Based on Elbow Position by a Monocular Camera**
Nohara Naoto, Shishido Hidehiko, Kitahara Itaru, Kameda Yoshinari

63 **Performance Analysis on Prediction Structure for Multi-view-based Light Field Video Coding**
Thuc Nguyen Huu, Vinh Van Duong and Byeungwoo Jeon

65 **Acquisition of Wiping Area Using SLAM for Visualization of Cleaning Area**
Wataru Yoneda, Yoshitsugu Manabe and Noriko Yata

71 **Evaluation for Harmonic Location Estimation System of Image Retrieval and SLAM**
Yamasaki Kohei, Shishido Hidehiko, Kitahara Itaru, Kameda Yoshinari

75 **HEVC intra prediction mode classification by deep learning**
Hanxiang Wang, Yanfen Li, L. Minh Dang, Hae Kwang Kim

Session 3C	Multimedia and Systems
Room	Griya Kresna at Mezanin Floor
Session Chair(s)	Hiroshi Watanabe

6 **An Estimation Method of the Camera Fluctuation for a Video-based Vibration Measurement**
Kazuyuki Miura, Takehiro Tsuruta and Atsushi Osa

	21	**Surface Segmentation on 3D Point Cloud of Unbroken Earthenware and Its Applications** Keita Hayasaka, Tsutomu Kinoshita and Kouichi Konno
	27	**Background Subtraction via Exact Solution of Bayesian L1-norm Tensor Decomposition** Taehyeon Kim and Yoonsik Choe
	29	**Generating 3D Model of Furniture from 3D Point Cloud of Room** Shunta Osakama, Yoshitsugu Manabe and Noriko Yata
	43	**A Selective Fusion Module for Video Super Resolution with Recurrent Architecture** Zichen Gong, Toshiya Hori, Hiroshi Watanabe, Tomohiro Ikai, Takeshi Chujoh, Eiichi Sasaki and Norio Ito

Session 3D	Image Processing II
Room	Sekar Jagad at 2nd floor
Session Chair(s)	Pao-Choo Chung

	18	**A Method of Finding Characteristic Ocean-Satellite-Image Groups using Autoencoder** Kenta Maruyama and Katsutsugu Matsuyama
	25	**Extraction of Distinctive Keywords and Articles from Untranscribed Historical Newspaper Images** Sora Ito and Kengo Terasawa
	66	**Head orientation detection with small camera in outdoor education using background images** Kosaku Hashimoto, Kazunari Haruta, Kantaro Tabiraki, Kazuhiko W. Nakamura, Ryo Sakurai and Takayuki Nakata
	69	**Control Method of Pseudo-force Intensity by Voltage Change using Software Signal** Tsukasa Asano, Takayuki Nakata and Yusuke Kazashi
	126	**Lymphocyte Detection in Liver Histopathological Image** Chao-Ting Li, Pau-Choo Chung, Kuo-Sheng Cheng, Hung-Wen Tsai, Nan-Haw Chow
16.00 - end	Prambanan Temple Tour Banquet Purawisata (Ramayana Ballet)	

08.30-10.00 **Papers Presentation : Parallel Session IV**

Session 4A	Computer Graphics and 3D
Room	Ballroom
Session Chair(s)	Mie Sato, Sayoko Takano

9 **A Study on 3D Modeling from Handwritten Maps for VR Environment Construction of Historical Town**
Ryunosuke Kumagai, Morimichi Furudate and Kouichi Konno

12 **An Examination on Shape Feature Extraction Based on the Elliptic Approximation for Spatial Arrangement of Earthenware Pieces by Using 3D Measured Point Clouds**
Kazuto Yoshikawa, Tsutomu Kinoshita and Kouichi Konno

16 **Speckle Based Pose Estimation for 3D Measurement of the Feature-less Environment by Two Cameras**
Hiroshi Higuchi, Hiromitsu Fujii, Atsushi Taniguchi, Masahiro Watanabe, Atsushi Yamashita and Hajime Asama

34 **Tree Growth Model for Simulation of Appearance Change**
Maya Takatsuka, Yoshitsugu Manabe and Noriko Yata

35 **Avatar's Facial Expression with "Manpu (Comic Symbols)" by Using Multiple Information**
Shu Gemba and Tokiichiro Takahashi

77 **Fringe Projection Profilometry: an Overview**
Xiaoyu He and Qian Kemao

Session 4B	Artificial Intelligence and Interdisciplinary Research II
Room	Sida Mukti at Mezanin Floor
Session Chair(s)	Alethea Suryadibrata, Naiwala P. Chandrasiri

36 **An Investigation of Machine Learning Methods for Predicting Bus Travel Time of Mongolian Public Transportation**
Nyamjav Jargalsaikhan and Katsutsugu Matsuyama

40 **Bayes Code for 2-dimensional Auto-regressive Hidden Markov Model and Its Application to Lossless Image Compression**
Yuta Nakahara and Toshiyasu Matsushima

42 **CNN-based Super-Resolution Adapted to Quantization Parameters**
Toshiya Hori, Zichen Gong, Hiroshi Watanabe, Tomohiro Ikai, Takeshi Chujoh, Eiichi Sasaki and Norio Ito

57 **A Study on Liver Tumor Detection from an Ultrasound Image using Deep Learning**
Takahiro Nakashima, Issei Tsutsumi, Hiroki Takami, Keisuke Doman, Yoshito Mekada, Naoshi Nishida and Masatoshi Kudo

112 **Robotic Path Planning using Evolutionary Neural Network**
Johanes Immanuel, Alethea Suryadibrata and Julio Christian Young

Session 4C	Multimedia Applications
Room	Griya Kresna at Mezanin Floor
Session Chair(s)	Shih Zen-Chung, Makoto Hirayama

49 **Extrinsic Parameters Calibration of Multiple Fisheye Cameras in Manhattan Worlds**
Weijie Chen, Sarthak Pathak, Alessandro Moro, Atsushi Yamashita and Hajime Asama

50 **Measurement of Abdominal Shape by Sampling Moiré Method and Extension to Video Processing**
Yuki Mochizuki, Norio Tagawa and Kann Okubo

51 **A Picking Interface Seamlessly Connecting Passive and Active Performance on Guitar**
Nozomu Yoshida and Katsutsugu Matsuyama

76 **A Display Panel System Using Spotlight Projector Mapping for Museum Exhibition**
Makoto J. Hirayama

78 **CSSNet: Image-Based Clothing Styles Switch**
Shao-Pin Huang, Der-Lor Way and Zen-Chung Shih

81 **Semantic Scene Modeling for Aquaculture Management Using an Autonomous Drone**
Hung-Yuan Chen, Shyi-Chyi Cheng and Chin-Chun Chang

Session 4D	Animation and Virtual Reality II
Room	Sekar Jagad at 2nd Floor
Session Chair(s)	Yoshiaki Shishikui

19 **Diminished Reality in Textureless Scenes**
Yuka Namboku and Hiroki Takahashi

85 **Assessing viewer satisfaction of CG programs as a substitute for real TV programs**
Yoshiaki Shishikui, Genki Nakamura and Masaki Hayashi

101 **A VR-based Support System of Self-learning Microscope Operation**
Hikaru Kaneko and Mitsunori Makino

102 **AR based Self Learning System of Japanese Calligraphy Skills**
Yusuke Ishihara and Mitsunori Makino

132 **Blocking Gamelan Instruments Frequency in Virtual Reality**
Clemens Felix Setiyawan and Mohammad Shobri

10.00-10.10 Coffee Break

10.00-11.00

Session 5A	Poster
Room	Foyer
Session Chair(s)	Hiroki Takahashi, M. Shobri, Shinji Mizuno

1 **Speckle Noise Reduction Technique for SAR Images using SRAD and Gradient Domain Image Filtering**

133 **3D model retrieval based on deep learning approach with weighted three-view deep features**
Xuemei Jiang, Yaqi Li, Jiwei Hu and Kin-Man Lam

134 **Discovering Inactive Students patterns and trends by applying Data Warehouse and Visualisation on Student Record**
Mohammad Shobri, Abba Suganda Girsang and Dyah Ayu Anggreini Tuasikal

Session 5B	Special Session Medical Imaging II
Room	Sida Mukti at Mezanin Floor
Session Chair(s)	Kim Jong Hyo

10 **Medical image fusion via discrete wavelet transform and fuzzy radial basis function neural network**
Zhen Chao and Hee Joung Kim

98 **Automatic Meniscus Segmentation using Cascaded Deep Convolutional Neural Networks with 2D Conditional Random Fields in Knee MR Images**
Uju Jeon, Hyeonjin Kim, Helen Hong and Joon Ho Wang

119 **A web based computer aided diagnosis system on liver disease**
Ju Huang, Huining Huang, Bijiang Li, Xuejun Zhang, Binmei Liang, Dongbo Wu

122 **Segmentation of Pulmonary Abnormalities on CT Images using Deep Learning**
Shoji Kido, Yuki Suzuki and Noriyuki Tomiyama

125 **Deep Learning Based Segmentation For Part Solid Nodule In Chest CT**
Young Jae Kim, Ye Rang Park, GyeongYun Yi, Kwang Gi Kim

131 **A Novel Low-Dose Scanning Method For Interventional CT**
Seungryong Cho*, Sanghoon Cho

11.45-12.30 IWAIT2020 Best Papers Award Announcement

	IWAIT Board Representative Closing Remark Room : Ballroom
12.30-13.30	Lunch
13.30-18.00	Borobudur Temple Tour (Group 2)

SPEAKERS

KEYNOTE SPEAKER 1

BAYU ARIA
Batik Expert, Entrepreneur, Academician
Lecturer, Indonesia Arts Institute, Yogyakarta
Hotwax Studio Yogyakarta Owner
October 21st, 2007 Until now
Yogyakarta City

A Classic Notes
ARTS ... a word that is meant to decorate the world
a bag of images and words scattered it with closeness
some go with purity of feeling
some are walking with logic
there is a materialistic degree and sparkling existence
there also who are arrogant without vein
for you art workers
the symbol is your instinct
not saying but telling stories
unconditional but a benefit
this is the art taught by nature and feeling to all
some understand that the heart is the most beautiful dictionary
some understands logic and words is the best dictionary
however ... the most important thing about art is
IDEALISM ... ideal for giving good around him
not selfishness that shackles art to numbness

thanks for God to give us this words ..
Happy Batik Day 2011
greetings culture for you friend :)

KEYNOTE SPEAKER 2

Chieteuk Ahn

Ph.D./Research Fellow

Broadcasting and Media Research Lab.

ETRI, The Future Wave

www.etri.re.kr

Chieteuk Ahn was a vice-president and a principal member of technical staff in the Digital Broadcasting Research Division of ETRI, Korea. He received the BS and MS from Seoul National University, Korea, in Feb., 1980 and 1982 respectively, and the PhD from the University of Florida, USA, in Aug. 1991. Since he joined ETRI in 1982, he has been involved in developing digital switching systems, MPEG standardization and broadcasting technology. His recent work has been focused on MPEG technology as well as on developing interactive multimedia technology. He has served as an HOD of MPEG-Korea and SC29-Korea since 1996. His main interests are in the areas of multimedia signal processing, broadcasting, and communications.

ABSTRACT

TABLE OF CONTENTS

Speckle Noise Reduction Technique for SAR Images using SRAD and Gradient Domain Guided Image Filtering

Hyunho Choi[1] and Jechang Jeong[1,*]
[1]Hanyang University, Department of Electronics and Computer Engineering, Seoul, South Korea

Email: hyunho0619@gmail.com, jjeong@hanyang.ac.kr

Abstract – In this paper, a novel algorithm is proposed using speckle reducing anisotropic diffusion (SRAD) and gradient domain guided image filtering (GDGIF) to reduce speckle in synthetic aperture radar (SAR) images. SRAD is suitable for reducing multiplicative noise in SAR images because it can directly process log-compressed data. Since GDGIF has edge-aware weighting, it is adaptively applied to SRAD result images to additionally reduce speckle noise. Experimental results demonstrate that the proposed algorithm, compared to existing filtering methods, shows excellent speckle noise reduction performance and a low computational complexity.

Photo Spot Recommendation for Theme Park Visitors Using Collage Images

Hayato Tsukushi[1], Tomohiro Fukuhara[1,2], Koichi Yamada[1], Hironobu Abe[1], Hidetaka Masuda1

[1] Tokyo Denki University, 5 Senju-Asahi-cho, Adachi-ku, Tokyo, 120-8551, JAPAN
[2] MULTISOUP CO,LTD, 3-11 Kanda-Nishiki-cho, Chiyoda-ku, Tokyo, 101-0054, JAPAN

Email: 18fmi18@ms.dendai.ac.jp, fukuhara@multisoup.co.jp, yamada@mail.dendai.ac.jp, hironobu.abe@mail.dendai.ac.jp, masuda@mail.dendai.ac.jp

Abstract – Social networking sites (SNS) are nowadays used for posting and seeing beautiful photographs. Such photographs are called instagrammable photographs, and the places where those photographs are taken are called instagrammable sights. Though many visitors click photographs in theme parks and post them to SNS sites, finding appropriate photo spots using SNS sites is not easy for the visitors because of the enormous number of images. The aim of this study is to help visitors find appropriate photo spots by using recommendation algorithms. As our test case, we chose Tokyo Disneyland (TDL), which many people visit and post their photographs from this place on SNS sites, especially Twitter. One of the characteristics of photographs found on Twitter is that several photographs are merged into a single image called collage. Collage images show the places visited and represent an excursion history of visitors. Based on those histories, we use a collaborative filtering algorithm to recommend photo spots to visitors. In order to design a photo spot recommendation system for theme parks, we have to know the intentions and preferences of theme park visitors. Therefore, we conducted a questionnaire survey. As a result, we can assume that male subjects prefer photo spots with good scenery and spots where they can pose. Female subjects tend to choose spots where they can be "instagrammable," "beautiful," and "cute" in the photographs. Based on this result, we will categorize photo spots in TDL and create a prototype recommendation system.

System for Searching Illustrations of Anime Characters Focusing on Degrees of Character Attributes

Yuta Koyama[1], Tomohiro Fukuhara[12], Koichi Yamada[1], Hironobu Abe[1], Hidetaka Masuda[1]

[1] Tokyo Denki University, 5 Senjuasahi-cho, Adachi-ku, Tokyo, JAPAN 120-8551
[2] MULTISOUP CO, LTD, 3-11 Kandanishiki-cho, Chiyoda-ku, Tokyo, JAPAN 101-0054

Email: 18fmi12@ms.dendai.ac.jp, fukuhara@multisoup.co.jp, yamada@mail.dendai.ac.jp, hironobu.abe@mail.dendai.ac.jp, masuda@mail.dendai.ac.jp

Abstract – In social networking sites (SNS) for posting illustrations of the anime characters, users use tags for searching illustrations. However, this requires the premise that users have to attach appropriate tags to the illustration beforehand. In the illustration of anime characters, even if specific information such as names of characters and titles, the obvious attributes (e.g. hairstyle, clothes, facial expression, posture, etc.) have a tendency that tagging is not performed. Therefore, though the users want to view only illustrations of characters with specific attributes, they can't search because tags are not attached. In addition, it is not possible to represent detail attributes such as the hair length. To tackle this issue, we propose a method for performing feature extraction focusing on specific attributes of characters drawn as illustrations as well as generating vector representations of illustrations based on the degrees of character attributes such as their hair length, hair color, and facial expression. With proposed method, the expression of detailed and diverse information of characters even tags are not assigned to illustrations. An overview of the method and results are described.

Wappen: Annotation System using Scene Matching with Multiple Terminals

Kouhei Maeda[1], Tomohiro Fukuhara[1,2], Koichi Yamada[1], Hironobu Abe[1], Hidetaka Masuda[1]

[1] Tokyo Denki University, 5 Senju-asahi-cho, Adachi-ku, Tokyo, JAPAN 120-8551
[2] MULTISOUP CO,LTD, 3-11 Kanda-nishiki-cho, Chiyoda-ku, Tokyo, JAPAN 101-0054

Email: 18fmi26@ms.dendai.ac.jp, fukuhara@multisoup.co.jp, yamada@mail.dendai.ac.jp, hironobu.abe@mail.dendai.ac.jp, masuda@mail.dendai.ac.jp

Abstract – An efficient system to support referencing external information using multiple mobile terminals is proposed. We employ an annotation system as the mechanism to display information to the user. We are developing a prototype system called Wappen to provide the following functions. First, annotation of any scenes on a primary terminal. Second, reference prepared scene annotations on a secondary terminal (when the user is viewing a scene on the primary terminal). The annotation method suggested here can add any information to any scenes of mobile applications. By adding useful information, it can display them corresponding to the status of progress. This method improves the efficiency of various tasks, such as gaming on a mobile terminal while viewing the supplementary information. We report experiment results using the system to confirm a difference in the time needed to obtain the desired information between using two terminals and a single terminal. The subjective evaluation of the system is also performed. 21 subjects participated to the experiment. We found that the time required to obtain the desired information by viewing annotations using two terminals was less than that using a single terminal. Subjective evaluation through a questionnaire also revealed that the use of two terminals obtained better evaluation results than the use of a single terminal.

Sub-window Median-like Filter in Constant Time

Daiki Isono[1], Xiaohua Zhang[1], Wenbo Jiang[2], Yuelan Xin[3]
[1]Hiroshima Institute of Technology, Hiroshima, Japan
[2]Xihua University, Chengdu, China
[3]Qinghai Normal University, Xining, China

Email: {bl16005,zhxh}@cc.it-hiroshima.ac.jp, caswenbojiang@gmail.com, xinyue001112@163.com

Abstract – Due to its simplicity, median filter is a very famous and useful tool in the fields such as image processing and computer graphics. Median filter is mainly for eliminating irrelevant details, especially removing salt-and-pepper noises in image. Median filter has the ability to preserve structural edges compared with box filter and Gaussian filter, however, this ability is very limited. When the radius of filter kernel becomes larger, the edge-preserving ability also becomes very weak. In this paper, we propose a median-like filter that removes small details including salt-and-pepper noises in image while having stronger edge-preserving ability than traditional median filter. The filter computes the output at the observed pixel using 8 sub-windows and a full window. Among these windows, 4 of them are built on the quadrants respectively, other 4 of them are on left, right, top, and bottom half plane. All of these sub-windows contain the observed pixel. Moreover, since medians are computed from histograms, we update column histograms and kernel histograms by simple subtraction and addition operations to accelerate the filtering. The computational complexity of the proposed median-like filter is independent of kernel size and thus is in constant time. A SSIM (Structural SIMilarity) evaluation demonstrates that the proposed median-like filter performs well.

An Estimation Method of the Camera Fluctuation for a Video-based Vibration Measurement

Kazuyuki Miura[1], Phooi Yee Lau[1], Takehiro Tsuruta[1], Atsushi Osa[2]
[1]Research & Development Institute, Takenaka Corporation, Chiba, Japan
[2]Graduate School of Sciences and Technology for Innovation, Yamaguchi Univ, Yamaguchi, Japan

Email: miura.kazuyuki@takenaka.co.jp, tsuruta.takehiro@takenaka.co.jp, osaa@yamaguchi-u.ac.jp

Abstract – A quick assessment of a building that subjected to an earthquake is very important to judge its safety. In some regions, the emergency safety evaluation should be conducted within 24 hours after a huge earthquake occurred. Some structural health monitoring systems enable the rapid evaluation, however they generally require a lot of vibration sensors. Our research group studies a video-based micro-vibration measurement system that can evaluate the safety of buildings without any vibration sensors. It is proposed in this manuscript that an estimation method of the camera fluctuation on the video-based micro-vibration measurement system. The proposed method estimates the camera fluctuation as global movement across the whole image. That is, the method finds a group of pixels with a mode of spatial motion using a time different of a spatial phase. Then, A time variant signals of the mode pixels is estimated as the camera fluctuation. It is demonstrated that the proposed method can estimate a camera vibration frequency in the conditions where multiple subjects exist within the angle of view. Our method works well for some vibrations on the order of 1 / 10 pixel.

Video Search for Ambiguous Requests

Ryunosuke Itabashi, Nobuyuki Yagi
Graduate School of Environmental and Information Studies, Tokyo City University, Yokohama, Japan 224-8551

Email: {g1883101, yagi}@tcu.ac.jp

Abstract – There are countless videos on the Internet with the spread of broadband network. And advances in video analysis technology make it possible to extract more exact metadata and make it easier to find the video we want. But provided search services are not suitable for ambiguous affective video search where a specific query cannot be given such as whether to want to watch at the time of relaxing. To solve the problem, video search method that can handle such ambiguous request utilizing existing search services is considered, and the method to replace an ambiguous request by queries that are group of multiple metadata which are something concrete such as name of objects is proposed. In order to confirm feasibility of the proposed method, classification experiments are performed for three ambiguous requests using multiple metadata automatically attached to the videos by Google Cloud Video Intelligence and so on. As a result, it is confirmed that automatic classification by machine learning using multiple metadata achieved performance close to manual classification. This suggests that the proposed replacing method is feasible.

Hybrid imaging technique of half ROI and full view scan for dose reduction

Dohyeon Kim1, Hyemi Kim2, Minjae Lee1, Hee-Joung Kim1,2*
[1]Department of Radiation Convergence Engineering, College of Health Science, Yonsei University, Wonju, Republic of Korea
[2] Department of Radiological Science, College of Health Science, Yonsei University, Wonju, Republic of Korea

Email: hjk1@yonsei.ac.kr

Abstract – Although many studies of iterative based reconstruction algorithm have been performed on dose reduction in medical applications, there is little research on how to approach the systematical acquisition method to reduce the dose. In this study, we proposed a hybrid imaging technique using half ROI and full view scan as new acquisition method. A prototype of the CT system (TVX-IL1500H, GERI, Korea) was used. The source-to-detector and the source-to-center of rotation distance were 1178 and 905 mm, respectively. A total of 720 projection data were obtained, from which full view CT reconstructed images were obtained as a reference image using a filtered back projection (FBP). Interior ROI CT images were reconstructed using the 720 truncated projection data. Proposed hybrid CT image was reconstructed using 360 truncated and variable number of nontruncated full-size projection data. We set a total of 6 acquisition parameters according to the number of full-size projection data and acquire hybrid images. Dose reduction can be achieved through half ROI and full view scan. As the number of nontruncated full-size projection data on the hybrid image increases, the CNR value increases. In conclusion, the hybrid image with 72 full views were similar to the reference image and the image showed the best image quality among various acquisition parameters.

