

UNCALIBRATED CAMERA BASED CONTENT GENERATION FOR 3D MULTI-VIEW DISPLAYS

MANSI SHARMA



DEPARTMENT OF ELECTRICAL ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY DELHI
MAY 2017

©Indian Institute of Technology, Delhi (IITD), New Delhi, 2017

UNCALIBRATED CAMERA BASED CONTENT GENERATION FOR 3D MULTI-VIEW DISPLAYS

by

MANSI SHARMA

Department of Electrical Engineering

Submitted

in fulfillment of the requirements of the degree of Doctor of Philosophy

to the



INDIAN INSTITUTE OF TECHNOLOGY DELHI

MAY 2017

To my mother & teachers...

CERTIFICATE

This is to certify that the thesis entitled, “**UNCALIBRATED CAMERA BASED CONTENT GENERATION FOR 3D MULTI-VIEW DISPLAYS**”, being submitted by Ms. Mansi Sharma for the award of the degree of Doctor of Philosophy is a record of bonafide research work carried out by her in the Department of Electrical Engineering, Indian Institute of Technology Delhi.

Ms. Mansi Sharma has worked under our guidance and supervision and has fulfilled the requirements for the submission of this thesis, which to our knowledge has reached the requisite standard. The results contained in it have not been submitted in part or full to any other university or institute for award of any other degree or diploma.

Dr. Santanu Chaudhury

Professor

Dept. of Electrical Engineering

Indian Institute Of Technology Delhi

New Delhi-110016, India

Dr. Brejesh Lall

Associate Professor

Dept. of Electrical Engineering

Indian Institute Of Technology Delhi

New Delhi-110016, India

ACKNOWLEDGMENTS

Foremost, I would like to express my sincere gratitude to my advisor Prof. Santanu Chaudhury who gave me the freedom to reveal myself, and reflect my own individuality throughout this work. His valuable comments for presenting and communicating the work effectively, guided me to recover when my steps faltered. His patience with my mistakes and constant encouragement help me to build confidence in the creative spirit.

I am indebted to all the teachers of Department of Mathematics, IIT Delhi, with whom I have the privilege to learn during the course of my post-graduate studies. Their teaching helped me to understand the concepts of hardcore mathematics and computing. I take this opportunity to acknowledge Prof. J. B. Srivastava, Prof. Suresh Chandra, Dr. Wagish Shukla, Prof. R. K. Sharma, Prof. S. Dharmaraja, Prof. S. C. S. Rao, Dr. Aparna Mehra, Dr. Anima Nagar, Prof. B. S. Panda, Prof. Niladri Chatterjee, Prof. B. Chandra, Prof. A. Tripathi, Dr. K. Sreenadh, Prof. B. R. Handa, and Prof. A. Nagabhushanam, for their unconditional support and entrusting my ability.

I would especially like to mention Prof. J. B. Srivastava, who has introduced me to computer vision. His special course on “Algebraic Geometry” inspired me to solve a formidable variety problem of vision, which form the basis of Chapter IV of this thesis. I also pay special thanks to Prof. Prem Narain, IIT Bombay, for his directions which opened my way to get into IIT.

I would like to thank my thesis committee members, Prof. S. D. Joshi and Prof. Prem Kumar Kalra, for their valuable comments. I am also thankful to Prof. Dinesh Manocha, University of North Carolina–Chapel Hill and Prof. A. N. Rajagopalan, IIT Madras, for reviewing this thesis and providing valuable suggestions which have led to significant improvement on the quality of this work. I am also grateful to Prof. Anil Kokaram, Trinity College Dublin, for recommending the ideas presented in Chapter V of this thesis to IEEE TCSVT.

My sincere thanks to Dr. Brejesh Lall & Dr. Sumantra Dutta Roy for being with me as an elderly figure throughout. Their personal helps and forbearing nature has always made me feel comfortable with them. I am also grateful to Prof. T. C. Kandpal, Prof. Mamidala Jagadesh Kumar, and Prof. Anshul Kumar that helped me adjust to a new structure of IIT-D during my initial days.

I must mention three old teachers, Dr. Pushpalata Sharma, late Ms. Asha Singh and Jiya Lal Sir, whom I learnt from in my school days. Without their blessing and initial support that helped to survive tough times, I was certainly not able to complete my secondary education.

The space is too small to acknowledge the generous effort of all my loyal friends, Venkatesh,

Prashanth, Rashi, Rashmi, Chakresh, Mahesh, Pratyush, Shalini, Vandana, Amrita, Shefali, Shweta, Ritu, Ravi, Kumar Shubham, Mrinal who made my stay at IIT-D pleasant & memorable. I want to especially thank Andleeb Zahra, my lifelong companion and ex-room partner, with whom I share a very special bond. Despite the distance and endured my rants, she has never judged me and stayed in touch.

Many physicians and spiritual teachers have helped me stay sane through all difficult year of continuous fight with arthritis. I deeply appreciate Dr. S. V. Tripathi, Dr. S. M. Tuli, Dr. Seema Ahlawat, Dr. I. P. Tripathi, Dr. Rajesh Malhotra, Dr. Deepak Gautam, Dr. Vijay Kumar D, Brahmakumari Dadi Janki and Sister Shivani for their love and care that helped me overcome setbacks. Their spiritual healing determined me to stay focused on my study.