A Study on 3D Modeling from Handwritten Maps for VR Environment Construction of Historical Town

Ryunosuke Kumagai, Morimichi Furudate,Kouichi Konno
Graduate School of Arts and science, Iwate University

Email:g0319057@iwate-u.ac.jp

Abstract – In recent years, virtual reality (VR) technology has been penetrating the world and some VR techniques are used for learning. In particular, some VR techniques are introduced for learning local histories because VR can reproduce old buildings and terrain at a low price, and many people can experience local history through such VR. One of the problems in creating VR for historical cityscapes is the difficulty of the preparation of 3D landscape models. Many 3D modeling techniques require the 3D shape measuring equipment and the special techniques to handle the software. In particular, materials from which old landform can be obtained are often hand-written old maps, and it is almost impossible to obtain numerical data such as GIS data or contour lines. To overcome the above problems, creating 3D models from such handwritten maps without specific numerical terrain information is required. Therefore, this paper proposes a method to create 3D terrain models from handwritten maps that do not contain any numerical terrain information such as height and gradient for reproducing VR contents of historical cityscape. First, Converting handwritten maps into the image. After scanning a map, the feature lines of the image are traced manually on a PC to delete unusable information and to help estimation of the contours of the topography. Second, The features of topography such as the height of the mountains and gradient are estimated as the numerical terrain data. Finally, According to the above steps, a 3D model for VR content can be created by using the contours and the features of topography estimated. In this paper, the 3D model created by the proposed method is shown as a result. Accuracy evaluation is the future work because it is difficult to compare old terrain and created a 3D model.

Medical image fusion via discrete wavelet transform and fuzzy radial basis function neural network

Zhen Chao[1], Hee-Joung Kim[1,2*]
[1]Department of Radiation Convergence Engineering, College of Health Science, Yonsei University, Wonju, Republic of Korea
[2]Department of Radiological Science, College of Health Science, Yonsei University, Wonju, Republic of Korea

E-mail: chaozhen1000@yonsei.ac.kr, hjk1@yonsei.ac.kr

Abstract – With the rapid development of computer technology and the advent of the information age, diverse medical imaging devices are emerging. However, limited by imaging principles, single-mode images have their own advantages and disadvantages, and it is difficult to fully express all practical information, causing the limitations of diagnosis. Accordingly, medical image fusion is inevitable trend which could integrate or highlight the complementary information, achieve enhanced image quality, reduce redundancy, and provide a reliable diagnosis. In the past, there were many methods that were proposed, but the effect was largely dependent on the experimental data. Based on this, in this study, we proposed a new image fusion method based on discrete wavelet transform (DWT) and fuzzy radial basis function neural network (FRBFNN). First, we analyzed the details or feature information of two images to be processed by DWT. Here, we used a 2-level decomposition, so that each image was decomposed into 7 parts including high frequency sub-bands and low frequency sub-bands. Subsequently, for the parts of the same position of the two images, we substituted them to the proposed FRBFNN. So, with the operation of these seven neural networks, we obtained seven fused parts in turn. Finally, through the inverse wavelet transform, we could get the final fused image. For the training method of neural network, we adopted the combination of error backpropagation algorithm and gravity search algorithm. The final experimental results demonstrated that our method performed significantly better than other algorithms.

Fast Seam Carving for Video Images

Yuta Muraki[1], Kengo Kitamura[2], Koji Nishio[1], Takayuki Kanaya[3], Ken-ichi Kobori[1]
[1]Osaka Institute of Technology, Faculty of Information Science and Technology, Osaka, Japan
[2]Tokyo System Research Corp. (TSR), Tokyo, Japan
[3]Hiroshima International University, Faculty of Health Services Management, Hiroshima, Japan

Email: yuta.muraki@oit.ac.jp, kenichi.kobori@oit.ac.jp, se.sacrifiseescape@gmail.com, t-kanaya@hirokoku-u.ac.jp

Abstract – In recent years, devices with various screen sizes have become widespread. Therefore, when a user browses video images, they are resized to fit the screen. However, depending on the resizing method, the impression of the object may be impaired. Avidan's Seam Carving(SC) is known as a resizing method that preserves the impression of video images. SC resizes an image by removing seams that are not visually important areas in the image. As a result, resizing is performed while maintaining the impression. However, when the background is complex, the visually important object region may be selected as the seam. This will result in resizing that will damage the impression. In addition, the method of Avidan et al. has a problem that processing takes time. Furuta et al. speeded up pixel and frame processing by changing from pixel to voxel. However, there is a problem that the processing cost increases as the number of input video frames increases. In addition, there is a problem that when the scene changes in the video, the background changes and the user feels uncomfortable. In this paper, we propose a resizing method that detects visually important regions from video images and retains them. Moreover, the input video is divided into shot areas, and the same seam is used in the area to reduce recursive calculation of seams. In addition, the cost map used for seam calculation can be processed at high speed by normalizing to 100x100 resolution. In the proposed method, the object obtained by object detection is regarded as an important area.

An Examination on Shape Feature Extraction Based on the Elliptic Approximation for Spatial Arrangement of Earthenware Pieces by Using 3D Measured Point Clouds

Kazuto Yoshikawa[1], Tsutomu Kinoshita[2], Kouichi Konno[1]
[1] Graduate School of Arts and Sciences, Iwate University, Japan
[2] Faculty of Engineering, Tohoku Gakuin University, Japan

Email: g0319169@iwate-u.ac.jp

Abstract – Restoration of earthenware is attempted by estimating the age, pattern, and size of the earthenware based on archaeological findings. However, restoration is more difficult if the number of earthenware pieces is not sufficient. This paper proposes a method to determine the position of height direction without adjacency. In the proposed method, the cross-section of earthenware is calculated and applied by elliptic approximation. Each elliptic curves derived from earthenware are compared to estimate the relative position of height direction. In our examination, the shape of both an elliptic curve derived by the whole cross-section, and an elliptic curve derived by part of the cross-section is estimated. As a result, we verified that our method is effective if the size of a part of the cross-section is more than 1/8.

Detecting Image Frames which Contain a Moving Object from a Severely Distorted Video Stream Using Dynamic Mode Decomposition

Daisuke Uchida[1], Atsushi Yamashita[1], Hajime Asama[1]
[1]Department of Precision Engineering, The University of Tokyo, Tokyo, Japan

Email: uchida@robot.t.u-tokyo.ac.jp, yamashita@robot.t.u-tokyo.ac.jp , asama@robot.t.u-tokyo.ac.jp

Abstract – There are few moving object detection techniques dealing with severely distorted video imagery, such as one taken from above the wavy water surface. In this paper, a method that identifies image frames containing a moving object from a video taken from above the wavy water surface is proposed. Considering the difficulty to apply common video processing techniques to such a video suffering from severe distortion, the proposed method utilizes dynamic mode decomposition, a data-driven method for analysis of dynamical systems, to develop an algorithm that extracts information of a moving object from a video stream. The experimental evaluation show that the proposed method is able to identify image frames containing a moving object from a severely distorted video stream.

A Method of Automatic Cage Generation by Variational Remeshing Method

Takayuki Kanaya[1], Naoyuki Awano[2], Yuta Muraki[3], Kenichi Kobori[3]
[1]Hiroshima International University, Faculty of Health Services Management, Hiroshima, Japan
[2]Osaka University of Economics, Faculty of Business Administration, Osaka, Japan
[3]Osaka Institute of Technology, Faculty of Information Science and Technology, Osaka, Japan

Email: t-kanaya@hirokoku-u.ac.jp, awano@osaka-ue.ac.jp, yuta.muraki@oit.ac.jp, kenichi.kobori@oit.ac.jp

Abstract – In recent years, with the improvement of computer performance, it has become possible to represent dense mesh models in computer graphics. However, performing a manipulation on the dense mesh models might be costly. In order to reduce the computational cost during manipulations, a dense model is often manipulated through coarse bounding cages that enclose the model. However, generating cages are usually tedious and time-consuming. In this paper, we propose a method of automatic cage generation by a variational remeshing method. We first evaluate the features, such as curvature, dihedral angles, of an original triangle model and then voxelize it. We extract and triangulate the outer faces of the voxels and transfer the features of the original model to the outer faces. Finally, we apply a variational remeshing method to this triangular mesh. The variational remeshing method is a method minimizing an energy functional correspond to good solution by global relaxation until convergence. An experiment result demonstrates that our method is effective.

Warping-based motion compensation for triangular patches

Keisuke Kamiya[1], Toshiyuki Yoshida[1]
[1]Dept. of Information Science, Graduate school of Eng., University of Fukui, Fukui 910-8507, Japan

Email: ji180065@u-fukui.ac.jp, tyoshida@fuis.fuis.u-fukui.ac.jp

Abstract – Currently, several efforts are being made to improve coding efficiency of the HEVC (High Efficiency Video Coding) standard, one of which includes an introduction of multi-parameter motion compensation(MC) technique. Warping-based [1] and affine-transformation-based [2-3] MC techniques have been known as promising approaches that overcome the limitation of the conventional motion-vector(MV)-based MC framework. This paper proposes a new MC technique based on triangular patch deformation using a warping technique, which adaptively selects MV-based translation and warping-based deformation. Overall MC efficiency is further improved by introducing hierarchical block partitioning, predictive MV coding, merge mode, and so on. The advantage of the proposed technique is demonstrated by comparing with the MC framework in HEVC.

Speckle Based Pose Estimation for 3D Measurement of the Feature-less Environment by Two Cameras

Hiroshi Higuchi[1], Hiromitsu Fujii[2], Atsushi Taniguchi[3], Masahiro Watanabe[4], Atsushi Yamashita[1], Hajime Asama[1]
[1]The University of Tokyo, Tokyo, Japan
[2]Chiba Institute of Technology, Chiba, Japan
[3]Hitachi, Ltd., Kanagawa, Japan (as of September 2019)
[4]Hitachi, Ltd., Kanagawa, Japan

Email: higuchi@robot.t.u-tokyo.ac.jp, hiromitsu.fujii@p.chibakoudai.jp, masahiro.watanabe.ubg@hitachi.com, yamashita@robot.t.u-tokyo.ac.jp, asama@robot.t.u-tokyo.ac.jp

Abstract – In this research, we propose the 3D measurement system combining structured light and speckle based pose estimation by introducing two different setting cameras. The proposed system consists of two lasers, namely spot laser and line laser, and two cameras, with and without lens, which can obtain both focused and defocused images at once. Local shapes are measured using focused images by structured light method. 3D positions of points projected by laser are calculated by triangulation. Pose changes are estimated from speckle information using defocused images. Displacements of speckle patterns are detected as optical flow by Phase Only Correlation (POC) method. Pose changes are estimated from speckle displacements by solving equations derived from the physical nature of speckle. The target shape as a whole is reconstructed by integrating the local shapes of each image into common coordinates using estimated pose changes. In the experiment, the texture-less flat board was measured with motion. From the experimental results, it is confirmed that the shape of the board was reconstructed correctly by the proposed 3D measurement system.

Human Performance of Face Recognition Inferred from Characteristics of Observing Eye Movement Patterns Learned by Hidden Markov Model

Shun Oue[1], Ryoko Yamada[1], Shigeru Akamatsu[1]
[1]Hosei University, Tokyo, Japan

Email: shun.oue.2x@stu.hosei.ac.jp, ryouko.yamada@akamatsu.info, akamatsu@hosei.ac.jp

Abstract – We investigated the relationship between the face recognition performance of individuals and their eye movement characteristics that were measured while each subject observed the faces. We formulated the statistical nature of the viewers' eye movements from a machine-learning perspective by applying a hidden Markov model (HMM). We used a set of computer-generated faces that included both the images of actual faces and synthetic images obtained by slightly transforming the impressions of the original faces. With these visual stimuli, we conducted a simple face recognition experiment, and participants judged whether they had seen the faces before. We obtained a quantitative hit rate score for each stimulus and subject. We also tracked their eye movements and recorded as temporal chains of their gaze fixation points using an eye-tracking system. For each class of face stimulus and subject, we estimated the HMM parameters from the training samples of the eye movement. For the given eye movement data as test samples, we conducted a classification test among the pre-defined classes based on the differences of the log-likelihood values obtained from each HMM. Better discrimination of the subjects by the HMM-based classification of the eye movement data corresponded to worse face recognition scores achieved by the subjects, suggesting that individually consistent eye movement patterns may lower the performance of face recognition by human.

A Method of Finding Characteristic Ocean-Satellite-Image Groups using Autoencoder

Kenta MARUYAMA[1], Katsutsugu MATSUYAMA[1]
[1]Iwate University, Graduate School of Arts and Sciences, Iwate, Japan

Email: g0319147@iwate-u.ac.jp, m18u@iwate-u.ac.jp

Abstract – We propose a visualization method that takes ocean-satellite-images and fishery data for understanding the sea conditions when catch amount is high. In this research, we focus on ocean-satellite-images with date information and fishery data which is a triplet list of (date, fish species, catch amount). We select the sea of Iwate prefecture in Japan as our target. Our method employs an autoencoder to calculate the similarity between images. Autoencoder is a learning technique of neural network that takes same data to the input and output of a network model. An autoencoder can be considered a pair of an encoder and decoder. Encoder transforms the high-dimensional input data into a low-dimensional feature vector and decoder recovers the data from the feature vector. We employ convolutional neural network model, giving our ocean-satellite-images to the input and output on learning stage. As a result, the feature vector of each image can be calculated. After using autoencoder, images are grouped by features. First, each image is given 'Positive', 'Negative' or 'None' label based on catch amount. For each Positive image, we find ten nearest neighbors in feature vector space. If the number of Positive in the group is greater than or equal to the threshold $\boldsymbol{\alpha}$ ($\boldsymbol{\alpha} = 6$ in this paper), we judge the group expresses sea conditions of high catch amount. Number of neighbor and threshold value $\boldsymbol{\alpha}$ are selected by trial and error. Our result shows that each extracted image group has high similarity and different group are visually distinctive. We expect that our result is helpful for examining sea conditions in which catch amount will be high.

Diminished Reality in Textureless Scenes

Yuka Namboku and Hiroki Takahashi
Graduate School of Informatics and Engineering, the University of Electro-Communications 1-5-1 Chofugaoka Chofu Tokyo, Japan

Email: yuka.namboku@uec.ac.jp, rocky@inf.uec.ac.jp

Abstract - Diminished Reality (DR), which removes real ob- jects visually, is expected in various scenes. It is, however, difficult to remove an object in some scenes with textureless backgrounds. This paper, therefore, proposes a method to re- alize DR in textureless scenes based on feature lines. Missing backgrounds are filled by each of divided background planes based on geometric information. The background is divided by boundaries of the background planes obtained by using Line Segment Detector (LSD), Hough transformation and the condition for concurrency of three straight lines. A removal object is tracked by feature point matching. Missing background are filled by image inpainting only in the first frame. In the second frame and after, removal regions obtained by projecting background boundaries are overlaid by the complemented background image. Proposed approach achieves removing an object in textureless scenes. However, stable DR cannot be realized with the influence of camera pose estimation error. In our future work, we will investigate more robust camera pose estimation, and refine estimates using bundle adjustment.

Automatic Classification of Manga Characters using Density-Based Clustering

Hideaki Yanagisawa[1], Kengo Kyogoku[1], Jain Ravi[2], Hiroshi Watanabe[1]
[1]Waseda University Graduate School of Fundamental Science and Engineering
Department of Communications and Computer Engineering, Tokyo, Japan
[2] Waseda University Graduate School of Fundamental Science and Engineering
Department of Computer Science and Communications Engineering, Tokyo, Japan

Email: bule-cosmo@ruri.waseda.jp, kengo.0602@akane.waseda.jp,
vainaijr114@akane.waseda.jp, hiroshi.watanabe@waseda.jp

Abstract – Manga (Japanese comic) is a globally popular content. In recent years, e-comics that converted from manga to electronic image data have become popular. Various services can be provided by automatically extracting and tagging elements in electronic comics such as characters or speeches. In order to automatically extract characters from manga, a character recognition method that does not use domain knowledge is required. One way to classify characters is to get image features from the character's faces cut out from manga images and cluster them. Previous research has shown that using the intermediate output of CNN fine-tuned with character face images is effective for character recognition. In order to classify face images without specifying the number of clusters, we proposed a clustering method using Density-Based Spatial Clustering of Applications with Noise (DBSCAN). However, DBSCAN is greatly affected by the hyperparameter setting. The purpose of this study is to automatically classify character face images without complicated hyperparameter setting. We examine the application of Ordering Points to Identify the Clustering Structure (OPTICS) and Hierarchical DBSCAN (HDBSCAN), which are density-based clustering algorithms that extend DBSCAN. OPTICS is an algorithm for finding clusters in spatial data, and HDBSCAN extracts flat partition from hierarchical cluster data. We also verify the effective CNN model as the feature extractor of facial images. Experimental results show that HDBSCAN is effective for character face image clustering. We also confirmed that the clustering performance improved as the number of CNN layers in the feature extractor increased.

Surface Segmentation on 3D Point Cloud of Unbroken Earthenware and Its Applications

Keita Hayasaka[1], Tsutomu Kinoshita[2], Kouichi Konno[1]
[1]Graduate School of Engineering, Iwate University
[2]Faculty of Engineering, Tohoku Gakuin University

Email: g0319129@iwate-u.ac.jp, kinoshita@mail.tohoku-gakuin.ac.jp, konno@cis.iwate-u.ac.jp

Abstract – Many kinds of research for point clouds have been studied in a wide range of fields. In archaeology, Jomon Potteries are various shapes depending on the areas and ages of production. In order to investigate the characteristics of the potteries, analysis methods of rim parts and surface patterns are required. When the shape of surfaces is analyzed, the point cloud which is measured by photogrammetry may be separated by the front and back surface. Therefore, this paper proposes a method to segment a point cloud of pottery, obtained by photogrammetry. In our method, point clouds are affine-transformed so that the z-axis is assumed as the central axis. In addition, the scale of the point cloud is changed based on the size of the actual earthenware measured with a ruler. After that, cross-section among the z-axis is calculated and analyzed whether the points belong to the inside surface or outside surface. Our method is verified by applying some point clouds.

Fast Total-Variation-based JPEG Artifact Removal via the Accelerated ADMM

Chihiro Tsutake[1], Ryoya Yamada[1], Ryoga Yamazaki[1], and Toshiyuki Yoshida[1]
[1]Department of Information Science, University of Fukui

Email: ctsutake@icloud.com

Abstract –JPEG artifact removal has been an active research topic, which focuses on restoring an image from the JPEG bitstream without artifacts such as blocking and ringing. Recently, TV-based techniques have been proposed in the literature, where a constrained minimization problem is solved based on iterative convex optimization methods, i.e., the projected subgradient descent (PSD) and the primal-dual algorithm (PDA). However, these techniques commonly require unrealistic computational cost because of the large number of iterations for the optimization. To overcome the difficulty, we utilize the accelerated Alternating Direction Method of Multipliers (ADMM) instead of the conventional optimization methods, and provide its explicit iterative process. The experimental results demonstrate that our technique outperforms existing ones in terms of computational cost.

Noise reduction in direct multi-material decomposition for dual-energy CT

Haenghwa Leea, Dong-Hoon Leeb, Pil-Hyun Jeona, Zhen Chaob and Hee-Joung Kima,b,*

[a]Department of Radiological Science, Research Institute of Health Science, Yonsei University, Wonju, Korea, 220-710; [b]Department of Radiation Convergence Engineering, 220-710

E-mail: hjk1@yonsei.ac.kr

Abstract – Dual energy computed tomographic (DECT) enhances tissue characterization by obtaining two or three material images from two measurements at two different X-ray source potential. Recently, developing for multi-material decomposition (MMD) in DECT has been studied to obtain decomposed material images. MMD need to reduce noise and maintain spatial resolution of decomposed images. However, no studies have reported total nuclear variation (TVN) as noise suppression method for MMD to improve decomposition accuracy. We proposed a noise suppression using TVN for the direct MMD. The TVN method was applied to CT data before material decomposition to reduce noise. Tissue characterization Model 467 phantom was employed as the test object in this study. To investigate the effect of various basis materials, we selected four materials as basis materials. The volume fraction (VF) value was calculated to quantitatively evaluate quality of decomposed images. The results are compared to direct MMD method and proposed method. In all decomposed images, VF accuracies using proposed method were better than the direct MMD method. Also, proposed method can provide decomposed images with a small difference in separated density. In conclusion, proposed method could provide better quantitatively accurate images.

Transform selection for video coding

Nam Uk Kim[1], Yung-Lyul Lee[*1]
[1]Department of Computer Engineering, DMS Lab, Sejong University, Seoul, Korea

Email: nukim@sju.ac.kr, yllee@sejong.ac.kr

Abstract – In video coding algorithm, a transform is one of the most important video compression tools. Transform converts residual signals into frequency domain data to get decorrelated signals. Decorrelated data using transforms is used for coding efficiency improvement in video compression. In this paper, an efficient transform selection method in term of video compression is described by using transform block size and intra coding mode on the top of the HEVC standard.

Extraction of Distinctive Keywords and Articles from Untranscribed Historical Newspaper Images

Sora Ito[1], Kengo Terasawa[2]
[1]Future University Hakodate, Graduate School of System Information Science, Hakodate, Japan
[2]Future University Hakodate, Department of Media Architecture, Hakodate, Japan

Email: g2118005@fun.ac.jp, kterasaw@fun.ac.jp

Abstract – This paper proposes a novel approach to extract distinctive keywords from historical newspaper images without using character recognition. We converted an image of the text block on an entire newspaper page into a sequence of codes based on discretization of the feature vectors, an approach that eliminated the errors introduced by optical character recognition (OCR). This conversion makes it possible to analyze untranscribed newspaper images by using text-processing methods. We examined the daily occurrence of every tri-gram string, and extracted strings with uneven appearance as distinctive keywords. In addition, we highlighted articles that contain distinctive keywords as distinctive articles. The proposed method was evaluated on an archive of Japanese newspaper images published in the 19th century, and the results were promising.

Background Subtraction via Exact Solution of Bayesian L1-norm Tensor Decomposition

Taehyeon Kim, Yoonsik Choe*
Yonsei University, School of Electrical and Electronic Engineering, Seoul, Korea

Email: pyomu@yonsei.ac.kr, yschoe@yonsei.ac.kr

Abstract – In background subtraction, Principal Component Analysis (PCA) based algorithms has shown a remarkable ability to decompose foreground and background in video acquired by static camera. The algorithm via closed form solution of L1-norm Tucker2 decomposition is one of the real-time background subtraction algorithms. The closed form solution can be obtained from linear combination of video frame vectors and coefficient vector composed of only +1 and -1. However, since the optimal coefficient vector is unknown, this method cannot help to be a complicated combinatorial optimization problem when the number of input frame is large. In this paper, to solve this problem, the proposed method utilizes the Bayesian optimization (BayesOpt) which is a black-box derivative-free global optimization is proposed. This method finds the optimal coefficient without considering the linear combination of all possible coefficient-combinations, using Bayesian statistical model and Expected Improvement (EI) acquisition function. Here the Bayesian statistical modeling is the method that measures the uncertainty of unsampled coefficient combination points and the EI function is a surrogate function which indicates the next sampling coefficient combination points. The experimental results confirm the efficiency of the proposed method.

GAN based Single-image Reflection Removal using Depth of Field Guidance

Miran Heo, Yoonsik Choe*
Yonsei University, School of Electrical and Electronic Engineering, Seoul, Korea

Email: pyomu@yonsei.ac.kr, yschoe@yonsei.ac.kr

Abstract – Eliminating reflections on a single-image has been a challenging issue in image processing and computer vision, because defining an elaborate physical model to separate irregular reflections is almost impossible. In fact, while human vision can automatically focus on the transmitted object, basic deep neural networks even have a limitation to learn the attentive mechanism. In this paper, to solve this problem, a Generative Adversarial Networks guided by using Depth of Field (DoF) is proposed. The DoF is formulated by using image statistics and indicates the focused region of image. Thus, by adding this information to both generative and discriminative networks, the generator focuses on the transmitted layer and the discriminator will be able to estimate the local consistency of the restored areas. Since it is intractable to obtain the ground-truth transmitted layer in real images, a dataset with synthetic reflection is considered for quantitative evaluation. The experimental results demonstrate that the proposed method outperforms the existing approaches in both PSNR and SSIM. The visual outputs indicate that the proposed network convincingly eliminates the reflection and produce sufficient transmitted layers as compared to the previous methods.

Generating 3D Model of Furniture from 3D Point Cloud of Room

Shunta Osakama, Yoshitsugu Manabe, Noriko Yata
Chiba University, Chiba, Japan

Email: shunta-0424@chiba-u.jp, manabe@faculty.chiba-u.jp, yata@chiba-u.jp

Abstract —This study proposes a method generating a 3D model of furniture from 3D point cloud data of a room captured by RGB-D camera in order to realize the layout simulation of the real room with furniture.

Environment Understanding During Walking via Modality Conversion from Visual to Haptic Information

Sinichiro Morita, Yoshihiro Kanehara, Kazuho Kito, Harumi Kobayashi, Naoki Mukawa, Akio Nakamura
Tokyo Denki University, Department of Robotics and Mechatronics, Tokyo, Japan
Email: {morita.s, nakamura}@is.fr.dendai.ac.jp

Abstract – The purpose of this research is to propose the method providing the visually impaired users with environment understanding via modality conversion from visual distance to haptic vibration information. According to studies on ecological perception, optical flow, which represents dynamical visual variation, plays significant roles in environment understanding for human. We have developed a head-mounted wearable device equipped with a 2-dimensional distance sensor and five vibro-motors arranged around the head. The vibration magnitude of vibro-motors is defined as not static distance to obstacles but dynamic distance variation caused from the user movement. The vibro-stimuli are generated based on the following optical flow characteristics: A) The magnitude depends on movement intensity. The movement intensity is defined as displacement of body center within average stride time 0.6 s. B) The magnitude depends on distance to the obstacles. The distance is defined as weighed distance; as the distance gets smaller, the distance obtains bigger weight. Thus, if the user moves to the obstacles dynamically, the user feels stronger vibration; in contrast, if the user pauses, the user feels no vibration since the distance to obstacles keeps constant. To evaluate basic performance of the proposed method, we asked five blindfolded subjects to walk toward the wall 50 times each. First 10 trials performed by each subject are considered as practice phase; hence, the remaining 40 trials are evaluated. In total 200 trials, the 97.5% of trials are considered as success ones; the almost subjects were able to perceive the wall existence and stopped without the collision with the wall. The experiment demonstrated the basic validity of our proposed method. The main contribution of this paper is utilizing dynamic distance variation to determine vibration magnitude and providing the user with vibro-stimuli simulated by optical flow to support the user's localization in the environment.

An approach to the multiplex reconstructing process for spatially multiplex projected holographic colored images using blue-violet laser light

Kunihiko Takano[1], Kazuki Audo[1], Miyu Osaka[1], Yuya Kawasaki[1], Koki Sato[2], Kikuo Asai[3]

[1] Tokyo Metropolitan College of Industrial Technology, 8-17-1, Minami Senjyu,Arakawa-ku,Tokyo,116-8523, Japan

[2] Former Shonan Institute of Technology, 1-1-25, Tsujido Nishi kaigan, Fujisawa City, Kanagawa,251-8511,Japan

[3] The Open University of Japan, 2-11,Wakaba,Mihama-ku, Chiba City,Chiba,261-8586,Japan

Email: ktakano@metro-cit.ac.jp

Abstract – In order to display the holographic reconstructed image in colors, we must sufficiently study the characteristics of the images reconstructed by blue-violet color laser light. But in general, in the reconstruction of the images, if we use the laser lights of the short wave length, the reconstruction may become more difficult. Moreover, as for the reconstruction adopted blue-violet color laser light, not so many results seem to have been reported earlier. It appears to lead us to an unfortunate problem of the restriction on the color region of the displayed holographic images. In this report, we shall study about the possibility of the holographic image reconstruction adopted blue-violet laser light (wavelength: 405nm) in order to make much efforts to enlarge the presentable color region of the reconstructed images. We studied a process which makes the two colored images obtained by time-sharing multiplex reconstruction using two colors: blue-violet, red (or green), by multiplexing on the screen spatially. As this result, we have confirmed that single color multiplex reconstructing process recovering the divided parts of the objects formed with multiple points is also applicable to the spatially multiplexed reproducing process using two colors. This seems to suggest some possibility of an extension of the color region for the recovered holographic images.

ElectroMagnetic Guitar: Chord Playing Support System on Guitar by Electromagnets

Nozomu Yoshida[1], Katsutsugu Matsuyama[1]
[1]Iwate University, Faculty of Science and Engineering, Iwate, Japan

Email: s0616063@iwate-u.ac.jp, m18u@iwate-u.ac.jp

Abstract – We introduce ElectroMagnetic Guitar (EMG), a new guitar interface that supports chord playing mainly for beginner guitar players. In case of right handed, we play the guitar with left hand fingers on a fretboard and pick strings by right hand fingers, our guitar guides the player's fingering of left hand by magnetic force. To realize such a concept, we made an original fretboard with controllable electromagnets mounted and attached it to the normal classic guitar. EMG 'draws' player's left hand fingers with magnetic material attached toward proper chord positions on the fretboard.

In our proposed method, an original fretboard with 18 electromagnets (6 strings * 3 frets) mounted is attached to the normal classic guitar. Each electromagnet can be switched ON/OFF by Arduino. We lift string height up by attaching another nut and saddle. Finally, small iron plates are attached on each finger of the player's left hand. 3V dry cells serve power to each electromagnet.

We conducted an informal study with 24 participants. This study showed a primary issue: weakness of magnetic force. The weakness seems to be caused by the exponential attenuation of magnetic force according to distance between the electromagnet and iron plate. As for magnetic force; 1, 16, and 3 participants reported it to be absent, weak, and adequate respectively. 19 people reported enough power to keep their chord form. The user study showed our EMG offered sufficient magnetic force to keep chord form though not enough power to properly draw fingering. Higher power voltage should be used to strengthen EMG draw force of fingering.

Automated detection of fundic gland polyps and hyperplastic polyps from endoscopic images using SSD

Koki Oshio[1], Nagito Shichi[1], Junichi Hasegawa[1], Tomoyuki Shibata[2]
[1]Faculty of Engineering, Chukyo University, Aichi, Japan
[2]Department of Gastrointerology, Hujita Health University, Aichi, Japan

Email:t416009@m.chukyo-u.ac.jp, hasegawa@sist.chukyo-u.ac.jp, shibat03@hujita-hu.ac.jp

Abstract – In recent years, for reducing diagnostic burdens in stomach screening, a computer aided diagnostic system (CAD system) for endoscopic stomach images is required. In our previous study, an automated polyp detection method from endoscopic images using the SSD (Single Shot MultiBox Detector) has been developed with 93.7% of detection rate. However, the detection target of this method has been limited only to fundic gland polyp. In this paper, we propose a method for automated detection and classification of two different types of polyp; fundic gland polyp (FGP) and hyperplastic polyp (HP) from endoscopic images using the SSD. In the experiment, 71 and 96 practical endoscopic images of FGP and HP were used. For training of SSD, 11210 and 5053 training images of FGP and HP were generated by data augmentation, respectively, and 20% of training images were automatically selected and used as verification images. As a result for test samples including 132 polyps (69 FGPs and 63 HPs), the detection rate for entire polyps was 96.2% (127/132), and the classification rate for two types of polyp was 88.6% (117/132). The number of false positive was only one all through the experiment.

Tree Growth Model for Simulation of Appearance Change

Maya Takatsuka, Yoshitsugu Manabe, Noriko Yata
Chiba University, Chiba, Japan

Email: takatuka.m@chiba-u.jp, manabe@faculty.chiba-u.jp, yata@chiba-u.jp

Abstract— In urban development, it is important to make a plan that takes into account the changes in the appearance of natural objects after decades. In particular, the appearance of trees varies greatly, so it is necessary to manage to log. This study proposes a simulation method of tree growth for the prediction of the appearance change of natural objects.

Avatar's Facial Expression with "Manpu (Comic Symbols)" by Using Multiple Biometric Information

Shu Gemba[1], Tokiichiro Takahashi[2 3]
[1] Tokyo Denki University Graduate School, Tokyo, Japan
[2] Tokyo Denki University, Tokyo, Japan
[3] ASTRODESIGN, Inc., Tokyo, Japan

Email: s-gemba@vcl.jp, toki@vcl.jp

Abstract – It is a challenging issue to generate avatar's natural facial expressions from user's facial images. One of the most difficult problems is to analyze user's facial images, and estimate his emotions. We have already proposed a method that estimated user's emotions combining both analysis on user's facial images taken from a video camera and measuring heart rate information, pNN50. Here, pNN50 is a percent of difference between adjacent heartbeat intervals greater than 50 *ms* calculated from heart rate information. Each user's emotion was estimated as either of 'Joy', 'Disgust', 'Sadness', or 'Anger'. Our previous method estimated user's emotions and generated appropriately avatar's face images. However, further improvement of measurement of each emotion is needed for the natural facial expressions of the avatars. In this paper, we propose a new method which measures the skin conductance level (SCL) by using a biometric sensor. We focus on "Russell's circumplex model of affect", in which each emotion is described as a 2D vector of the 'valence' and 'arousal'. Our method calculates the 'valence' by measuring pNN50, and the 'arousal' from the SCL. The four quadrants of "Russell's circumplex model of affect" correspond to 'Joy', 'Disgust', 'Sadness', and 'Anger', respectively. Based on results of multiple biological information, i.e., pNN50 and SCL, the proposed method estimates the accuracy of emotions and generates the avatar's natural facial expressions by adding comic symbols (called "Manpu") successfully and appropriately. "Manpu" can emphasize emotions. By using "Manpu", human emotions can be more easily communicated.

An Investigation of Machine Learning Methods for Predicting Bus Travel Time of Mongolian Public Transportation

Nyamjav Jargalsaikhan, Katsutsugu Matsuyama
Graduate School of Arts and Sciences, Iwate University, Iwate, Japan

Email: jnyamjav@gmail.com, m18u@iwate-u.ac.jp

Abstract – Many public bus services have their timetable which provide time information of arrival and/or departure at waypoints along the route. On the other hand, there are some bus services that do not have fixed time schedule, such as in Ulaanbaatar, the city of Mongolia. In this case, a lot of confusion occurs for the passengers. Prediction of bus travel time can help to provide services such as efficient scheduling for the passengers of their trips by avoiding to wait a long time. For this purpose, we investigate some machine learning methods to predict bus travel time. Concretely, for bus travel data, we employ three regression methods: linear regression (LR), support vector regression (SVR) and artificial neural network (ANN) to predict travel time. The performances of these machine learning methods are estimated and compared using conventional measures such as mean absolute error and root mean squared error. In a quantitative study, the artificial neural network is the best model having errors less than 1 minute in most cases. We also performed a qualitative study to investigate the details of our prediction results by using heatmap visualizations. Our visualization results offer easily grasping the tendency of travel time and error values.

A Study on Thermal Image Generation Based on Deep Learning and Abnormal Temperature Detection

Ziyun Zhang, Makoto Hasegawa
Tokyo Denki University, Tokyo, Japan

Email: 19kmc14@ms.dendai.ac.jp, hasegawa@mail.dendai.ac.jp

Abstract – A thermal image taken by a thermal camera and an RGB color picture were arranged in a dataset pair; the datasets were learned using a deep learning algorithm called pix2pix. After the sufficient training in the machine learning, it was possible to generate a thermal image from a color image taken with a digital camera. This paper provides a new method of generating thermal image without a thermal camera; the thermal camera is required when we create teaching materials. Furthermore, a method for detecting abnormal temperatures using deep learning is proposed. Features of thermal images are concerned and evaluated the results of our method.

Articulation awareness with 3D tongue using VR system

Sayoko Takano [1], Makoto J. Hirayama[3]
[1]Media informatics, Kanazawa Institute of Technology (KIT), Kanazawa, Japan
[2] Information Science and Technology, Osaka Institute of Technology, Osaka, Japan

Email: tsayoko@neptune.kanazawa-it.ac.jp, makoto.hirayama@oit.ac.jp

Abstract – We propose an articulation awareness system with 3D tongue using VR. Human speech sound is made from combination of vocal fold vibration (voice source) and tongue/lip motion (articulation). Articulation are mainly related to movement of jaw, tongue, and lip. One important aspect is speaker should pay deep attention to the importance of these movement. In this research, the tongue shape is visualized to give awareness for speech organ. Subjects observe inside and outside of the mouth as if he/she is the size of thumb. The models of the oral area are made from Magnetic Resonance Image (MRI) data during vowel production. The subjects reported that they had awareness to the articulators after experiencing the 3D-tongue using VR system.

Development of An Interface Simulating Swimming Motion for Virtual Space Walk-Through

Kenta Uchida[a], Keito Tanizaki[a], Tomoaki Moriya[a], Tokiichiro Takahashi[ab]
[a]Tokyo Denki University, Senju-Asahi-Cho 5, Adachi-Ku, Tokyo, Japan 120- 8551;
[b]ASTRODESIGN, Inc., 1-5-2 Minami-Yukigaya, Ohta-Ku, Tokyo, Japan 145-006

Email: k-uchida@vcl.jp, moriya@vcl.jp, toki@vcl.jp

Abstract – Virtual Reality (VR) contents have been becoming popular. However, VR would certainly cause VR motion sickness. One of the main factors in VR motion sickness might be the discrepancy between the predicted physical experiences and the actual physical experiences. To prevent such discrepancy, an approach is effective that synchronizes the user's physical movements in the virtual space with the user's somatosensory movements. There has been developed a VR system which provides such a somatosensory interface. However, it is too huge, expensive and difficult to operate.

In this paper, we propose a simple somatosensory interface for moving virtual space by imitating swimming movements. The users of our system move the virtual space like to swim in the water by the kicks of both foot and the strokes of both arms. We developed a sensory interface that measures user's imitating swimming movements by four gyro sensors. Each gyro sensor is a 9-axis gyro sensor, including a 3-axis angular velocity sensor, a 3-axis acceleration sensor, and a 3-axis geomagnetic sensor. Four gyro sensors measure the movements of the strokes of both arms and kicks of both feet, respectively.

Five students evaluated our interface. They use our interface to navigate the specified course. After that, they answered the questionnaires. Experimental results verified that our interface was useful and effective for moving the virtual space without causing VR motion sickness.

Bayes Code for 2-dimensional Auto-regressive Hidden Markov Model and Its Application to Lossless Image Compression

Yuta Nakahara[1] and Toshiyasu Matsushima[2]
[1]Waseda University, Center for Data Science, Tokyo, Japan
[2]Waseda University, Department Pure and Applied Mathematics, Tokyo, Japan

Email: yuta.nakahara@aoni.waseda.jp and toshimat@waseda.jp

Abstract – For general lossless data compression in information theory, researchers have repeated expansion of stochastic models to express target data and design of codes for the expanded models. In this paper, we apply this approach to lossless image compression. We expand an auto-regressive hidden Markov model to a 2-dimensional model to express images containing a diagonal edge. Then, we design a Bayes code with an approximative parameter estimation by variational Bayesian methods. Experimental results for both of synthetic and real images show that the proposed model is sufficiently flexible for the target images and the parameter estimation is accurate enough.

Depth Estimation with Tilted Optics by Multi-Aperture Using Color Filter

Hiroshi Ikeoka[1], Takayuki Hamamoto[2]
[1]Fukuyama University, Faculty of Engineering, Hiroshima, Japan
[2]Tokyo University of Science, Faculty of Engineering, Tokyo, Japan

Email: ikeoka@fukuyama-u.ac.jp, hamamoto@ee.kagu.tus.ac.jp

Abstract – We have been investigating a depth estimation system for real-time applications. Stereo camera method is too sensitive to slight variations of baseline length due to vibration and temperature, and it has occlusion problem. Conversely, monocular camera method by focusing cannot provide a balance between wide-area estimation and real-time estimation. Therefore, we proposed a novel method that adopts tilted lens optics. In this method, the plane of sharp focus (POF) lies and the depth of field (DOF) enlarges toward the depth direction. Herein, we can obtain depth values at each pixel, which are calculated from the ratio of sharpness values of two tilted optics images, by using monocular camera system that includes two image sensors and spectroscopic. Moreover, our method can allocate different range and different resolution at each y-coordinate. In this paper, we introduce new optical setting for our depth estimation method, which uses only one pair of an image sensor and a tilted optics. It is realized by multi-aperture using color filters. In addition to a normal aperture, our optics has another smaller aperture using green color filter. Therefore, two types of blur images are obtained by only one shot without spectroscopic. An image of the green channel has larger DOF than the red and blue channels. Therefore, it is possible to calculate sharpness ratio between the green channel image and the red-blue channels image pixel by pixel; the red-blue channels image is created by simple average of the red and blue intensity. Hence, our method can obtain the depth value from the sharpness ratio by the same way as before. Consequently, it is succeeded by our proposal to realize estimating depth in real-time with the simple imaging device.

CNN-based Super-Resolution Adapted to Quantization Parameters

Toshiya Hori1, Zitchen Gong1, Hiroshi Watanabe1, Tomohiro Ikai2, Takeshi Chujoh[2], Eiichi Sasaki[2] , Norio Ito[2,]

[1]Waseda University, 3-14-9 Okubo, Shinjuku-ku, Tokyo, Japan

[2] Sharp Corporation, 1-9-2, Nakase, Mihama-ku, Chiba, Japan

Email: tsy-h@fuji.waseda.jp, zichen.gong@fuji.waseda.jp, ikai.tomohiro@sharp.co.jp, chujoh.takeshi@sharp.co.jp, eiichi.sasaki@sharp.co.jp, itoh.norio@sharp.co.jp

Abstract – In video transmission, the videos are encoded and decoded. At that time, bit control is performed by specifying the quantization parameter (QP). The video undergoes various processing to remove redundancy and then orthogonally transforms the video signal into the frequency domain. The frequency domain coefficients are then quantized and transmitted. At that time, by specifying QP, the quantization step is changed, and the amount of data can be changed. Now we are thinking about a coding system using super-resolution. In normal super-resolution, the image pairs used as datasets are not encoded. Thus, trained weights can be applied to any image in the training environment. However, for encoded images, the degradation changes at each image frequency. Therefore, the super-resolution effect will change even if the input image is encoded with the same QP and the QP of the training image and the QP of the input image do not match. In this paper, we will examine the cause of the problem of weight mismatch. It solves the weight mismatch problem by training super-resolution with image pairs with the same degradation rate instead of encoded image pairs.

A Selective Fusion Module for Video Super Resolution with Recurrent Architecture

Zichen Gong[1], Toshiya Hori[1], Hiroshi Watanabe[1], Tomohiro Ikai[2], Takeshi Chujoh[2], Eiichi Sasaki[2], Norio Ito[2],
[1]Waseda University, 3-14-9 Okubo, Shinjuku-ku, Tokyo, Japan
[2]Sharp Corporation, 1-9-2, Nakase, Mihama-ku, Chiba, Japan

Email: zichen.gong@fuji.waseda.jp, tsy-h@fuji.waseda.jp, ikai.tomohiro@sharp.co.jp, chujoh.takeshi@sharp.co.jp, eiichi.sasaki@sharp.co.jp, itoh.norio@sharp.co.jp

Abstract - As an important subtask of video restoration, video super-resolution has attracted a lot of attention in the community as it can eventually promote a wide range of technologies, e.g., video transmission system. Recent video super resolution model [1] achieves cutting-edge performance. It efficiently utilizes recurrent architecture with neural networks to gradually aggregate details from previous frames. Nevertheless, this method faces a serious drawback that it is sensitive to occlusion, blur, and large motion changes since it only takes the previous generated output as recurrent input for the super resolution model. This will lead to undesirable rapid information loss during the recurrently generating process, and performance will therefore be dramatically decreased. Our works focus on addressing the issue of rapid information loss in video super resolution model with recurrent architecture. By producing attention maps through selective fusion module, the recurrent model can adaptively aggregate necessary details across all previously generated high-resolution (HR) frames according to their informativeness. The proposed method is useful for preserving high frequency details collected progressively from each frame. This improves the average quality of the super resolution video. Furthermore, when the recurrent model is used for video compression (ground truth HR image is given in the beginning), the proposed selective fusion makes a significant improvement that lower the decreasing speed of the related quality index.

A Method for Enhancing Playground Equipment Experience using VR Technology with Smartphone

Yoichi Shinchi[1], Shinji Mizuno[1]
[1]Aichi Institute of Technology, Japan

Email: b19717bb@aitech.ac.jp, s_mizuno@aitech.ac.jp

Abstract – We propose a method to make children's playground equipment more thrilling using VR technology only with smartphones. In this research, the children's playground equipment we used includes a seesaw, a globe jungle gym, a swing, and so on. A user plays a playground equipment while wearing smartphone VR goggles. Another smartphone is put on the playground equipment to acquire its movement. The same playground equipment used in the real space is prepared in the VR space, and the user can play with it in the VR space as well as the real space. At this time, the size and the movement of the playground equipment can be changed in the VR space. By increasing the size or accelerating the movement of the playground equipment, the equipment for children can become more thrilling in the VR space. Several studies have been reported that extend or exaggerate real experiences by visual manipulation using VR technology . They require dedicated facilities and equipment. On the other hand, our method uses only common playground equipment and smartphones. As long as there is a seesaw or a globe jungle gym in the park, the park will turn into an amusement park with more thrilling equipment by visual exaggeration using only two smartphones. We implemented a prototype system and tested our method.

A Study of AR Advertising to Estimate the Volume of Invisible Contents in a Packaged Product

Natsumi Motoura[1], Zentaro Kimura[2], Mie Sato[3]
[1] Utsunomiya University, Graduate School of Regional Development and Creativity, Utsunomiya, Japan
[2] Utsunomiya University, Graduate School of Engineering, Utsunomiya, Japan
[3] Utsunomiya University, Faculty of Engineering, Utsunomiya, Japan

Email: more46v@gmail.com, 142929zkimura@gmail.com, mie@is.utsunomiya-u.ac.jp

Abstract – In recent years, advertising with Augmented Reality (AR advertising) has gathered attention as a new method of information presentation for sales promotion. Since AR advertising can present more information than conventional advertising written on paper and can flexibly create diverse contents, many companies create AR advertising for their products. Promotions with AR advertising range from those that can be experienced at home to those that are performed at stores. On the other hand, when a consumer purchases a product packaged in a box or bag, there is a problem that the volume of its contents the consumer expects differs from the actual volume. Because the consumer cannot see the contents in the packaged product, the consumer relies on information on the product's package, such as weight. However, it is difficult for consumers to estimate the volume of the invisible contents in the packaged product based on such information when they purchase the product for the first time. The purpose of this study is to help consumers estimate the volume of the contents in the packaged product that they cannot see at the time of purchase. We develop AR advertising with a virtual model of the invisible contents in the packaged product that makes consumers perceive the sense of their volume.

An Approach to Metric Rectification Using Sequential Estimation of Spatial Adjacent Planes Appeared in Single View Image

Haruka KUROSE[1], Akio KIMURA[2]
[1]Division of Science and Engineering, Graduate School of Arts and Sciences, Iwate University
[2]Faculty of Science and Engineering Iwate University

Email: krs@mips.cis.iwate-u.ac.jp, kimura@cis.iwate-u.ac.jp

Abstract — In this paper, we propose a metric (Euclidean) rectification method for a target spatial plane showed up in a single viewpoint image. Our approach is realized from homography matrix calculated without using four points correspondence, unlike general well-known metric rectification methods. The method first estimates projection matrix for the pre-defined world coordinates system on a specific plane regarded as the base plane, and then sequentially estimates adjacent spatial planes appeared in the input image using the method of [1]. Once our method can estimate the target plane, the world coordinates system on the base plane is moved to on the estimated target plane, and it is regarded as the new world coordinates system. By doing this, it can be showed that the homography matrix for the target plane is calculated only from the original projection matrix and the transformation parameters between the old and the new world coordinates systems (rotation and translation). Finally, metric rectification can be realized using the calculated homography matrix. That is, the proposed method can obtain 3D information such as length and angle on a plane where four points correspondence is not easily determined, only using a single view image. Experimental results showed the validity and effectiveness of our approach.

Robotic Motion Generation for Realization of the Target Task Using Function and Poses of Objects

Takuma Terasawa[1], Kentaro Kozai[1], Manabu Hashimoto[1]
[1]Chukyo University

Email: terasawa@isl.sist.chukyo-u.ac.jp, mana@isl.sist.chukyo-u.ac.jp

Abstract – Everyday objects vary in shape and size, even if they are of the same type (e.g., spoons and cups), and they are normally placed in various poses (e.g., face up or face down). Robots must decide on robust motion parameter to handle everyday objects of various shapes, sizes and poses. Motion parameter is necessary information to generate robot motion, such as the approach position and the approach direction of the robot hand to the object. Previously, motion parameter was decided by the following two methods. In the first method, motion parameter is decided by comparing the object with a 3D model with set in advance motion parameter. This method has a problem of increase development cost because it is necessary to set motion parameter to each object. In the second method, motion parameter is decided by using CNN with RGB image. This method has a problem that it is difficult to recognize the poses of the object. In this research, we propose a method to decide on motion parameter to execute target tasks such as "scooping powdered tea and putting it into a teacup", by using no information about the shape, size and poses of each object. Our method uses 3D-DNN techniques to recognize an objects' function (e.g., "scoop" and "grasp") and poses (e.g., face up or face down). The robotic motion is generated on the basis of the motion parameter decided by using the recognition results. In the experiment, we evaluated the performance of the method by testing it on five spoons of various shapes and sizes. The system had a success rate of approximately 86%.

Method for Recognizing Objects of Unknown Size Using Surface Primitives

Takahiro Suzuki[1], Kentaro Kozai[1], Shuichi Akizuki[1], Manabu Hashimoto[1]
[1]Chukyo University
Email: suzuki@isl.sist.chukyo-u.ac.jp, mana@isl.sist.chukyo-u.ac.jp

Abstract –In this paper, we propose a method for recognizing items of different sizes sold in convenience stores. In recent years, the need for robots to operate at stores instead of people is increasing. Such robots are required to estimate the 6DoF pose of more than 2,000 types of items in order to display and dispose of them. Previous methods for 6DoF pose estimation use 3D-CAD models of objects. However, these methods are high in cost because more than 2,000 models of items need to be prepared. In this paper, we utilize their shapes, that is, “cuboid”, “isosceles triangular prism”, “cylinder” and “regular triangular prism”. The method consists of a module that generates candidates for the shape and size of an object, and a module that extracts an optimal hypothesis from among hypotheses. First, in the method, many hypotheses of various sizes and 6DoF pose are generated using the numbers of types of surfaces of each shape and the positional relationship between the surfaces. Second, the sizes and 6DoF pose of items are defined in the hypotheses are evaluated in depth images. Finally, the best hypothesis of the evaluation is determined as the optimal hypothesis. In the evaluation experiment of proposed method, we generated 100 objects of different sizes in a virtual space and applied this method to them. In result of the experiment, the recognition rate of isosceles triangular prisms was 84%, that of cuboids 93%, and that of cylinders 94%. Thus, the items were recognized without using 3D-CAD models.

Extrinsic Parameters Calibration of Multiple Fisheye Cameras in Manhattan Worlds

Weijie Chen[1], Sarthak Pathak[1], Alessandro Moro[1], Atsushi Yamashita[1], and Hajime Asama[1]
[1]Department of Precision Engineering, The University of Tokyo, Tokyo, Japan

Email: {chen, parthak, moro, yamashita, asama}@robot.t.u-tokyo.ac.jp

Abstract –With the advantage of having a large field of view, fisheye cameras are widely used in many applications. In order to generate a precise view, calibration of the fisheye cameras is very important. In this paper, we propose a method of extrinsic parameters calibration of multiple fisheye cameras working in man-made structures. A Manhattan Worlds space assumption is used, which describes man-made structures as sets of planes that are either orthogonal or parallel to each other. The orientation of the cameras is obtained by extracting vanishing points that denote orthogonal principal directions in different images captured by the cameras at the same time. With the proposed method, the calibration of extrinsic parameters is very convenient and the system can be recalibrated remotely.

Measurement of Abdominal Shape by Sampling Moiré Method and Extension to Video Processing

Yuki Mochizuki[1], Norio Tagawa[1], Kann Okubo[1]
[1]Tokyo Metropolitan University, Department of Electrical Engineering and Computer Science, Tokyo, Japan

Email: yu.mochizuki.dev@gmail.com, tagawa@tmu.ac.jp, kanne@tmu.ac.jp

Abstract – To detect abnormal breathing in a patient, it is necessary to take an image of the patient's chest and abdomen and analyze their movement. The ultimate goal of this study is to develop a system that determines the triage level by measuring the 3D deformation of the patient's abdomen and chest in real time and analyzing the results. This time, considering that there is little texture information in the abdomen and chest, we focused on Moiré analysis that is effective for 3D measurement in such cases. In this research, dynamic 3D measurement using image sequences is necessary. Therefore, the sampling Moiré method, which is often used in recent years because of high calculation efficiency, was adopted. In the Moiré analysis, it is necessary to determine the absolute fringe order of the Moiré fringes by another method. In static measurement, the distance to the reference fringe can be measured in advance, but it is difficult in dynamic measurement. Therefore, we propose a method of projecting spot light together with slit light for Moiré analysis and using the dynamic stereo method together. It can be adaptively adjusted for each frame so that the spot light irradiation position on the object is on the reference stripe. The depth distance to that position can be calculated using the dynamic stereo method, so 3D measurements can be performed independently for each frame. The effectiveness and accuracy of this method were confirmed through experiments on real image sequences.

A Picking Interface Seamlessly Connecting Passive and Active Performance on Guitar

Nozomu Yoshida[1], Katsutsugu Matsuyama[1]
[1]Iwate University, Faculty of Science and Engineering, Iwate, Japan

Email: s0616063@iwate-u.ac.jp, m18u@iwate-u.ac.jp

Abstract – We have designed an interface that allows a guitar player to manipulate another channel, such as the manipulation of visual effects, while playing the guitar. This interface connects passive and active performance seamlessly. For example, when the performer is NOT playing the guitar, he/she can switch image effects by pressing the pick with various degrees of forces or by shaking his/her wrist (active). On the other hand, when playing the guitar, the picking motion itself is sensed to manipulate the parameter of the effect (passive). Our goal is to create an interface that immediately reflects the performer's intentions without turning away from the guitar playing motion.

We propose a method to distinguish 3 picking states (Chord stroke/Single note/Non-picking) by sensing pressure on the pick and the rotation angle of the player's index finger. For our goal, recognizing whether picking occurs or not is needed firstly. We think it is also useful to recognize playing style, i.e. stroking a chord or picking a single note. Our interface consists of a pressure sensor and an acceleration sensor. The pressure sensor is fixed onto a pick. It senses pressure force between thumb and index finger of the player's picking hand. The acceleration sensor is fixed on the first knuckle joint of the index finger. By combining these two factors: pressure and rotation angle, we judge the current picking state. Our brief experiment shows that our method can correctly identify the picking state with high accuracy.

Finally, we introduce an application example which manipulates visual and sound effects through our interface. For example, in the non-picking state, the performer can manipulate CG sphere effects by associating the pressure with the expansion rate, and the rotation angle with the colorfulness.

Examination of Group Head Angle Acceleration Analysis Method for Learning Evaluation in Outdoor Education

Kazunari haruta[1], Ryo Sakurai[2], Kazuhiko W. Nakamura[3], Takayuki Nakata[1], Kousaku Hashimoto[1]

[1]Toyama Prefectural University, Information System Engineering, Toyama, Japan
[2]Ritsumeikan University, Policy Science, Kyoto, Japan
[3]University of Tokyo, Agricultural and Life Sciences. Tokyo, Japan

Email: t955012@st.pu-toyama.ac.jp, ryosak@fc.ritsumei.ac.jp, k_nakamura@csis.u-tokyo.ac.jp, nakata@pu-toyama.ac.jp, t955011@st.pu-toyama.ac.jp

Abstract – In the field of outdoor education, education is evaluated by subjective evaluation. However, these methods have a problem that human subjectivity is mixed. Therefore, there is an attempt to evaluate learning from sensor's data. Since it can be considered that the moment the learner moves his head during the lecture has important information, we measure the concentration of the learner using the angular velocity information of the head. When a learner performs head movements, it cannot clear form only one learner's information whether the cause of the movement is concentrated or not. Therefore, we propose a method for discriminating the concentration of learners by comparing the moment of head movements in the learning group. In order to confirm the effectiveness of the proposed method, we measured the angular velocity of the head in outdoor education by attaching an angular velocity sensor to the learner's head. In the lecture induced gaze direction, the head movement timing of many leaners coincided after gaze induction timing. In addition, we found that the concentrated learners did not move their heads except for the gaze induction timing.

Accuracy Enhancement in Intra- and Inter-Frame Example Search for Lossless Video Coding

Koji Nemoto[1], Yusuke Kameda[1], Ichiro Matsuda[1], Susumu Itoh[1], Kyohei Unno[1, 2], Sei Naito[2]

[1] Department of Electrical Engineering, Tokyo University of Science
2641 Yamazaki Noda-shi, Chiba 278-8510, JAPAN

[2] KDDI Research, Inc.

Email: nemoto@itohws01.ee.noda.tus.ac.jp

Abstract – Most efficient lossless image coding techniques consist of two processing stages: decorrelation by prediction and entropy coding of the prediction residuals. On the other hand, we previously proposed a lossless coding method which seamlessly integrates both the stages for lossless video coding. In this method, several examples, i.e. a set of pels whose neighborhoods are similar to a local texture of the target pel to be encoded, are searched by template matching from already encoded areas of the current and previous frames. Then the probability distribution of video signals is directly modeled as a weighted sum of the Gaussian functions whose center positions are given by the individual examples on a pel-by-pel basis. Furthermore, model parameters that control shapes of the Gaussian functions are defined as parametric functions of local feature values including the matching cost, and optimized so that the resulting coding rate can be a minimum. In this paper, the above example search process is enhanced from integer- to fractional-pel accuracy for more reliable probability modeling. Specifically, we perform the parabola fitting technique with respect to the matching costs around the integer position obtained by the template matching. For calculating the estimation video signals at the resulting fractional-pel position, bicubic interpolation is used. Simulation results indicate that coding rates are lowered for most sequences by introducing the above accuracy enhancement. As a result, our coding method outperforms the H.265/HECV-based lossless video coding scheme by 1.5 % in averaged coding rates.

A MR-based Visualization System of IoT Security

Yuma Akasaka[1], Mitsunori Makino[2]
[1.2]Department of Information and System Engineering
Chuo University
Tokyo, Japan

Email: a14.4r5t@g.chuo-u.ac.jp, makino@m.ieice.org

Abstract – IoT devices increases rapidly. In addition, there are more opportunities for users who do not understand security status of Internet of Things (IoT) devices to use these. Therefore, this article proposes a visualization system with mixed reality (MR) technology, which helps users understand security status of IoT devices around them. The proposed system introduces an image recognition technology into obtaining type and location of the device. Based on both of location of the devices and the relevant security information stored in database, the proposed system overlays color as security level and messages of recommended action on the devices. If the user reacts correctly, the reactions are stored as the action history for the device in the system. Then, the proposed system updates the security status of the devices based on the action history and the relevant security information stored in database. Installed on a head-mounted-display (HMD), the proposed system expresses appropriate security status of IoT devices more intuitively than the conventional visualization such as log-text or graph. Also the system can express the vulnerabilities related to positional relation of the devices. Such intuitive and position-related representation of the security is expected even for non-expert to act adequately.

Haze Removal in Outdoor Images

Ryo Akamatsu, Yuta Muraki, Koji Nishio, Kenichi Kobori
Osaka Institute of Technology, Graduate School of Information Science and Technology, Osaka, Japan

Email: m1m18a01@st.oit.ac.jp, yuta.muraki@oit.ac.jp, koji.a.nishio@oit.ac.jp, kenichi.kobori@oit.ac.jp

Abstract – Digital cameras are used in various scenarios; however, the sharpness of images captured outdoors may be reduced due to bad weather conditions such as fog and haze. Therefore, to obtain a clear image, it is essential to remove haze. In this study, we propose a haze removal method by separating the sky and foreground regions and applying different processes to the sky region because it has different features from the foreground region; this improves the visibility of the image. We assume that the sky region is a bright region that does not change much throughout the image and extract multiple sky region candidates, which are merged according to color distance. Next, we estimate atmospheric light and transmittance of haze. Atmospheric light is the light scattered in the entire image, and transmittance of haze is the amount of scattered light. The sky region determines the brightness of atmospheric light, and the impression of the entire image determines the color of the atmospheric light. Transmittance of haze is estimated by dark channel features and morphological operations. The conventional method uses a fixed-sized patch due to which a smooth transmission map may not be generated. Our method generates a smooth transmission map for any image by changing the patch size. The haze in the image is removed using the estimated atmospheric light and transmittance; however, the resulting image is dark for which brightness correction is performed and a clean image without haze is obtained.

Improvement of Accuracy of Wide-surrounded Multi-projection in Indoor Spaces

Kentaro Miura, Naoki Hashimoto
The University of Electro-Communications, Tokyo, Japan

Email: miura@ims.cs.uec.ac.jp, naoki@cs.uec.ac.jp

Abstract – With the spread of projection mapping in recent years, image production for a wide area of a theater or a museum is performed using multiple projectors. In order to achieve image projection in such various places, which are not ideal screens, it is necessary to grasp the shape of the space. In addition, because of the large space for the projection, the necessary equipment for the acquisition will become huge. Therefore, in this research, we aim to improve the accuracy of wide area projection using multiple projectors in an environment surrounded by walls such as indoor spaces. In previous efforts, we have realized a method for large-scale image projection easily, by projecting a pattern onto the indoor space and measuring with a combination of a normal lens and a fish-eye lens. However, when a pattern is projected in an indoor space surrounded by walls, an area in which measurement cannot be performed correctly by indirect reflected light occurs, which causes distortion in the projection result. In this paper, the influence of indirect reflected light is reduced by dividing the projected pattern image. Focusing on the fact that indirect reflected light loses high-frequency components in its generation process, it is greatly reduced by a simple difference process based on the similarity of indirect reflected light of a sufficiently divided pattern image. On the other hand, such division of the pattern image increases the time required for grasping the shape of the space. Therefore, the divided patterns are designed to reduce the number of them as much as possible while effectively removing indirect reflected light. In this way, both high accuracy and short setup time are achieved for wide area projection in indoor spaces.

A Study on Liver Tumor Detection from an Ultrasound Image using Deep Learning

Takahiro Nakashima†, Issei Tsutsumi†, Hiroki Takami†, Keisuke Doman†, Yoshito Mekada†, Naoshi Nishida‡, Masatoshi Kudo‡
† School of Engineering, Chukyo University, Toyota, Japan
‡ Faculty of Medicine, Kindai University, Osaka, Japan

Email: {nakashima.t, tsutsumi.i, takami.h}@md.sist.chukyo-u.ac.jp, {kdoman, y-mekada}@sist.chukyo-u.ac.jp

Abstract – The difficulty of ultrasound examination is high because the doctor operates equipment and interpret images at the same time. Therefore, it's concerned that some tumor is overlooked. To prevent overlooking tumors, here we study a liver tumor detection method using convolutional neural networks toward realizing computer-assisted diagnosis systems. In this report, we propose a liver tumor detection method and false-positive reduction method. We used YOLOv3 as the object detection algorithm in order to finding candidate regions, and then we used ResNeXt for reducing false positives. In the results using YOLOv3 and ResNeXt, the recall for cysts was relatively high (0.88), whereas that for metastatic liver cancer was low (0.36). Experimental results showed that our method could detect lesions that frequently appeared in training data. Future work includes increasing data from more hospitals and using them to improve the detection accuracy.

Texture Reconstruction Based on Underlying Pattern Modification that Reflects User's Intention

Mami Nagoya, Tomoaki Kimura and Hiroyuki Tsuji
Graduate School of Engineering, Information and Computer Sciences, Kanagawa Institute of Technology,
Atsugi, Kanagawa, Japan

Abstract – When creating a CG, texture images are used to express textures of various objects' surfaces. Texture image samples are often available on commercial texture collections or websites. Although it is relatively easy to find a texture image similar to the designer's idea, it is difficult to obtain what the designer exactly wants, that is, the texture image which has both ideal patterns and ideal textures. Additionally, it is even more difficult to create texture images by oneself. The purpose of this study is to propose an intuitive method for reconstructing an ideal texture image by editing the available texture image at hand. In general, most of texture images provided as samples are composed of a certain basic pattern and a fine texture. In such images, it is difficult to separate textures from basic patterns and to modify only the underlying patterns. In this paper, we propose a method to achieve such difficult tasks by using deep learning. The proposed method is a deep neural network that improves pix2pix which can learn a conversion process between input and output images, but is weak in keeping the output texture as you want. For this reason, we add some restriction to the proposed method, so that the learned process should not change the original texture of the input image. The effectiveness of our modification is verified through several computational experiments, by using many texture images with underlying patterns and fine textures as a test data set. As a result, the generated image is able to keep the input color and edge information. However, the generated image is blurred and the input texture information cannot be reproduced in some cases. These problems should be improved in near future.

A Tuna Dismantling Education System in Virtual Reality

Hikaru Yokoyama, Mitsunori Makino
Department of Information and System Engineering
Chou University
1-13-27 Kasuga, Bunkyo-ku, Tokyo 112-8551, Japan
a14.4a57@g.chuo-u.ac.jp, makino@m.ieice.org

Abstract – This article proposes a virtual reality (VR) system To Equip knowledge and skill of the tuna dismantling, they have to learn it in hands-on style as well as in conventional lecture and self-study styles. However, high cost of tuna gives them less chances of the practical work. Therefore, a VR based self-study system is proposed as a supplemental tool, in which the students can study and practice dismantling the tuna in a virtual space. The proposed system consists of PC, a HMD for VR and controllers for both hands. One controller is assumed to be a cutting knife, and the another is used for tracking the position of the other hand. Representing a whole virtual tuna in front of a user having the two controllers, the system encourages the user to cut the virtual tuna with the controller as the cutting knife. From position and direction of the knife when the knife is cut in, the system determines the correctness of the cutting. If the cutting is within normal range, the system represents the cutting and encourage the user to proceed. If not, the system give caution by message and/or vibration to the user and encourages him/her to re-cut in. The system has another mode called "test mode." In the test mode, the system gives no caution/hints to the user, while the system determines the correctness of the user's action (cutting). The accumulated determination is converted to score, so that the user can easily understand his/her level of achievement. Thus, the system is expected to be effective as a supplemental and continuous study tool.

Pitching Form Evaluation Based on Elbow Position by a Monocular Camera

Naoto Nohara[1], Hidehiko Shishido[2], Itaru Kitahara[2], Yoshinari Kameda[2]
[1]University of Tsukuba, Graduate School of System and Information Engineering, Ibaraki, Japan
[2]University of Tsukuba, Center for Computational Science, Ibaraki, Japan

Email: {nohara.naoto, shishido.hidehiko, kitahara.itaru, kameda.yoshinari}@image.iit.tsukuba.ac.jp

Abstract – Pitching in the correct form is essential for preventing injury and improving skills. It is not easy for athletes and instructors to check whether a pitcher is throwing in the correct form. In this study, we record a pitcher from the direction of the catcher by a monocular camera and estimate the skeleton pose of the pitcher by using OpenPose. We propose a new method to evaluate whether the pitcher can pitch in the correct form by examining the estimation result. We use SSE(Shoulder, Shoulder, Elbow)-line as an evaluation index. When the upper body of the pitcher faces the batter, it is desirable that the SSE-line be straight.In our preliminary study, we took videos of a pitcher at 30 fps at Full-HD resolution. The participant was a male high school student who belongs to the baseball club. Since the movement of the pitcher was fast, it was found that it was difficult to find the right frame at which the pitcher body turns squarely to the batter. So we have introduced a 240 fps camera and analyze the relationship between the SSE-line which is estimated from the pose and the pitcher body rotation over time. It was found that 240 fps videos are valid to find the right frame at which the upper body of the pitcher faces the batter.

Detection and Motion Analysis of Knee Joint in 3-D Point Cloud Data Measured by Depth Camera

Wataru Morozumi, Norio Ragawa, Kan Okubo, Hiromichi Fujie
Tokyo Metropolitan University, Department of Electrical Engineering and Computer Science, Hino, Tokyo

Email: wataru.newpc@gmail.com, tagawa@tmu.ac.jp, kanne@tmu.ac.jp, fujie@tmu.ac.jp

Abstract – There are many things to be improved in total knee arthroplasty. This is because patients are often dissatisfied with the post-operative course and require high medical costs. Therefore, a "3D joint instability analysis system" that generates basic data for navigation in total knee arthroplasty has been proposed. This system performs instability analysis by associating "knee joint movement detected from images" with "weight / moment measured by 6-axis force sensor." In this study, we focused on measuring three-dimensional displacement of the knee joint using images. In order to acquire 3D knee joint data, we used an Intel RealSense depth camera, which has a built-in depth sensor and can get depth information dynamically. In the experiment, using the RealSense depth camera, template data of only the knee joint part of the patient was acquire, and the shape similar to the acquired template data was searched from the 3D point cloud data of the entire leg to detect the knee joint. Using two algorithms, RANSAC and ICP, template shape detection was performed with high efficiency and high accuracy. Next, we examined the measurement of knee joint angles. By setting the knee joint detection result as the center position of the 3D point cloud of the entire leg, the point cloud data can be divided into thigh part and shin part. The shape of the cylinder created by each obtained point group was determined, the angle of two degrees of freedom formed by the axial direction of the two cylinders was calculated, and used as the measurement result of the knee joint angle. Experiments using real images confirmed that our method has sufficient accuracy for this navigation system.

A Museum Video Guide Creation System with CG Made from Scripts on the Server

Akira Asada[1], Masaki Hayashi [2], Makoto J. Hirayama [1]
[1] Faculty of Information Science and Technology , Osaka Institute of Technology , Osaka , Japan
[2] Department of Game Design , Uppsala University , Gotland, Sweden
Email: asdakr0830@gmail.com, hayasim77@gmail.com, makoto.hirayama@oit.ac.jp

Abstract – Audio guides are used in many museums, and recently, some museums start providing video guides for visitors. However, the video guide need video production that requires a lot of manpower. We have been studying and developing a system which enables users to produce TV-like art programs made by computer graphics in a way of writing simple scripts instantly. We use TVML (TV program Making Language) to create CG animations. Based on the TVML technology, we have also developed the system which converts internet-based blog posts to CG animations automatically. Our video guide creation system is achieved by combining those technologies to construct a web-based system that enables creators to upload art program scripts to the server and allows museum visitors to access the scripts for play back on their smartphones.

In our system, movie guide creators write a script on a special web site constructed by Wordpress and upload the script simply pressing the upload button, then the script is uploaded to the server. The reason to use Wordpress is that it is already established article posting system to be used as an editing system for our purpose. On the other hand, we developed a special player application designed for play back of the script on the server is used by museum visitors to visualize the scenario. It allows the visitors to access the script with a smartphone app and watch it as a CG animation on the spot. One of the advantages of our method is that guide movies can be easily modified and updated by the creators.

As for the next step, we plan to make our museum guide creation system open to the public, and will allow many people to create and register movie guide scripts voluntarily just like the Wikipedia.

Performance Analysis on Prediction Structure for Multi-view-based Light Field Video Coding

Thuc Nguyen Huu, Vinh Van Duong, and Byeungwoo Jeon
Department of Electrical and Computer Engineering

Email: thuckechsu@skku.edu, duongvinh@skku.edu, bjeon@skku.edu

Abstract – Light Field (LF) image/video data provides both spatial and angular information of scene but at the cost of tremendous data volume for their storage and transmission. At the moment, the Multi-view Video Coding (MVC) is one of promising compression solutions for LF video data, so it deserves more investigation regarding its good prediction structure to effectively reduce the redundancy in LF video data. Several prediction structures have been investigated but only with limited experimental evaluations due to the lack of dataset and different testing configurations. This practical problem can be mitigated now by new datasets and common test condition for LF Video recently proposed by MPEG. As a first step for the design of a good compression method for LF video data, in this paper, we re-evaluate the performance of existing prediction structures for MVC-based LF video coding methods following the MPEG common test condition and its dataset.

Comparison of Data Costs for Depth Estimation from Compressed Light Field Images

Vinh Van Duong, Thuc Nguyen Huu, and Byeungwoo Jeon
Department of Electrical and Computer Engineering, Sungkyunkwan University

Email: duongvinh@skku.edu, thuckechsu@skku.edu, bjeon@skku.edu

Abstract – Depth estimation is one of the core applications of light field (LF) image/video data and there are numerous algorithms developed for it. While compression is inevitably required for light field image data in practice due to its huge data amount, most depth estimation methods have not yet paid sufficient attention to the compression effect on it. Thus, we investigate various light field depth estimation methods to design more robust and better compression method of light field data in the context of depth estimation. By noting that building data cost is a very first step in most depth estimation algorithms, and the data cost computation is a crucial stage that significantly affects the accuracy of the disparity map, in this paper, we present an in-depth analysis of data cost computation in light field depth estimation problem in the context of compression. Experimental results show that the data cost building on epipolar plane image (EPI) outperforms other methods in most evaluation criteria.

Acquisition of Wiping Area Using SLAM for Visualization of Cleaning Area

Yoneda Wataru, Yoshitsugu Manabe, Noriko Yata
Chiba University, Chiba, Japan

Email: aeca2631@chiba-u.jp, manabe@faculty.chiba-u.jp, yata@chiba-u.jp

Abstract—Cleaning is inseparable in life, but it is impossible to see with the naked eye where the room was actually cleaned. For this reason, if information on the location where the cleaning was performed cannot be shared when cleaning by multiple people, there is a possibility that an unclean area is remained. Therefore, if Augmented Reality (AR) can be used to visualize the passing area of the hand or cleaning tool being cleaned, it will lead to improve cleaning efficiency and increase motivation by visualizing the cleaning area. The purpose of this research is to obtain and superimpose the location information of the passing area using Simultaneous Localization and Mapping (SLAM) in order to visualize the passing area of the hand or the cleaning tool using AR.

Head Orientation Detection With Small Camera In Outdoor Education Using Background Images

Kosaku Hashimoto[1], Kazunari Haruta[1], Kantaro Tabiraki[2], Kazuhiko W. Nakamura[3], Ryo Sakurai[4], Takayuki Nakata[1]
[1]Toyama Prefectural University, Information Systems Engineering, Toyama, Japan
[2]Matsumoto University, Department of Tourism and Hospitality, Nagano, Japan
[3]The University of Tokyo, Graduate School of Agricultural and Life Science, Tokyo, Japan
[4]Ritsumeikan University, College of Policy Science, Osaka, Japan

Email: t955011@st.pu-toyama.ac.jp, kantaro.tabiraki@t.matsu.ac.jp, nakamura@fr.a.u-tokyo.ac.jp, ryosaku@fc.ritsumei.ac.jp, nakata@pu-toyama.ac.jp

Abstract – In the outdoor education, the educational evaluation generally has been performed by using questionnaires or observation by evaluators. However, these methods require a lot of manpower and time to obtain the result of the assessment. We assume that the evaluation of education can be performed from gaze information such as an attention object and fixation time. In this paper, to estimate the learner's gaze in the learning activity, we propose the method that detect the learner's head orientation time-sequentially by calculation the position and the posture of the small camera on the head by Structure from Motion (SfM). To match feature points that extracted from only stationary objects, this method requires background images besides head camera images. The experimental result is obtained from the video be recorded in the actual educational site. Input head mount camera images are 16 images that be extracted from a part of the video that be kept recording while the learning activity. Also, background images are 781 images without the moving subject such as human. These are taken in the place same as the educational site after the learning activity. Among 16 head mount camera images that extracted from the video every second, 13 images are detected with the mostly correct posture.

Moving Obstacle Tracking and Estimation on Crosswalk for Blind-People Navigation System

Kongpob Maksap, Kosin Chamnongthai
Department of Electronic and Telecommunication Engineering
King Mongkut's University of Technology Thonburi
Bangkok, Thailand

Email: kongpob.maksap@mail.kmutt.ac.th, kosin.cha@kmutt.ac.th

Abstract – When a blind person starts to cross a crosswalk, he/she may need to track moving objects such as pedestrians, pets, and so on in in order to avoid collision. This paper proposes a method of moving-obstacle tracking on a crosswalk for blind-person navigation system. In the method, borders of a crosswalk in front of the blind person is first determined using straight white lines on an intensity image, and a depth image is simultaneously used to detect candidate moving obstacles using differences among neighboring frames. The moving objects are then determined by height of the candidates, and positions of moving objects, which are assumed to be pedestrians and pets, are obtained from the depth image. This position information on previous image frames is finally used to estimate movement vector based on the moving object trajectory, and rectangular windows are employed to track moving objects in the next frame. To evaluate performance of the proposed method, experiments with pedestrians at a crosswalk were performed, and results showed effectiveness of the proposed method.

Lossless Coding of HDR Color Images in a Floating Point Format Using Block-Adaptive Prediction with Exponent Equalization

Yuya Kamataki, Yusuke Kameda, Ichiro Matsuda, Susumu Itoh
Department of Electrical Engineering, Tokyo University of Science
2641 Yamazaki Noda-shi, Chiba 278-8510, JAPAN

Email: 7319521@ed.tus.ac.jp

Abstract – Recently, with the advancement of image sensors and display devices, high dynamic range (HDR) images are becoming much popular. The Radiance RGBE format, which represents RGB components by floating point numbers, is widely used for recording HDR image contents. In this format, three mantissa parts of R, G and B components as well as a common exponent part E are represented in 8-bit depth, respectively. Thus, it provides practically enough dynamic range for color information in a relatively simple data structure (32 bits per pel). To compress image data in this format, we formerly applied the block-adaptive prediction technique independently to each mantissa or exponent part by regarding it as a grayscale image. In this preliminary study, we found that different exponent values between adjacent pels lowered a spatial correlation and resulted in poor coding performances for the mantissa parts. To cope with this problem, we already proposed the Exponent Equalization (EE) technique which adjusts values of reference pels (mantissa part) used in the prediction so that their exponent part can be regarded as same as that of the target pel to be predicted. In this paper, we integrated the EE technique into Inter-Color Prediction (ICP) which exploits not only the same color but also other color information as far as they are already encoded. Simulation results indicate that the combination of both the EE and ICP techniques can considerably improve the coding performance.

Control Method of Pseudo-Force Intensity by Voltage Change using Software Signal

Tsukasa Asano, Yusuke Kazashi, Takayuki Nakata
Toyama Prefectural University, Information Systems Engineering, Toyama, Japan

Email: t955002@st.pu-toyama.ac.jp, nakata@pu-toyama.ac.jp

Abstract - In this study, we propose a control method of pseudo-force intensity using software signal for pseudo-force device. Pseudo-force is defined as tractive force that does not act in reality but is perceived. We perceive this force by holding the vibrator that is asymmetrically vibrating. The previous study showed that asymmetric-vibrating vibrator changes pseudo-force intensity by changing voltage. By realizing this method using the only digital signal, it is possible to save changing supply voltage directly. We change voltage by PWM control, and try to control pseudo-force intensity. PWM controls output voltage by switching applied voltage for every constant period. In this study, we use the pulse signal for pseudo-force presentation that include the other for changing voltage. Pseudo-force intensity was evaluated as weight's tractive force. The vibrator and weights are connected by a wire, and a subject hold the vibrator and experiences pseudo-force from it; additionally, this force presents to the direction of pulling weights. On this condition, we change duty ratio of the pulse signal for changing voltage. The result showed that pseudo-force intensity increases as high frequency pulse duty ratio increases. We confirmed the effectiveness of PWM controlling to control pseudo-force intensity by voltage change.

Efficient Bin Allocation for Chroma Intra Mode Coding

Jeeyoon Park and Byeungwoo Jeon
Department of Electrical and Computer Engineering, Sungkyunkwan University, Suwon, Korea

Email: {jiyoonpark, bjeon}@skku.edu

Abstract –The Versatile Video Coding (VVC) is a new state-of-the art video compression technology that is being under standardization. It targets for about two times higher coding efficiency against the existing HEVC, supporting HD/UHD/8K video, and 10 bit/high dynamic range (HDR) video. It also targets for versatilities such as screen content coding, adaptive resolution change, and independent sub-pictures. However, in return, its encoding and decoding complexity has been noted to increase significantly. To develop an effective coding method for chroma intra prediction mode, in this paper, we investigate its binarization process in CABAC (context adaptive binary arithmetic coding) and test a method which assigns shorter bins to more frequent chroma intra modes and longer bins to the less frequent ones based on the its mode statistics.

Evaluation for Harmonic Location Estimation System of Image Retrieval and SLAM

Kohei Yamasaki[1], Shishido Hidehiko[2], Itaru Kitahara[2], Kameda Yoshinari[2]
[1] University of Tsukuba, Grad, School of System and Information Engineering, Ibaraki, Japan
[2] University of Tsukuba, Center for Computational Science, Ibaraki, Japan

Email: yamasaki.kohei @image.iit.tsukuba.ac.jp, {shishido.hidehiko, kitahara.itaru,,kameda.yoshinari}@image.iit.tsukuba.ac.jp

Abstract – Robustness of pedestrian localization is one of the important issues to realize a reliable navigation system for pedestrians in daily scenes. A new pedestrian localization system that utilizes both the image-retrieval approach we have recently developed and a SLAM (Simultaneous Localization and Mapping) approach is proposed in this paper. It relies on a single camera as a sensor and estimates the location by computer vision technology. The problem here is that high processing cost is needed to operate two approaches simultaneously and it could be impractical to run these two on a wearable computing unit. The problem is solved by running the two approaches on separate computers that are connected with a computer network. We have implemented a preliminary system that harmonizes the two approaches over two computers and measured its performance in typical daily scenes on our campus. The result is promising for further implementation.

Proposal of a Rescue Operation Support System based on 3D Reconstruction, GPS, and Digital Pen

Wei-Chung Chen, Shuuichi Ihara, Makoto Hasegawa
Tokyo Denki University, Tokyo, Japan

Email: 18kmc12@ms.dendai.ac.jp, hasegawa@mail.dendai.ac.jp

Abstract – Three-dimensional computer graphics (3DCGs) of disaster scenes can be a visualization method of the situation of disaster areas. The 3DCG created from photographs taken by victims or rescue teams is a useful tool for disaster relief. We propose a system based on the Structure from Motion (SfM) technology to reconstruct the 3D models of disaster areas. And it's also based on the attached GPS data within collected photographs taken by smartphones at the disaster sites to localize the reconstructed 3D models on the global map. We describe how it registers 3DCGs to a global map and how it estimates the positions and orientations of 3DCGs here. Moreover, users can edit the reconstructed 3D models with a digital pen and a dot screen in this system; use a projector to project 3DCGs on the dot screen. This paper shows the use of GPS information and presents the simulations that took place in the disaster experience facility "Sona Area Tokyo of The Tokyo Rinkai Disaster Prevention Park" and the "921 Earthquake Museum of Taiwan in National Museum of Natural Science."

Privacy-Preserving Machine Learning Using EtC Images

Ayana Kawamura, Yuma Kinoshita, Hitoshi Kiya
[1]Tokyo Metropolitan University, Tokyo, Japan

Email: kawamura-ayana@ed.tmu.ac.jp, kinoshita-yuma@ed.tmu.ac.jp, kiya@tmu.ac.jp

Abstract – In this paper, we propose a novel privacy-preserving machine learning scheme with encrypted images, called EtC (Encryption-then-Compression) images. Using machine learning algorithms in cloud environments has been spreading in many fields. However, there are serious issues with it for end users, due to semi-trusted cloud providers. Accordingly, we propose using EtC images, which have been proposed for EtC systems with JPEG compression. In this paper, a novel property of EtC images is considered under the use of z-score normalization. It is demonstrated that the use of EtC images allows us not only to protect visual information of images, but also to preserve both the Euclidean distance and the inner product between vectors. In addition, dimensionality reduction is shown to can be applied to EtC images for fast and accurate matching. In an experiment, the proposed scheme is applied to a facial recognition algorithm with classifiers for confirming the effectiveness of the scheme under the use of support vector machine (SVM) with the kernel trick.

HEVC Intra Prediction Mode Classification by Deep Learning

Han Xing Wang, Yan Fen Li , Lien Minh Dang, Hae Kwang Kim
Sejong University, Department of Computer Engineering, Seoul, Korea

Email: hanxiang@sju.ac.kr,1826535091@sju.ac.kr, minhdl@sju.ac.kr , pfonseca@ipma.pt, hkkim@sejong.ac.kr

Abstract – In High Efficiency Video Coding (HEVC) standard, the best intra prediction mode is decided by choosing the smallest rate-distortion cost of actual encoding among the total of 35 modes with the MPM (Most Probable Mode) scheme for compression purpose of mode encoding with reference to the adjacent reference blocks of the current prediction unit. This causes to heavy computational complexity. In this paper, a deep neural network is conceived and experimented as a probable module for the intra prediction mode decision process inside of the HEVC encoding scheme. The neural network is trained and tested with a ground-truth data set constructed from actual HEVC Intra encoding of original images. For the performance of the test, accuracy is used as the percentage of the correct mode output by the designed neural network to the ground-truth mode. the experimental results show that the neural network does not give high accuracy for the correct mode. However, accuracy increased when neighboring similar angle modes are considered as the correct mode. Also, the special modes of DC and Planar for MPM are analyzed in this paper.

A Display Panel System using Spotlight Projector Mapping for Museum Exhibition

Makoto J. Hirayama[1]
[1] Information Science and Technology, Osaka Institute of Technology, Osaka, Japan

Email: makoto.hirayama@oit.ac.jp

Abstract – In museums, title, artist, year, and additional explanations are put on a small panel near each exhibit and large poster panels are placed on a wall for overall explanations of the exhibition, traditionally. Sometimes, video contents are played using a display monitor for more detailed lectures about these exhibits. Or, an audio guide using a small audio player with an earphone can be rent at an entrance. Nowadays, video guides may be available using a handy video terminal or smartphones. To present flexible information around a showcase, a method, which uses projection mapping onto showcase floor, back wall, and front glass, is proposed. A half transparent film is pasted onto the front glass of a prototype showcase which is made for testing purposes. Explanation contents include texts, figures, and pointers to exhibit objects, and they are changed in time like slide show manner. These flexible display contents are also useful to support multiple languages on demand. Contents are made using drawing and presentation software and then reshaped for mapping using graphics software. For explanation contents projection, a spotlight projector which can be used as lighting, is used. In the current prototype implementation, a spotlight projector is placed on the ceiling of the showcase. Then, projection to the backwall, projection to the front grass with transparent screen, and projection to an exhibit as lighting and explanations around the exhibit, are evaluated. By using the proposed method, visitors can watch exhibits and informative explanations simultaneously in one view.

Fringe Projection Profilometry: an Overview

Xiaoyu He[1], Qian Kemao[2]
[1]Interdisciplinary Graduate School, Nanyang Technological University, 639798, Singapore
[2]Computer Science and Engineering, Nanyang Technological University, 639798, Singapore

Email: mkmqian@ntu.edu.sg

Abstract – Getting 3D profile is required everywhere. Measuring manufactured pieces for quality control serves as one example. A recent example is FaceID, which captures the 3D face geometry for personal identification. Fringe projection profilometry, as an accurate, automatic and accelerated structured light technique, has been attractive for about half a century. We thus provide a brief overview on this technique, explaining its basic principle and the development trajectory, but emphasizing more on the following: (i) its advantages and disadvantages comparing with other sister techniques, including stereo vision, speckle projection and time of flight; (ii) existing and potential application areas; (iii) challenges in research and development; and (iv) new possibilities such as machine learning involvement. Along the explanation, some of our technical innovations on better pattern generation and processing will be introduced.

CSSNet: Image-Based Clothing Style Switch

Shao-Pin Huang[1], Der-Lor Way[2], Zen-Chung Shih[1]
[1]Institute of Multimedia Engineering, National Chiao Tung University, Hsinchu 300, Taiwan
[2]Department of New Media Art, Taipei National University of the Arts Taipei 112, Taiwan
Email: ken4797482@gmail.com, adlerway@gmail.com, zcshih@cs.nctu.edu.tw

Abstract –We propose a framework, the CSSNet to exchange the upper clothes across people with different pose, body shape and clothing. We present an approach consists of three stages. (1) Disentangling the features, such as cloth, body pose and semantic segmentation from source and target person. (2) Synthesizing realistic and high resolution target dressing style images. (3) Transfer the complex logo from source clothing to target wearing. Our proposed end-to-end neural network architecture which can generate the specific person to wear the target clothing. In addition, we also propose a post process method to recover the complex logos on network outputs which are missing or blurring. Our results display more realistic and higher quality than previous methods. Our method can also preserve cloth shape and texture simultaneously.

Super Resolution for 8K Endoscope

Seiichi Gohshi[1], Kenkichi Tanioka[2], Hiromasa Yamashita[3]
[1]Department of Informatics Kogakuin University Tokyo, [2]Medical Consortium Network Group, [3]Department of Informatics Kairos Co., Ltd

Email: gohshi@cc.kogakuin.ac.jp, taniokah@jb3.so-net.ne.jp, yamashita-h@ymail.plala.or.jp

Abstract—In endoscopic surgery, a 0.02 mm thread is used. The 0.02 mm thread is not visible when using a 4K endoscope. However, when an 8K endoscope is employed, the thread is visible on fine-focus. However, adjusting the focus of 8K by using the common tools only, such as view finders or small monitors, is very difficult. Although auto-focus functions are common, the central areas are the focus points. The central areas are not always region of interesting. A doctor is assigned to adjust the focus during the endoscopic operation. It is always very difficult to adjust the focus accurately. Super resolution (SR) has been proposed to sharpen out-of-focus images. However, a real-time SR technology is necessary for the 8K endoscope. In this study, a nonlinear signal processing super resolution (NLSP) is introduced to improve the resolution of 8K endoscope cameras. NLSP can enhance the 8K endoscope images and improve the camera's focus depth.

Self-supervised Depth Completion with Attention-Based Loss

Yingyu Wang[a], Yakun Ju[a], Muwei Jian[b], Kin-Man Lam[c], Lin Qia[a], Junyu Dong[a]
[a]Ocean University of China, Qingdao, China; [b]Shandong University of Finance and Economics, Jinan, China; [c]Hong Kong Polytechnic University, Hong Kong

Email : {qilin, dongjunyu}@ouc.edu.cn

Abstract – Deep completion which predicts dense depth from sparse depth has important applications in the fields of robotics, autonomous driving and virtual reality. It compensates for the shortcomings of low accuracy in monocular depth estimation. However, the previous deep completion works evenly processed each depth pixel and ignored the statistical properties of the depth value distribution. In this paper, we propose a self-supervised framework that can generate accurate dense depth from RGB images and sparse depth without the need for dense depth labels. We propose a novel attention-based loss that takes into account the statistical properties of the depth value distribution. We evaluate our approach on the KITTI Dataset. The experimental results show that our method achieves state-of-the-art performance. At the same time, ablation study proves that our method can effectively improve the accuracy of the results.

Semantic Scene Modeling for Aquaculture Management Using an Autonomous Drone

Hung-Yuan Chen, Shyi-Chyi Cheng, Chin-Chun Chang
National Taiwan Ocean University, Department of Computer Science and Engineering, Keelung, Taiwan

Email: xz8624550@gmail.com, csc@mail.ntou.edu.tw, cvml@mail.ntou.edu.tw

Abstract – This paper presents a semantic scene modeling technique for constructing a cloud-based aquaculture surveillance system using an autonomous drone. The emergence of low-cost drones has created opportunities to find new solutions for a number of problems of computer vision and artificial intelligence based internet-of-things (AIoT). However, vision based activity detection using a mobile RGB camera still remains as a challenging task since the activities in different regions of the scene to be monitored are quite different. Moreover, the sizes of detected objects using a drone are often very small. In this work, the 3D model of an aquaculture environment is first constructed using the calibrated intrinsic camera parameters, the depth maps and the pose parameters of frames in the captured video using a drone. Next, our semantic scene modeling algorithm represents the visual and geometrical information of the semantic objects which defines the checkpoints for routine data gathering and environmental inspection. To associate each checkpoint with the GPS signal and the altitude value of the drone, our approach combines the automatic drone navigation, computer vision and machine learning algorithms to detect the checkpoint specific activities. The scene modeling algorithm transfers the essential knowledge to the mobile drone through the aquaculture cloud for monitoring the fish, persons, nets and feeding systems in an aquaculture site on a daily basis. Thus, the drone becomes a flyable intelligent robot that helps the manager of an aquaculture site to automatically collect valuable data that are important in optimization fish production using further decision making algorithms. Experiments show that our approach attains very high performance yielding significant semantics-based activity recognition accuracy without sacrificing the operation speed.

Integrated and Scalable Augmented Reality Multiplayer Robotic Platform

Budianto Tandianus[1], Hock Soon Seah[1], Li Wang[2], Chee Kwang Quah[3]
[1]School of Computer Science and Engineering, Nanyang Technological University, Singapore
[2]State Key Laboratory of Robotics and System Harbin Institute of Technology, China
[3]ST Electronics (Training & Simulation Systems) Pte Ltd, Singapore

Email: btandianus@ntu.edu.sg, ashsseah@ntu.edu.sg, 15b908017@hit.edu.cn, quah.cheekwang@stengg.com

Abstract – We propose a scalable AR (Augmented Reality) multiplayer robotic platform, which enables multiple players to control different machines (a drone and a robot) in shared environments, i.e virtual and real environments. We use state-of-the-art visual SLAM (Simultaneous Localization and Mapping) algorithms for tracking machine poses based on camera and IMU (Inertial Measurement Units) inputs. Players will observe consistent AR objects between them thanks to our backend system, which synchronizes the AR objects between players. Moreover, the system is scalable in term of hardware (e.g. IMU, camera, machine type) and software (SLAM algorithm) as we utilize ROS for communication between modules. We demonstrate our system on a game developed in Unity, a robust and widely used popular game engine. We present some statistics of the game such as its frames-per-second performance.

A Hybrid Nonlinear and Linear Approach for Content-Aware Image Downscaling

Takumi Owada, Yusuke Kameda, Ichiro Matsuda, Susumu Itoh
Department of Electrical Engineering, Tokyo University of Science
2641 Yamazaki Noda-shi, Chiba 278-8510, JAPAN

Email: owada@itohws01.ee.noda.tus.ac.jp

Abstract – In recent years, content-aware image resizing methods have attracted attention in order to support various display devices with different sizes and aspect ratios. Seam carving, which removes or duplicates connected paths of unimportant pixels called seams in an iterative manner, is a pioneer work for such a purpose and many researchers have followed this approach. However, the approach often suffers from visually annoying distortions when scaling ratios given by users are too low. This is mainly because penetration of the seams into some important objects becomes unavoidable at the latter stage of the iterative operations. As a solution of this problem, we previously proposed a nonlinear downscaling technique which iteratively performed a DCT-based locally linear scaling operation within 'belt-like seams', i.e. seams with a certain width. To enhance the above idea, in this paper, we replace the latter processing stage with a globally linear scaling operation. In this framework, an appropriate transition point from the nonlinear to linear operations is highly depend on the image content and must be determined automatically for practical use. In order to allow automatic determination of the transition point, we introduce a preservation measurement for the important objects. The measurement is defined based on integral amount of a mask image which designates shapes of the important objects. We employ the neural network-based object detection and segmentation method 'Mask R-CNN' to obtain the mask image. Simulation results show that our approach can produce subjectively better results than the conventional nonlinear downscaling methods.

Pose Estimation of Excavators

Hinako Nakaura[1], Yumeno Tsukada[2], Toru Tamaki[1], Bisser Raytchev[1], Kazufumi Kaneda[1]

[1] Department of Information Engineering, Graduate School of Engineering, Hiroshima University, Japan

[2]School of Engineering, Hiroshima University, Japan

Email: hinako-nakamura@hiroshima-u.ac.jp, yumeno-tsukada@hiroshima-u.ac.jp, tamaki@hiroshima-u.ac.jp, bisser@hiroshima-u.ac.jp, kin@hiroshima-u.ac.jp

Abstract – In current pose estimation studies, bottom-up approaches estimating the pose from detected parts of the bodies and their relations have been major methods. This approach enabled us fast and accurate pose estimation. In other hands, there are many fields that the pose estimation could be applied. In this paper, we propose a bottom-up method for excavator pose estimation. For pose estimation, we generate ground truth confidence maps according to the annotation and evaluate loss function for training. We evaluated the model by calculating the loss between the estimated map and true map.

Assessing viewer satisfaction of CG programs as a substitute for real TV programs

Yoshiaki Shishikui[1], Genki Nakamura[1], Masaki Hayashi[2]
[1]Meiji University, Tokyo, Japan
[2] Uppsala University, Visby, Sweden
Email: sisikui@meiji.ac.jp, masaki.hayashi@speldesign.uu.se

Abstract – We are investigating a system that provides a computer graphics (CG)-based television program to viewers by sending a script to the terminal using CG and voice data. In this study, we assess the viewer satisfaction of CG programs produced using a text-based TV program making language (TVML). To verify whether CG programs can be used as a substitute for real programs, we conducted a comparative evaluation experiment between real and CG programs. In addition, to verify the best device for CG program viewing, we compared the same CG program viewed on a PC display and a smartphone. The results suggest that CG programs are acceptable substitutes for real news and information-related programs, and that smartphones might be more suitable than PC displays for viewing CG programs.

Visual Analysis of Fish Feeding Intensity for Smart Feeding in Aquaculture Using Deep Learning

Jui-yuan Su[1,2], Pei-Hua Zhang[1], Sin-Yi Cai[1], Shyi-Chyi Cheng[1], Chin-Chun Chang[1]
[1]National Taiwan Ocean University, Department of Computer Science and Engineering, Keelung, Taiwan
[2]Department of New Media and Communications Administration, Ming Chuan University, Taipei, Taiwan

Email: rysu@mail.ntou.edu.tw, yvonne09751228@gmail.com, csc@mail.ntou.edu.tw, cvml@mail.ntou.edu.tw

Abstract – This paper presents a novel deep-learning approach to analyze the fish feeding intensity based on the images of fish tanks during the fish feeding process. The grade of the fish feeding intensity is an important indicator on fish appetite. On the design of a smart feeding system in aquaculture, this information is of great significance for guiding feeding and optimizing the fish production. However, conventional fish appetite assessment methods are inefficient and subjective. To solve these problems, in this study, based on a space-time two-stream 3D CNN, a deep learning approach for grading fish feeding intensity is proposed to evaluate fish appetite. The flow of the approach is implemented as follows. First, a fixed RGB camera is setup to capture the videos from the fish tanks during the feeding process. This also constructs a dataset for training the two-stream neural network, and the fish appetite levels are graded using the trained neural network model. Finally, the performance of the method is evaluated and compared with other CNN-based deep learning approaches. The results show that the grading accuracy reached 91.18%, which outperforms the compared CNN-based approaches. Thus, the model can be used to detect and evaluate fish appetite to guide production practices.

Two-Stream Deep Learning Architecture for Action Recognition by Using Extremely Low-Resolution Infrared Thermopile Arrays

Igor Morawski[1], Wen-Nung Lie[1,2,3]
[1]Department of Electrical Engineering, National Chung Cheng University, Chia-Yi, Taiwan
[2]Center for Innovative Research on Aging Society (CIRAS)
[3]Advanced Institute of Manufacturing with High-tech Innovations (AIM-HI)
Email: igor.morawski@hotmail.com, ieewnl@ccu.edu.tw

Abstract – This paper presents a new method for action recognition using an extremely low-resolution infrared imaging sensor. Thermopile arrays give users privacy but this comes at the price of limited information captured. The question of what methods are applicable to this sensor remains open. In our work, we adopt a two-stream deep learning architecture with feature-level fusion and optical flow to represent the motions. To the best of our knowledge, this is the first optical-flow-based method used in combination with extremely low-resolution thermal image sequences. We use a 16x16 pixel dataset introduced by a related work to directly compare the results and demonstrate the validity of optical flow for motion representation even in extreme low resolutions.

Recognition of Japanese Connected Cursive Characters Using Multiple Candidate Regions

Kazuya Ueki[1], Kohei Yamada[1], Ryou Mutou[1], Rostam Sayyed Nezhad[1], Tomoka Kojima[1], Yasuaki Hagiwara[1], Takuya Taketomi[1]
[1]School of Information Science, Meisei University, Tokyo, Japan
Email: kazuya.ueki@meisei-u.ac.jp

Abstract – It is difficult to recognize cursive characters (*kuzushiji*) in classic Japanese literature because multiple characters are connected. In this study, we propose a method for correctly recognizing consecutive *kuzushiji* characters by using multiple candidate regions as input to a neural network even if character regions are misaligned. An evaluation using an image database of three consecutive *kuzushiji* characters demonstrated that the proposed method had a higher accuracy rate than a method in which character detection preceded character recognition.

A Simple Refinement for Depth Information Predicted with DNN

Ren Sato[1], Akio Kimura[2]
[1]Division of Science and Engineering, Graduate School pf Arts and Science, Iwate University
[2]Faculty of Science and Engineering, Iwate University

Email: ren@mips.cis.iwate-u.ac.jp, kimura@mips.cis.iwate-u.ac.jp

Abstract – This paper proposes a refinement method for depth information estimated with deep neural network (DNN) from single-view RGB image, by applying simple image processing techniques. There have been proposed a lot of effective depth prediction methods based on DNN, such as ResNet-UpProj. However, such a learning-based method sometimes estimates unclear or uncertain depth information especially at around its edges, even if enough learning has been done. This paper aims to improve depth values especially around edges by applying simple image processing based on the position information of edges in original RGB image to the depth image obtained from the predicted depth information with DNN. The proposed approach is based on the following idea: we first estimate edge positions of the depth image properly, where corresponds to the edges of the original RGB image, and then modify depth values around edges of the depth image so as to make their change steeper. We did several experiments with NYU Depth v2 data-set, and showed that our simple approach can decrease about 14.3% of the root mean square error of depth values.

Handwriting Feature Extraction Method for Writer Verification Independent of Charactor Type by using AdaBN and AdaIN

Kimiya Murase[1], Shunsuke Nakatsuka[1], Mariko Hosoe[2], Kunihito Kato[1]
[1]Faculty of Engineering, Gifu University, Gifu, Japan
[2]Forensic Science Laboratory, Gifu Prefectural Police Headquarters, Gifu, Japan

Email: murase@cv.info.gifu-u.ac.jp, nakatsuka@cv.info.gifu-u.ac.jp, hosoe@cv.info.gifu-u.ac.jp, kkato@gifu-u.ac.jp

Abstract – Writer verification is usually conducted by using similarity between same type characters written by known and unknown writers. However, in the case of not present same type characters in each writer's documents, writer verification is so difficult. In this paper, we propose a method to extract the handwriting feature independent of the character type to solve this problem. The proposed model is based on AutoEncoder, and applying Adaptive Batch Normalization (AdaBN) or Adaptive Instance Normalization (AdaIN) for each layer to extract the objective features. We conducted a writer verification experiment using handwriting images pairs between different character types of ETL-1 Character Database (ETL-1). As a result of the experiment, we confirmed that the proposed method could perform writer verification with high accuracy even in the case.

Fast and Effective Object-Aware Domain Enhancement and Adaptation for Semantic Segmentation

Kuo-Liang Chung[1], Ya-Yun Cheng[1], Arie Tando[1], Don-Kai Chiang[2]
[1]Department of Computer Science and Information Engineering, National Taiwan University of Science and Technology, Taipei 10672, Taiwan R.O.C.
[2]KUS Qian Jun Technology Ltd., Taiyuan 1st, Zhubei City, Hsinchu County 302, Taiwan R.O.C.

Email: klchung01@gmail.com, yyiesfff@gmail.com, arie.tando91@gmail.com, dkchiang@gmail.com

Abstract – In this paper, we propose an effective object-aware domain enhancement and adaptation (ODEA) approach for semantic segmentation to increase the segmentation accuracy. In the proposed ODEA approach, first, we analyze the warning elements in the testing step, such as the falling-leaves, manhole covers, cirrus, and advertisements, etc, which caused invalid segmented objects. Then, we create a new GTA5-like (Grand Theft Auto V-like) dataset containing the scenarios including these warning elements. Further, we perform a domain adaptation on the created GTA5-like dataset to generate a photo-realistic GTA5-like dataset. Finally, we combine the generated dataset with the original photo-realistic GTA5 dataset and the realistic Camvid dataset to constitute a more diverse training dataset. The comprehensive experimental results have confirmed the semantic segmentation accuracy improvement of the proposed ODEA approach relative to existing domain adaptation methods by Tsai *et al.*

Halide Implementation of Weighted Median Filter

Akari Ishikawa[1], Hiroshi Tajima[1], Norishige Fukushima[1]
[1] Nagoya Institute of Technology, Nagoya, Japan

Email: a.ishikawa.914@stn.nitech.ac.jp, h.tajima.929@stn.nitech.ac.jp, fukushima@nitech.ac.jp

Abstract – With the recent extension of camera applications, image filtering is essential in image processing, such as image denoising and detail enhancement. Median filtering is one of the image denoising methods. The median filter replaces a pixel with the median value of the pixel's kernel. Also, weighted median filtering is an extention of the median filter. The filter had weights to each pixel in the kernel for statical computing values. The difference between the two filters is the removal capability of the impulsive noise. The weighted median filter can be more useful for removing noise and blurring correction. The drawback of the weighted median filter is its computational complexity.

Approaching the end of Moore's law, CPU microarchitectures become more complex year by year; therefore, it is difficult to write programs suited to each execution environment. Halide is one of a solution to this problem. The Halide is a domain-specific language for image processing and a pure functional language. Halide code is modularized as algorithm and scheduling parts. The algorithm parts show the image processing algorithm, and the scheduling parts reveal the computational order and computational method, e.g., vectorization, parallelization, and loop unrolling. By using Halide, we can easily optimize the code of image processing. In this paper, we present the weighted median filter with the Halide code. The implemented code is easy for us to write the weighted median filter code. Also, by adding the Halide scheduling code on this weighted median filter, we can efficiently parallelize the filter by slightly modifying the code.

Automated Classification Method of Lung Tumor Type using Cytological Image and Clinical Record

Ayumi Yamada[1], Atsushi Teramoto[1], Yuka Kiriyama[2], Testsuya Tsukamoto[2], Kazuyoshi Imaizumi[2], Masato Hoshi[1], Kuniaki Saito[1], Hiroshi Fujita[3]
[1]School of Medical Sciences, Fujita Health University, Toyoake, Japan
[2] School of Medicine, Fujita Health University, Toyoake, Japan
[3] Faculty of Engineering, Gifu University, Gifu, Japan

Email: ayumi926@fujita-hu.ac.jp

Abstract – Recently, as chemotherapy has advanced, it is important to accurately diagnosis the histological type (adenocarcinoma, squamous cell carcinoma and small cell carcinoma). In previous study, automated classification method for lung cancers in cytological images using a deep convolutional neural network (DCNN) was proposed. However, its classification accuracy is approximately 70%, therefore improvement in accuracy is required. In this study, we focus on liquid-based cytology images and clinical record. In this study, we aimed to improve the classification accuracy of lung cancer type by combining cytological images and electronic medical records. We aimed to develop of classification method of lung tumor type by combining cytological images and clinical record. First, the cytological images were collected. The original microscopic images were first cropped to obtain images with resolution 256 × 256 pixels. And then, we collected personal clinical data (age, gender, smoking status, laboratory test values, tumor markers and so on) corresponding to cytological images. Next, image features were extracted from cytological images using VGG-16 model pretrained on the ImageNet dataset. 4096 features before the fully connected layer were extracted. Then, these features were reduced dimensions by PCA. Image features obtained from the DCNN and clinical data corresponding to cytological images were given to the classifier. Finally, classification result of 3 histological categories was obtained. Evaluation results showed that classification by combining cytological images and clinical record improved classification accuracy than by cytological images alone. These results indicate that the proposed method may be useful for histological classification of lung tumor.

A Method for Rendering Wavelength-dependent Phenomena Using Spectral Image-based Lighting

Toru Nagai[1], Bisser Raychev[1], Toru Tamaki[1], Kazufumi Kaneda[1]
[1]Department of Information Engineering, Graduate School of Engineering,
Hiroshima University,
1-4-1 Kagamiyama, Higashi-hiroshima 739-8527, Japan

Email: {toru-nagai | bisser | tamaki | kin}@hiroshima-u.ac.jp

Abstract – Traditionally, RGB rendering that calculates the intensity of light in only three components has been often used for generating photorealistic images in global illumination environment, but the method cannot render wavelength-dependent phenomena accurately. On the other hand, Spectral rendering generates photorealistic images in various kinds of scenes including wavelength-dependent phenomena such as interference and fluorescence. However, the method is computationally expensive especially in global illumination environment compared to RGB rendering, because the spectral intensity of light over the range of the visible light should be calculated each time light rays collide with scene objects in the raytracing process. To reduce the computational cost of the spectral rendering, we introduce Image-based Lighting (IBL) where target objects are rendered without a number of iterations of the ray bouncing with scene objects by using a light probe image as an ambient light. We extend the IBL into a spectral IBL in order to combine IBL with spectral rendering, that is, a spectral image that includes the spectral intensity of ultraviolet in addition to visible light is used for the light probe image, and the spectral intensity of light is calculated to render the target objects. The proposed method is able to render wavelength-dependent phenomena realistically in shorter time, because the number of intersections between rays and objects is much smaller than that of the case without IBL. We have implemented the proposed method to a PBRT renderer and rendered scenes including the effects of fluorescence to demonstrate the usefulness of the proposed method.

Elastic Net with Adaptive Weight for Image Denoising

Jun Xiao [1], Rui Zhao [1], Kin-Man Lam [1]
[1] Department of Electronic and Information Engineering, the Hong Kong Polytechnic University, Hong Kong, China

Email: jun-gwansiu.xiao@connect.polyu.hk, rick10.zhao@connect.polyu.hk ,enkmlam@polyu.edu.hk

Abstract – Sparse models have been widely used in image denoising, and have achieved state-of-the-art performance in past years. Dictionary learning and sparse code estimation are the two key issues for sparse models. When a dictionary is learned, sparse code estimation is equivalent to a general least absolute shrinkage and selection operator (LASSO) problem. However, there are two limitations of LASSO: 1). LASSO gives rise to a biased estimation. 2). LASSO cannot select highly correlated features simultaneously. In recent years, methods for dictionary construction based on the nonlocal self-similarity property and weighted sparse model, relying on noise estimation, have been proposed. These methods can reduce the biased gap of the estimation, and thus achieve promising results for image denoising. In this paper, we propose an elastic net with adaptive weight for image denoising. Our proposed model can achieve nearly unbiased estimation and select highly correlated features. Experimental results show that our proposed method outperforms other state-of-the-art image denoising methods.

Operator Overloading for cv::UMat Converted to Equivalent Function Calls at Compile Time

Yusuke Kameda, Yoshihiro Maeda, Takayuki Hamamoto
Department of Electrical Engineering, Tokyo University of Science, Katsushika-ku, Japan

Email: kameda@ee.kagu.tus.ac.jp, ymaeda@rs.tus.ac.jp, hamamoto@rs.tus.ac.jp

Abstract – OpenCV is an open source programming library for computer vision and image processing, and has been used worldwide in industry and academia as a de facto standard for more than 10 years, from large-scale computers to embedded devices such as smartphones. OpenCV has image data representation classes cv::Mat and cv::UMat as image processing APIs, and the latter can implement parallel computation using heterogeneous computing such as the OpenCL framework. Operator overloading is a technique that enables intuitive programming using arithmetic operations and assignment operators, and is defined in cv::Mat. On the other hand, operator overloading is not defined in cv::UMat, and it is necessary to call functions appropriately. Therefore, programming using cv::UMat is difficult, and there is a problem that the code is not compatible with cv::Mat using operator overloading. In addition, cv::Mat operator overloading has a problem of extra memory reallocation at runtime, so it is not appropriate to apply it directly to cv::UMat. Therefore, in this paper, we propose a method to realize operator overloading that does not require extra memory reallocation at runtime and can be equivalently converted to cv::UMat function calls at compile time. This method enables programming that achieves intuitive operator overloading without any overhead at runtime.

Temporal and Contextual Aggregation for Road Making Semantic Segmentation

Hiroaki Aizawa[1], Yuto Ura[1], Kunihito Kato[1]
[1]Gifu University, Gifu, Japan

Email: aizawa@cv.info.gifu-u.ac.jp, ura@cv.info.gifu-u.ac.jp, kkato@gifu-u.ac.jp

Abstract – The road scene understanding is an important task in the fields of automatic driving and a road maintenance system. Recently, semantic segmentation approaches which perform pixel-wise object classification allows us to recognize the road scene accurately. However, in the road scene, some classes are difficult to recognize with a single static image, such as a white center line and a white dash line. Therefore, we introduce novel convolutional neural networks for semantic segmentation on road markings in the situation where sequential segmentation ground truth masks are available. The proposed model unifies the temporal information in video data and the context information in a static image using a 3D convolution and a skip-connection for accurate segmentation. Moreover, we employ CGNet as the backbone network to reduce trainable parameters. In the experiment, we evaluate the model using the Gifu-city road making segmentation dataset, which includes road markings of open roads in Gifu city. As a result, the segmentation performance such as a white center line and white dash line is an improvement.

Automatic Meniscus Segmentation using Cascaded Deep Convolutional Neural Networks with 2D Conditional Random Fields in Knee MR Images

Uju Jeon[1], Hyeonjin Kim[1], Helen Hong[1,*], Joon Ho Wang[2]
[1]Department of Software Convergence, Seoul Women's University, Seoul, South Korea
[2]Department of Orthopedic Surgery, Samsung Medical Center, Seoul, South Korea

Email: ujujeon@swu.ac.kr, hyunjinkim@swu.ac.kr, hlhong@swu.ac.kr, mdwang88@gmail.com

Abstract – Meniscus segmentation of MR images is necessary to generate a patient-specific model of meniscus for surgical planning and 3D printing of knee joint disease. However, automatic meniscus segmentation is challenging due to thin shape, shape variance among patients, and similar signal intensity to nearby organs such as cruciate and collateral ligaments. Our cascaded segmentation network consists of two steps: meniscus localization and meniscus segmentation. First, to localize the medial and lateral meniscus, the meniscus is roughly segmented using a U-Net based 2D CNN with six classes including femoral and tibial bones and cartilages, meniscus and background. Then, 2D CRF is performed to prevent the under-segmented area of the roughly segmented meniscus, and the medial and lateral meniscus are localized by the segmentation results. Second, to accurately segment the meniscus, a U-Net based 3D CNN with two classes including a meniscus and background focusing on local information and considering spatial information is performed. The 2D segmentation network showed under-segmentation inside the meniscus. The under-segmentation was prevented after 2D CRF, but over-segmentation occurred in nearby ligaments with similar intensity. The 3D segmentation network prevented under- and over-segmentation due to considering local and spatial information, and showed the best performance. The medial meniscus shows higher accuracy than the lateral meniscus due to less leakage into the collateral ligament. The average dice similarity coefficient of medial and lateral meniscus were 92.27% and 90.27%, showed better results of 4.78%

and 9.96% at medial meniscus, 3.94% and 9.58% at lateral meniscus compared to the segmentation method using 2D U-Net results and combined 2D U-Net and 2D CRF, respectively.

(This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT (NRF-2017R1A2B3007362),
*Corresponding author: Helen Hong)

Indonesian Culture Recognition Portal based on Crowd Sourcing Contents

Winarno[1], Rudy Pramono[2], Wella[1]
[1]Universitas Multimedia Nusantara, Faculty of Engineering and Informatics, Tangerang, Indonesia
[2]Pelita Harapan University, Tangerang, Indonesia

Email: pmwinarno@umn.ac.id, rudy.pramono@uph.edu. wella@umn.ac.id

Abstract – The people of Indonesia experiences high cultural crisis, especially the younger generation who prefer foreign cultures rather than local cultures, so that local cultures fade away. The aim of this research is to develop a portal website about Indonesian cultures based on crowd sourcing contents to rise the recognition and conservation of local cultures. This research seeks to change the perception of old fashioned local cultures to become new fashioned local cultures by combining local cultures with the implementation of information and communication technology which is currently being used by the people of all walks of life. The method used in this study include system development life cycle and content crowdsourcing, because it requires the help of all members of the communities to fill in the content and managed data which later have a large volume. The portal website was successfully created, and later it needs website contents such as video, photographs and narration via crowd sourcing mechanism.

Deep Skip Connection and Multi-Deconvolution Network for Single Image Super-Resolution

Shiyu Hu[1], Muwei Jian[2], Guodong Wang*[1], Yanjie Wang[1], Zhenkuan Pan[1], Kin-Man Lam[3]

[1]College of Computer Science and Technology, Qingdao University, PR China, 266071.
[2]School of Computer Science and Technology, Shandong University of Finance and Economics, Jinan, China.
[3]Centre for Signal Processing, Department of Electronic and Information Engineering, Hong Kong Polytechnic University, Hong Kong.

Email: hudhiyu@qdu.edu.cn, jianmuweihk@163.com, doctorwgd@gmail.com, wanyanjie@qdu.edu.cn, zkpan@qdu.edu.cn, enkmlam@polyu.edu.hk

Abstract –In this paper, we propose an efficient single image super-resolution (SR) method for multi-scale image texture recovery, based on Deep Skip Connection and Multi-Deconvolution Network. Our proposed method focuses on enhancing the expression capability of the convolutional neural network, so as to significantly improve the accuracy of the reconstructed higher-resolution texture details in images. The use of deep skip connection (DSC) can make full use of low-level information with the rich deep features. The multi-deconvolution layers (MDL) introduced can decrease the feature dimension, so this can reduce the computation required, caused by deepening the number of layers. All these features can reconstruct high-quality SR images. Experiment results show that our proposed method achieves state-of-the-art performance.

A VR-based Support System of Self-learning Microscope Operation

Hikaru Kaneko, Mitsunori Makino
Department of Information and System Engineering, Chuo University, Tokyo, Japan

Email: 14.8jtj@g.chuo-u.ac.jp, makino@m.ieice.org

Abstract – Acquiring microscope operation skills has been contained by Japanese secondary education. Although many chances of observation with the microscope operation must inspire students to study deeply science, the curriculum allows short time and small budget. In this paper, a virtual reality (VR) system of self-learning for the microscope operation is proposed, which makes inexperienced students equip fundamental skill of the operation. The system, composed by a computer, a head-mounted-display (HMD) and a dial type input device, displays the student a virtual microscope in the center of the virtual world. In the virtual world, the student can manipulate the virtual microscope with controllers and a dial type input device, as well as he/she can see the procedures of the operation. Since the operation to focus with focus knob is especially important in the procedure, the system adopts a dial type input device to experience under the condition more similar to the reality by the sense of touch. The system displays how to operate a microscope by visual information in texts and figures and judges if user's operation is correct by comparing result of user's operation and the correct one stored in the system such as the position and inclination of virtual objects. If the user operates incorrectly, the system represents an alert. From the above, the system is expected to help users equip microscope operation skills practically at relatively low cost.

AR based Self Learning System of Japanese Calligraphy Skills

Yusuke Ishihara, Mitsunori Makino
Department of Information and System Engineering, Chuo University, Tokyo, Japan

Email: a14.kwhr@g.chuo-u.ac.jp, makino@m.ieice.org

Abstract – While Japanese calligraphy has been contained by Japanese elementary and secondary education curriculum, teachers in charge of the calligraphy education are not always proficient, since skill of the calligraphy is unnecessary for teachers' license. To overcome the problematic situation, not only training of teachers but also an ICT-based support system for the students to learn the skill by themselves. Therefore this article proposes a system that combines a self-study support system for inexperienced people to learn Japanese calligraphy skills, and a virtual writing system for practicing calligraphy without using ink in a virtual environment. The proposed system is a kind of the augmented reality (AR) system, which consists of a computer, a head-mounted-display (HMD) and a non-contact motion sensor. Firstly, the system introduces AR technology into displaying the dynamic brush stroke example to the user with CG animation, which is previously recorded from an expert's action. Secondly, on the system users practice calligraphy by imitating example brush stroke on virtual paper. To obtain the user's motion, the motion sensor attached to the HMD acquires the position of the tip and end of the brush at each frame. Simultaneously, the system simulates the footprint of brush from these data and displays it at the tip of the brush using AR technology. After the users' practice, the system calculates the differences in brush position, tilt and speed between the expert's brush stroke and users' brush stroke by using DP matching and encourages users to improve their brush stroke, finally. The system is expected to help users learn Japanese calligraphy skills for themselves.

Enhanced Combined Inter-Intra Prediction (CIIP) in Versatile Video Coding

JiHoon Do, Beomhee Cho, Jae-Gon Kim, and Dae-Gwon Jeong
Korea Aerospace University, School of Electronics and Information Engineering, Goyang-si, Korea

Email: {jhdo, run4elephant}@kau.kr, {jgkim, dgjeong}@kau.ac.kr

Abstract – Joint Video Experts Team (JVET) are developing a new video coding standard beyond High Efficiency Video Coding (HEVC) named as Versatile Video Coding (VVC). In VVC, various new prediction modes have been adopted compared to HEVC and Combined Inter-Intra Prediction (CIIP) is one of them. CIIP is a new prediction mode adopted in VVC which combines inter prediction and intra prediction with derived weights to form a final prediction. In this paper, we propose methods to enhance CIIP with more accurate weights for the combination of both predictions as well as extending intra prediction modes to be combined based on the adjacent blocks' coding modes. According to empirical observations, the below-left block and above-right block are correlated with left and above blocks in terms of prediction mode, respectively. So, the first proposed method is to derive finer weight values by using prediction modes of more adjacent blocks up to 3 blocks, and weight values is derived by the blocks of left, above and above-left than the blocks of below-left and above-right. The second proposed method is to use intra-coded modes of adjacent blocks of left and above blocks which are used to derive MPM candidates to be combined instead of planar mode used in the current CIIP. Experiment results show that the proposed methods slightly improve the performance of CIIP in the VVC Test Model (VTM).

Radiomic Feature-based Prediction Model of Lung Cancer Recurrence in NSCLC Patients

Soomin Lee[1], Julip Jung[1], Helen Hong[1], Bongseok Kim[2]
[1] Department of Software convergence, Seoul Women's University, Seoul, South Korea
[2] Hematological Oncology, Central Veterans Hospital, Seoul, South Korea

Email: soominlee@swu.ac.kr, jjulip@swu.ac.kr, hlhong@swu.ac.kr, seog@bohun.or.kr

Abstract – Non-small cell lung cancer (NSCLC) is the most commonly diagnosed type of lung cancer, with the majority of adenocarcinoma and squamous cell carcinoma. Recently, image analysis techniques have been used successfully to provide a more accurate personalized prognosis and treatment plan. The aim of this study is to investigate the potential of a radiomic feature-based prediction model for two-year relapse-free survival (RFS) in NSCLC patients on chest CT images. Our method is composed of the major three steps. First, tumor areas are defined as intra-tumoral areas that have been manually segmented by a radiologist. Second, a total of 68 radiomic features are extracted within the tumor area. Then, eight features with weights that are clearly distinguished from other weights are defined as significant features using the reliefF algorithm. Finally, to predict lung cancer recurrence, random forests are trained for the classification of two groups representing recurrence and non-recurrence within two years. 156 NSCLC patients of lung adenocarcinoma and lung squamous cell carcinoma were used in this study and the number of recurrence and non-recurrence patients was 79 and 77. In the experimental results, the accuracy, sensitivity, specificity, and AUC were 70.97, 56.25, 86.67, and 0.78, respectively.

(This research was supported by the Bio & Medical Technology Development Program of the National Research Foundation (NRF) funded by the Ministry of Science &ICT(2015-2015M3A9A7029725), Corresponding author: Helen Hong)

Super-resolution Image Generation for Improvement of Orbital Thin Bone Segmentation

Hee Rim Yun [1], Min Jin Lee [1], Helen Hong [1,*], Kyu Won Shim [2]
[1] Seoul Woman's University, Dept. of Software Convergence, Seoul, Korea
[2] Severance Children's Hospital, Dept. of Pediatric Neurosurgery, Seoul, Korea

Email: heerimyun@swu.ac.kr, minjin@swu.ac.kr, hlhong@swu.ac.kr, SHIMKYUWON@yuhs.ac

Abstract – Orbital bones are composed of high-intensity, large-area cortical bones, and low-intensity, narrow-area thin bones. To make a customized orbital bone plate modeling, it is necessary to segment accurately the orbital bone in head-and-neck CT images. However, since the slice thickness is two to three times larger than the pixel spacing, the segmentation accuracy is degraded and the reconstructed thin bones of the orbital bone show an aliasing effect. Thus, we generate super-resolution orbital bone CT images using 3D SRGAN network to improve the accuracy of segmentation of thin bones with large slice thickness. Our method is composed of two major steps. First, for data preparation, the thin bone area of the orbital bone is cropped to size of 164x252x80 in head-and-neck CT images. Second, super-resolution thin bone volume is generated using a 3D SRGAN network consisting of generator and discriminator. We experimented with 14 orbital bone CT images and measured the peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM) scores. Experimental results show that the generated images using 3D SRGAN are clearly observed in thin bone areas and are more similar to the original images than other images using bilinear and bicubic interpolations. 3D SRGAN was the best in terms of PSNR and SSIM and there was a significant difference in performance between 3D SRGAN and bilinear interpolation (p-value < 0.001) and 3D SRGAN and bicubic interpolation (p-value < 0.001). (This research was supported by Basic Science Research Program through the National Research Foundation of Korea funded by the Ministry of Education(No. 2017R1D1A1B03034927) and the Technology Innovation Program (or Industrial Strategic Technology Development Program) (20001811, International standardization on deep learning based human tissue 3D modelling) funded by the Ministry of Trade, Industry & Energy (MOTIE, Korea), *Corresponding author: Helen Hong)

Identifying Blurry Car License Number Plate Using Machine Learning

Erika Sawada[1], Ayaka Fujima[1], Seiichi Gohshi[1], Naiwala P. Chandrasiri[1]
[1] Department of Informatics Kogakuin University

Email: em19006@ns.kogakuin.ac.jp, j215109@ns.kogakuin.ac.jp,
gohshi@ns.kogakuin.ac.jp,, chandrasiri@ns.kogakuin.ac.jp

Abstract – Equipment used in a crime is a clue to solve cases. Cars are often used in crimes as a transportation tool. Security cameras are set in everywhere of downtowns. Drive recorders are set in many cars in Japan. The image of the license number plate is often recorded in the security camera and in the driver recorder. However, the license number plate is generally recorded in a small part of the image. It is necessary to enlarge the image to read the number. However, the enlarged image is blurry. Since the edges of the blurry image is not crisp, it is difficult to read the number plate from the blurry image. Machine learning is a practical tool to identify the number in the blurry image. Performance of machine learning changes depending on classifiers. In this paper, performances of five classifiers for identifying the license number plate are compared. The conclusion of the paper has possibility to contribute solving criminal cases.

Development of a Database of Radar-shadow Cast by a SAR Satellite Using High Resolution DEM Data

Makoto Ide, Toshikazu Samura and Katsumi Tadamura
Graduate School of Sciences and Technology for Innovation, Yamaguchi University, Ube, Japan

Email: b005vg@yamaguchi-u.ac.jp, samura@yamaguchi-u.ac.jp, tadamura@yamaguchi-u.ac.jp

Abstract – Synthetic aperture radar (SAR) aboard a satellite irradiates observation microwave toward the earth and the microwave arriving at the surface of the earth scatters every direction. The backward scattering component of it is received on the antenna of the satellite. The satellite SAR allows us to observe the earth during 24 hours without being influenced by the weather because of this observation principle. On the other hand, since the observation microwave needs to be irradiated toward obliquely downward, radar shadow area, at which the observation microwave does not arrive due to obstruction of the other higher place, appears. There is no observation data at the radar shadow area by nature. However, there is no clear indication about the radar shadow area in a satellite SAR's observation data (satellite SAR image). Therefore, when we use the satellite SAR image to detect occurrence of a natural disaster, we might carry out useless calculation about a steep valley bottom region which has a higher possibility of occurrence of the land slide. This is because the steep valley bottom region is often included in the radar shadow area, that is the no-data area. If we can know about the radar shadow area corresponding to the latest observation data, improvement of the estimation precision of occurrence of a natural disaster by using satellite SAR image can be expected. In this paper, we propose a method for efficiently developing a database for radar shadow cast by ALOS-2/PALSAR-2 (a satellite operated by JAXA) by using high resolution (5 meter mesh) DEM data as a pre-process.

Camera – LiDAR Calibration Using ICP-based Automatic Plane Extraction Method

Seiyoul Jung[1], Jeyeon Kim[2], Hongjun Lee[2], Seunghyun Kim[2], Whoi-Yul Kim[2]
[1]Department of Intelligent Robot Engineering Hanyang university Seoul, South Korea
[2]Department of Electronics and Computer Engineering Haynag university Seoul, South Korea

Email: syjung@vision.hanyang.ac.kr, jykim@vision.hanyang.ac.kr, hongjunlee@vision.hanyang.ac.kr, shkim@vision.hanyang.ac.kr, wykim@vision.hanyang.ac.kr

Abstract – In the autonomous driving systems, camera and LiDAR are widely used to recognize surrounding environment. For the fusion of camera and LiDAR, it is necessary to conduct calibration. The existing approach has disadvantages of manually extracting planes from 3D LiDAR point cloud to conduct calibration. Therefore, we propose method for automatically extracting planes from LiDAR point cloud using Iterative Closet Point. Also, by defining new loss terms for optimization, we improve the calibration accuracy.

PSF Optimization for Motion Deblurring using Temporally Coded Light

Jaelin Lee, Gyoheak Oh, and Byeungwoo Jeon,
Department of Electrical and Computer Engineering, Sungkyunkwan University, Suwon, Korea

Email: {jaelin, dhrygur, bjeon}@skku.edu

Abstract – Motion blurs inevitably occur in an image photographing fast moving objects, and its removal, known as motion deblurring, is one of the most well-known ill-posed problems. In this paper, we investigate the deblurring problem of motion blurs by using a modulated external light. Noting that the motion blurs depend both on camera exposure time and modulation of external light, we investigate how to optimize the point spread function (PSF) pattern considering not only the external light but also the ambient light. More particularly, we generate a new PSF pattern by adding the external modulated light pattern and analyzing the ambient light pattern. Using the proposed method, the noise is found to be much reduced compared to the conventional methods considering only external modulated light.

Robotic Path Planning using Evolutionary Neural Network

Johanes Immanuel[1], Alethea Suryadibrata[2], Julio Christian Young[3]
Universitas Multimedia Nusantara, Department of Informatics, Tangerang, Indonesia

Email: johanes.immanuel@student.umn.ac.id, alethea@umn.ac.id,
julio.christian@umn.ac.id

Abstract – Path planning is a well-known problem in mobile robot. The robot needs to move from the starting point to destination point and avoid obstacles. To create an agent that able to reach some destinations in the given environment, such agent need some abilities to processing a collection of information obtained by its sensors and roam freely in the environment. In this paper, we design mobile agents to solve local path planning problems in 3D environment by using Evolutionary Neural Network (ENN) algorithm. ENN combines Evolutionary Algorithm (EA) and Neural Network algorithm. We chose Genetic Algorithm (GA) for the EA part and designing a simple feed forward neural network for the neural networks part. We evaluate what kind of ENN configuration values that works best in a local path planning problem. Experiment results show that the lowest iteration rate is 1.8 with one hidden layer and 50 hidden nodes when the population size is 50.

Deep Neural Network for Joint Light Field Deblurring and Super-Resolution

Jonathan Samuel Lumentut[1], In Kyu Park[1]
[1]Inha University, Department of Information and Communication Engineering, Incheon, Korea
Email: jlumentut@gmail.com, pik@inha.ac.kr

Abstract – The recent works on the light field (LF) image enhancement are focused on specific tasks such as motion deblurring and super-resolution. State-of-the-art methods are limited with the specific case of 3-degree-of-freedom (3-DOF) camera motion (for motion deblurring) and straight-forward high-resolution neural network (for super-resolution (SR)). In this work, we proposed a framework that utilizes the deep neural net to solve LF spatial super-resolution and deblurring under 6-DOF camera motion. The neural network is designed with end-to-end fashion and trained in multiple stages to perform robust super-resolution and deblurring. Our neural network achieves superior result in terms of quantitative and qualitative performance compared to the recent state-of-the-art recent LF deblurring and SR algorithms.

Effective Binarization for Historically Degraded As-built Drawing Maps Using Convolutional Neural Networks

Kuo-Liang Chung *[a], De-Wei Hsieh[a], Chi-Huang Liao[b]
[a] Computer Science and Information Engineering National Taiwan University of Science and Technology Taipei, Taiwan;
[b] System Online Co., Ltd. Taipei, Taiwan

Abstract – Binarizing historically degraded as-built drawing (HDAD) maps is a challenging job, especially in removing noise, yellowing areas, and folded lines while preserving the foreground components. This paper first proposes a convolutional neural networks-based (CNN-based) color classifier to determine the dominant color class of each HDAD block. Then, a dominant color driven- and CNN-based binarization method is proposed, producing a high-quality binarized HDAD map. Based on real HDAD dataset, the thorough experiments have been carried out to show that in terms of F-measure and perceptual effect, our binarization method substantially outperforms existing state-of-the art binarization methods.

Multi-frame Interpolation of Bayer Images Using Optical Flow

Toshiki Sato, Norishige Fukushima
Nagoya Institute of Technology, Nagoya, Japan
Email: t.sato.793@stn.nitech.ac.jp, fukushima@nitech.ac.jp

Abstract – A camera's image sensor contains millions of pixels. The pixel itself records only the intensity of a single light of red, green, or blue in Bayer patterned cameras. When the camera moves position slightly, however, the different kinds of light signals can be recorded in a different pixel. In other words, we can complement the missing color signals in the Bayer image pattern by the captured ray from the moved the camera. Consequently, a full-color image can be generated without demosaicing, which may cause a false-color. In our proposed method, we take more than 100 images to complement the pixels because the camera position is moved artificially rather than mechanically. Besides, we use the dense optical flow to track the movement of all pixels between the reference image and the other images. All images are converted from Bayer images to grayscale images to calculate the optical flow. Then, the pixels are linked according to the optical flow, the missing color pixel values in reference Bayer image are complemented from the other corresponding Bayer images. In our experiments, we generated sharper images than the demosaiced method.

A Web Based Computer Aided Diagnosis System on Liver Disease

Xuejun Zhang[1,2], Ju Huang[1], Huining Huang[1], Bijiang Li[1], Binmei Liang[1], Dongbo Wu[3]
[1]School of Computer, Electronics and Information, Guangxi University, Nanning, Guangxi, China
[2]Guangxi Key Laboratory of Multimedia Communications and Network Technology, Nanning, China
[3]Liuzhou Worker's Hospital, Liuzhou, China

Email: xjzhang@gxu.edu.cn

Abstract –Computer aided diagnosis (CAD) system has been proven to be useful in clinical routine. However, different kinds of software installed in different machine limits the widely usage of CAD by doctors. With the rapid development of Internet technology, medical image processing system has developed from the past stand-alone mode to network application. We transfer our previous CAD system on liver disease into a web based program to enable users to diagnose potential hepatic abnormalities through internet, by using Xojo platform which is easy to make web application under BASIC programming language and provide virtual server function when in running mode. Integrating different functions like image data collection without privacy information, liver segmentation, tumor detection and fibrosis classification into one website, our web applications can be run either as stand-alone Web servers or as CGIs on specific ports, using HTTP and/or HTTPS for secure communication. The result shows that our CAD system can be easily open on different OS and any location with network connection. Such convenience makes diagnosis less time consuming while significantly collecting datasets via internet from different hospitals or even patient him/herself without raising privacy and network security issues. We believe that our CAD is possible to come to clinical practice among many hospitals with its web manipulability and the continuous improving web speed by 5G communication.

Multimodality breast mass classification using CNN-based similarity estimation

Chisako Muramatsu[1], Mikinao Oiwa[2], Takako Morita[2], Tomonori Kawasaki[3], Hiroshi Fujita[4]
[1]Shiga University, Faculty of Data Science, Hikone, Japan
[2]Nagoya Medical Center, Nagoya, Japan
[3]Saitama Medical University International Medical Center, Hidaka, Japan
[4]Gifu University, Dept of Electrical, Electronic and Computer Engineering, Gifu, Japan

Email: chisako-muramatsu@biwako.shiga-u.ac.jp

Abstract – An intelligent image analytic system can support radiologists' accurate and efficient image interpretation. Presentation of images similar to a new lesion as a reference can help in image diagnosis and treatment planning. In this study, we investigated an image retrieval method based on the similarity estimation using convolutional neural networks (CNN) for diagnosis of breast masses on mammograms and ultrasound images. Using CNN, diagnostically relevant images can be retrieved without lesion segmentation. With the multimodality data, retrieval precision was improved. Similar reference images with relevant clinical data can be helpful in diagnosis of breast masses.

Automated Retinopathy Detection Based on Convolutional Neural Network on Retinal Images

Yuji Hatanaka[1], Phooi Yee Lau[1], Paulo Correia[2], Paulo Fonseca[3], Aida Campos[3]
[1]Department of Electronic Systems Engineering, School of Engineering, The University of Shiga Prefecture
Hikone, Japan

Email: hatanaka.y@e.usp.ac.jp

Abstract – Retinopathy is a complication of systemic diabetes or hypertension, and it may cause vision loss. Diabetic retinopathy is the leading cause of blindness; thus, it is necessary to detect a retinopathy in the early stage. Thus, we develop the detection methods based on a convolutional neural network (CNN) of microaneurysms and exudates, which are the signs of diabetic retinopathy. Many labeled samples need for training of CNN, but it is hard work to collect many lesions and to label them. Therefore, we propose a method using a small number of labeled lesion samples. In order to train the CNN using a small number of microaneurysm images, the enhanced images were applied. As enhancement filters, we applied Gabor filter, Shape index of Hessian matrix and our original double-ring filter. Moreover, the false positives (FPs) were reduced by three-layer neural network with 36 image features. For training CNN, 246 microaneurysms were augmented to 2460 images by the general data augmentation technique. By testing 100 microaneurysms included 61 retinal images, the sensitivity was reached to 0.89 with 2.0 FPs per image. Moreover, data augmentation using Wasserstein Generative Adversarial Networks Gradient Penalty (WGAN-GP) was applied in order to overcome the lack of the sample of exudates. CNN for exudates detection was trained using 3,000 real exudates image, 27,000 artificial exudate images by WGAN-GP. The sensitivity and specificity were reached to 0.75 and 0.96, respectively. The sensitivity and specificity were 0.70 and 0.98 when CNN was trained without artificial images, so that WGAP-GP maybe useful for CNN training. The CNN for the lesions of diabetic retinopathy could work effectively by application of image enhancement filter or data augmentation using WGAN-GP.

Segmentation of Pulmonary Abnormalities on CT Images using Deep Learning

Shoji Kido[1], Yuki Suzuki[1], Noriyuki Tomiyama[2]
[1] Department of Artificial Intelligence Diagnostic Radiology, Osaka University Graduate School of Medicine, Suita, Japan
[2]Department of Radiology, Osaka University Graduate School of Medicine, Suita, Japan

Email: kido@radiol.med.osaka-u.ac.jp, y-suzuki@radiol.med.osaka-u.ac.jp, tomiyama@radiol.med.osaka-u.ac.jp

Abstract – Recently, deep-learning based approaches have achieved the state-of-the-art performance in image segmentation. In this presentation, we introduce an overview of deep learning-based approaches of segmentation for pulmonary abnormalities on CT images. We used U-Net to segment each opacity pattern of diffuse lung diseases. U-Net has an encoder-decoder structure. We also used residual U-Net, using residual units to improve training problems in deep layers. The opacities we used were consolidation, ground-glass opacity, honeycombing, emphysema, diffuse nodular and normal lungs. As the training data, we used coincided areas annotated by three radiologists. The structure of U-Net proposed by us has 19 convolution layers and four max-pooling layers. The structure of residual U-Net proposed by us has 11 convolution layers and the two max-pooling layers, were configured to be shallower than U-Net. The results were in good agreement with the radiologist's annotation. We segmented lung nodules from 3D images using Deconvolution Network and V-Net. V-Net is a method proposed for the purpose of medical image segmentation as well as U-Net. While U-Net targets two-dimensional images, V-Net targets 3D images. As the difference between the two models, Deconvolution Network uses the pooling layer and the unpooling layer during image compression and decompression. However, V-Net uses a convolution layer of two kernel sizes and strides and a deconvolution layer. As training images, the images annotated under the guidance of a radiologist were used, which were cut out into 128 × 128 × 64 pixels centering on the gravity of the nodule. The extraction results of Deconvolution Network and V-Net coincided well with the radiologist's annotation. Segmentation is an important technique for quantitative analysis of

pulmonary abnormalities on CT images. Deep learning-based approaches are able to segment pulmonary abnormalities accurately and robustly, so they will be able to improve the diagnostic accuracy of pulmonary abnormalities.

Breast Ultrasound Computer-aided Diagnosis Using Deep Learning

Ruey-Feng Chang[1,2]
[1]Department of Computer Science and Information Engineering
[2]Graduate Institute of Biomedical Electronics and Bioinformatics, National Taiwan University, Taipei, Taiwan

Email: rfchang@csie.ntu.edu.tw

Abstract – Breast ultrasound is the common examination for tumor detection and classification. Computer-aided diagnosis (CAD) system is a useful tool to diagnose a tumor. With the growth of computing power, particularly GPU computing, deep learning has been applied in many different domains and especially in image recognition. Recently, we have used a computer-aided diagnosis (CADx) system based on the convolutional neural network (CNN) method to diagnose the breast tumors using 2-D breast ultrasound and ultrasound elastography. Automated whole breast ultrasound (ABUS) has been widely used as a screening modality. Reviewing hundreds of ABUS slices is time-consuming. Therefore, a fast and effective computer-aided detection (CADe) system based on 3-D CNN to accelerate this reviewing. Also, a 3-D CNN-based CADx system is proposed to diagnose the tumors of ABUS images. The proposed ABUS CADx method is utilized two 3-D CNNs with different architectures to obtain the texture and the morphology features, respectively.

Halide and OpenMP for Generating High-Performance Recursive Filters

Yuta Tsuji and Norishige Fukushima
Nagoya Institute of Technology, Nagoya, Japan

Email: y.tsuji.635@stn.nitech.ac.jp, fukushima@nitech.ac.jp

Abstract – Halide is a domain-specific language for image processing. Halide can separate code into an algorithm part and a scheduling part. We can write how to work image processing in the algorithm part, and how to compute it in the scheduling part. The scheduling part has a restriction, which does not change the calculation result. The restriction prevents efficient code generation for some kinds of image processing. In this paper, we propose high-performance recursive filters with Halide and OpenMP. The recursive filter is one of the difficult algorithms for optimizing code with Halide. In our implementation, we divided an input image into multiple tiles, and then each tile is processed with Halide code. Also, each processing for a tile in Halide is parallelized by OpenMP. The processing has an approximated computation in boundary conditions for forcefully cutting the effect from the next tiles; thus, the resulting image has slightly degraded. The closed tile, however, improves the cache efficiency in computation. In the experiment, the processing time of the box image filtering with and without tiling was compared. Box filtering is the most simple recursive filter. The box image filtering uses an integral image technique to process in constant time, and its calculation includes recursive filters. Code with tiling, which was the proposed method, the tile size is 128×128. The experimental results showed that the proposed method had a better computational time performance than the code without tiling.

Deep Learning Based Segmentation for Part Solid Nodule in Chest CT

Young Jae Kim [a], Ye Rang Park [b], GyeongYun Yi [b], Kwang Gi Kim *[a]
[a]Dept. of Biomedical Engineering, Gachon University, Incheon, South Korea;
[b]Dept. of Health Sciences and Technology Gachon Advanced Institute for Health Sciences and Technology (GAIHST), Gachon University, Incehon, South Korea

Email: youngjae@gachon.ac.kr, kimkg@gachon.ac.kr

Abstract – In this study, we aimed to segment two types of part solid nodules (solid, non-solid) using the deep convolutional neural network architecture. A total of 25,993 images and 9,159 images obtained from CT of 1,477 patients were used for experiments on the non-solid and solid regions, respectively. Deep learning was performed on the non-solid and the solid regions using the KERAS API based on the tensorflow in the Ubuntu system. In this study, we have validated the deep learning models for the segmentation of the non-solid and the solid regions with test sets. As a result, the non-solid region showed an average sensitivity of 82 % and an average DSC of 0.8, and the solid region showed an average sensitivity of 91 % and an average DSC of 0.9. In chest CT, deep learning based models showed relatively high accuracy for all non-solid and solid types of nodule.

Lymphocyte Detection in Liver Histopathological Image

Chao-Ting Li [1], Pau-Choo Chung [1], Kuo-Sheng Cheng [2], Hung-Wen Tsai [2], Nan-Haw Chow [2]

[1]National Cheng Kung University, Institute of Computer and Communication Engineering,

[2] National Cheng Kung University (NCKU), Department of Biomedical Engineering, NECK Hospital

Email: chting658@gmail.com, pcchung@ee.ncku.edu.tw, kscheng@mail.ncku.edu.tw, hungwen@mail.ncku.edu.tw, chownh@mail.ncku.edu.tw

Abstract – Medical images segmentation is a fundamental challenge in pathological image analysis. However, it is commonly occurred extreme class imbalance problem. To address the problem, this paper proposed an active learning (AL) framework to solve the extreme class imbalance problem by both under-sampling majority class and over-sampling minority class. Experimental results show that the proposed method can achieve annotation-effective solution in extremely imbalanced class segmentation. The contribution of the proposed method has two-folds, (1) we proposed an AL framework for solving the extreme class imbalance problem by both under-sampling majority class and over-sampling minority class. (2) the proposed framework achieves good performance for lymphocyte detection in histopathological image with fewer labeled samples.

Deep Residual Convolutional Neural Network with Curriculum Learning for Source Camera Identification

Animasahun Idowu Oluwaseun, Kin-Man Lam
Department of Electronic and Information Engineering, The Hong Kong Polytechnic University, Hong Kong, China

Email: oluwaseun.animasahun@connect.polyu.hk, kin.man.lam@polyu.edu.hk

Abstract – Source camera identification is a fundamental area in forensic science, which deals with attributing a photo to the camera device that has captured it. It provides useful information for further forensic analysis, and also in the verification of evidential images involving child pornography cases. Source camera identification is a difficult and challenging task, especially in cases involving small-sized query images. Recently, many deep learning-based methods have been developed for camera identification, by learning the camera processing pipeline directly from the images captured by the camera under consideration. However, most of the proposed methods have considerably good identification accuracy for identifying the camera models, but much less accurate results on individual or instance-based source camera identification. In this paper, we propose to train an accurate deep residual convolutional neural network (ResNet) for small-sized query images, with the use of curriculum learning (CL) and preprocessed noise residues of camera images, so as to suppress the contamination of camera fingerprints and extract highly discriminative features for camera identification. The proposed ResNet consists of five convolutional layers and two fully connected layers with residual connections. For curriculum learning, we propose a manual and an automatic curriculum learning algorithm. Furthermore, after training the proposed ResNet with CL, the flattened output of the last convolutional layer is extracted to form deep features. These deep features are then used to learn one-vs-rest linear support vector machines for predicting the camera classes. Experimental results on 10 cameras from the Dresden database show that the proposed method achieves better identification accuracy when compared with some state-of-the-art methods.

A Novel Tracking Scheme for Norway Lobster and Knee Replacement Surgery Phase Recognition with Wearable Camera

Syoji Kobashi[1], Shoichi Nishio[1], Belayat Hossain[1,2],
[1]Graduate School of Engineering, University of Hyogo, Hyogo, Japan
[2]JSPS Postroctoral Fellowships, Japan

Email: kobashi@ieee.org

Abstract – Artificial intelligence (AI) assisted surgical operation is one of the six major expected fields of AI application in medicine. Surgical operation is composed of some steps (phases). Especially, scenario of knee replacement surgery is almost fixed, and it consists of about 30 phases. AI may assist not only surgeons but also nurses to perform the scheduled surgery. This paper proposes automated surgery phase recognition from wearable camera. It plays the fundamental role of AI assisted surgical operation system. The method is based on convolutional-long short-term memory (LSTM) network with MobileNet. The proposed method has been evaluated by using six different patient video images. The validation method was leave-one-out-cross-validation (LOOCV). The mean phase recognition rate was 53.7%.

Digital Acquisition and Reconstruction of Artworks

Kyudong Sim[1], Jong-Il Park[1]
[1]Hanyang University, Department of Computer and Software, Seoul, Korea

Email: kdsim@mr.hanyang.ac.kr, jipark@hanyang.ac.kr

Abstract – This paper deals with the digitization and restoration methods based on the physical properties of artwork. Types of physical properties that can be obtained from artworks include color information, geometric information, and reflection characteristics, and artworks can be largely classified into 2D, 2.5D, and 3D in order to divide main physical properties according to artworks. 2D artwork is an artwork like a painting on a plane, and high resolution and accurate color information are important. For high resolution and color information, multiple images are stitched, and multichannel color information is acquired. In 2.5D artwork, BRDF, which is a reflection property, is acquired to express brush strokes in art such as oil painting. 3D artwork is a three-dimensional artwork that is digitized to three-dimensional geometric information. The digital information of the acquired art can be printed by a 2D or 3D printer or virtually reconstructed on a display. The proposed classification and digitization method in this paper is expected to increase the utilization of art works.

A Drone-Projected Image Stabilization: Consideration of Tilt and Scaling

Eunbin Choi, Younghyeon Park, and Byeungwoo Jeon
Department of Electrical and Computer Engineering, Sungkyunkwan University, Suwon, South Korea
Email: {eunbin544, neversky, bjeon}@skku.edu

Abstract – In order to make a movable display projecting onto high-rise walls or arbitrary objects in physical space as a digital signage, a system, called a drone-projector system has been investigated with a beam projector mounted on it. The vibration during hovering caused by the motor of propeller in the drone brings distortion in the projected image. Noting the tilt of the projected image as one source of distortion, in our previous work, we introduced its sensor-based stabilization method. In this paper, we extend the previous work. The contribution of this paper is two-fold. One is the scaling of the projected image by employing LiDAR sensor which measures distance from our drone-projector system to projected surface, and the other is light-weighted hardware configuration of the drone-projector system by removing the gimbal. To evaluate the proposed stabilization method, we use the stabilization assessment method proposed in our previous work. Our experimental result shows that the distortion of the projected image is made smaller by using our stabilization method in this paper.

A Novel Low-Dose Scanning Method for Interventional CT

Seungryong Cho*, Sanghoon Cho
Department of Nuclear and Quantum Engineering, Korea Advanced Institute of Science and Technology, Daejeon, Korea 34141

Email: scho@kaist.ac.kr

Abstract – Interventional procedures in medicine often require CT imaging in operation. Low-dose scanning in such applications is important in consideration of patient dose. We have recently developed a novel low-dose scanning method based on a slip-ring type gantry CT scanner. Not only radiation dose reduction but also dual-energy data acquisitions are possible with the proposed method. A moving beam-filter is placed in between the x-ray source and the patient. We purposely made a slanted-angle multislit filter and translated it along the bore axis direction while scanning. The sinogram would contain streaks as a result, and we developed a new image reconstruction method that can reconstruct a base image free of streak-related artifacts. A notch filter-based masking technique was applied for removing streaks. This will result in quantitatively inaccurate sinogram data. We used the contaminated data though to reconstruct a structure based anatomical image based on which quantitatively more accurate image reconstruction for both low- and high energies are performed in an iterative reconstruction framework. Finally, a material decomposition is performed. Preliminary results are very promising, and the proposed method is thought to be able to provide a low-dose dual-energy CT scanning option in an interventional room.

Blocking Gamelan Instruments Frequency in Virtual Reality

Clemens Felix Setiyawan[1], Mohammad Shobri[1]
[1]Universitas Multimedia Nusantara, Indonesia

Email: clemens.felix@umn.ac.id, shobri@umn.ac.id

Abstract – This study was an experiment that looked at the possibility of VR relationships with the frequency of gamelan in animated puppet shows that were considered based on the blocking instrument. On the other hand, gamelan instruments provide the clearest picture of the frequency spectrum that is affected by various types of treatment, regardless of the ability of the instrument itself to produce sound in many characteristics of each instrument. The key experiment in this study was to arrange the position of the instrument in a different semicircle with a conventional arrangement that is lined up. This experiment is believed by the author to bring its own impact in relation to VR. Gamelan are used Slendro and pelog to get two different data. It can be assumed that this blocking semicircularis able to produce a more focused frequency distribution for the proportion of the listening room and VR users. And in connection with theexperimentation blocking semicircularit can be concluded that sound is more effective and easier to identify.

3D Model Retrieval Based on Deep Learning Approach with Weighted Three-View Deep Features

Xuemei Jiang[1], Yaqi Li[1], Jiwei Hu[1], Kin-Man Lam[2]
[1]School of Information Engineering, Wuhan University of Technology, Hubei Key Lab. of Broadband Wireless Communication and Sensor Network,China
[2]†Department of Electronic and Information Engineering, The Hong Kong Polytechnic University, Hong Kong

Email: jxm2001@whut.edu.cn, lyq199746@hotmail.com,hujiwei@whut.edu.cn, kin.man.lam@polyu.edu.hk,

Abstract–With the development of computer graphics and three-dimensional (3D) modeling technology, 3D model retrieval has been widely used in different applications, such as industrial design, virtual reality, medical diagnosis, etc. Massive data brings new opportunities and challenges to the development of the 3D model retrieval technology. However, with the emergence of complex models, traditional retrieval algorithms are not applicable to some extent. One important reason for this is that the traditional content-based retrieval methods do not take the spatial information of 3D models into account during feature extraction. Therefore, how to use the spatial information of a 3D model to obtain a more extensive feature has become a significant issue. In our proposed algorithm, we first normalize and voxelize the model, and then extract features from different views of the voxelized model. Secondly, deep features are extracted by using our proposed feature learning network. Then, a new feature weighting algorithm is applied to our 3D-view-based features, which can emphasize the more important views of the 3D models, so the retrieval performance can be improved. The experimental results on the standard 3D model dataset, Princeton ModelNet10, show that our model can achieve promising performance.

Discovering Inactive Students patterns and trends by applying Data Warehouse and Visualisation on Campus Student Record

Mohammad Shobri*[a], Abba Suganda Girsang[a], Dyah Ayu Anggreini Tuasikal[b]
[a] BINUS University, Indonesia; [b]Multimedia Nusantara University, Indonesia

Email: mohammad.shobri@binus.ac.id, agirsang@binus.edu, dyah.ayu@umn.ac.id

Abstract - The problem of inactive students is a problem that is often experienced by several campuses in Indonesia. There are several factors that cause students to be inactive which can be categorized into two, namely not officially active and not official. This informal, inactive student can cause other problems. Therefore, the campus requires inactive student data processing to read and provide a visualization of trends so that it can be used as supporting data to make policies so as not to infer other problems. With the application of data warehouse where the ETL process is carried out as a backend process that can provide supporting data for the recording and processing of inactive student data. Visualization of inactive student trends is beneficial for the campus to help display data in an easily understood model so that the campus can learn trends and help the campus determine strategic policies so as to reduce the level of students not being active in lectures.

ADDITIONAL INFORMATION

YOGYAKARTA TOURISM MAP

Conference Venue
Hotel Santika Premiere Jogja

Address: Jl. Jend. Sudirman No.19, Cokrodiningratan, Kec. Jetis, Kota Yogyakarta, Daerah Istimewa Yogyakarta 55233
Phone: (+62 274) 563036

Official Taxi

- **Jas Taxi (+62 274) 373737**
- **Rajawali Taxi (+62 274) 512976**
- **Blue Bird Taxi (+62 274) 6411234**

EMERGENCY CONTACT
BY CALL OR WHATSAPP

Rangga Winantyo
+62 816873438

Andy Firmansyah
+62 8561679472