The person with the greatest indirect contribution in my life, particularly to this work, is my mother, Dr. Urmila Sharma. She has patiently endured long hours alone while I worked on my thesis. Without her genuine care and persistent, I may not have gotten to where I am today, at least not sanely.

Despite life's challenges and the differences that followed, this thesis could always be my number one priority. At any rate, I have completed, but not alone. I could not have succeeded without the invaluable support of my father, Mr. Udai Chand Sharma, my elder brother, Mayank and my sister-in-law, Anamika. My father, even after his retirement, remained a supporter and provided insight right up to the end. For this, I cannot thank them enough. I am forever grateful.

Lastly, I must mention about my little nephew, Divik, who constantly asked me "Have you done your homework yet ?" and mimicked me to be very slow and affectionately referred to me as his best friend. His adorable derision kept things light and me smiling.

Mansi Sharma

mansi sharma

ABSTRACT

The demand for 3D TV systems is going high and technology is rapidly improving. High quality 3D content production is crucial for working on novel ways to show glasses-free 3D. For creating a truly immersive experience, it is essential to support advanced functionalities like free-viewpoint viewing of natural video. Other multi-media features which increase user interactivity with television content, like editing or mixing of scene components, virtual panning, tilting or zoom-in, a video featuring visual 3D effects as frozen movement, etc., must also be realized. It is also desirable for users to enjoy 3D vision with an increased field of view.

Supporting these requirements is one of the key issues, using only a limited number of real cameras. Capturing arbitrary cameras using huge multi-camera rigs is tedious (and costly). It causes inflexibility in the acquisition of the natural environment. Although standardised Multi-view video plus depth (MVD) data format, allows to create virtual views using depth-image-based rendering (DIBR) in calibrated setting. But synthesizing a large number of intermediate views, for high-quality free-viewpoint video, is challenging. DIBR is known to suffer from inherent visibility and resampling problems. This resulted in appearance of artifacts like holes, cracks, corona, etc., in synthesized images. Existing techniques reduces rendering artifacts only from nearby viewpoints, but are inadequate to deal with arbitrary virtual camera movements. Besides, DIBR fails to provide multimedia functionalities such as graphical realism.

This thesis proposed novel architectures based on uncalibrated cameras for multi-view 3D TV systems that supports the above mentioned features. We first proposed a signal representation that improves the content generation pipeline and the interactivity of DIBR. It is based on MVD which is being standardized by MPEG. We introduced novel DIBR algorithms based on a segmentation cue that address the rendering problems encountered in creating special effects and wide baseline extensions. We also proposed a 3D warping scheme for reduction of computation cost of DIBR rendering.

We designed another image-based system for multi-view 3D TV based on a full-perspective parameterized variety model. An efficient algebraic scheme is proposed that addresses the

problem of characterizing the set of images of a 3D scene. The system is flexible to take the input from uncalibrated handheld cameras. Image-based parameterization of scene space allows to render high-quality virtual views from arbitrary viewpoints without 3D model and using few sample images. The proposed signal representation and free-viewpoint rendering method overcome major shortcomings of geometry-based methods, where performance is seriously affected due to scene complex conditions.

Further, we proposed a new high-quality multi-view rendering scheme for glasses-free 3D TV by integrating Kinect with the parameterized variety model. This scheme does not need dense depth, any hardware modification or active-passive depth imaging modalities to recover missing depth. In fact, it presents a new way to deal with the challenges inherent with DIBR and noisy depth estimates.

Towards the end, we proposed a new approach to 3D images that promises glasses-free multiple-perspective 3D. We answered an important question “How to represent the space of a broad class of perspective and non-perspective stereo varieties within a single, unified framework?”. We presented a representation that combines the geometric space of multiple uncalibrated perspective views with the appearance space of manifolds in a globally optimized way. Our approach works for uncalibrated static/dynamic sequences and render high-quality content for multi-perspective 3D TV. Finally, we addressed the challenges in creating good quality composite 3D for postproduction visual-effects. Our proposed DIBR compositing technique is regularized to handle large warps, vertical disparities, and stereo baseline changes. Overall, all proposed schemes are backwards-compatible with state-of-the-art MPEG tools and existing image-based rendering systems.

सार

3 डी टीवी सिस्टम की मांग बढ़ रही है और तकनीक तेजी से सुधर रही है। उच्चगुणवत्ता वाले 3 डी सामग्री का उत्पादन चश्मे से मुक्त 3 डी दिखाने के लिए और नई तरीकों पर काम करने के लिए महत्वपूर्ण है। सही मायने में अनुभव बनाने के लिए, फ्री विएवपॉइंट वीडियो का समर्थन करना आवश्यक है। अन्य बहु-मीडिया विशेषताएं जो बढ़ाती हैं उपयोगकर्ता की इंटरैक्टिविटी टेलीविज़न सामग्री के साथ, जैसे दृश्य घटकों के संपादन या मिश्रण, वर्चुअल पैनिंग, झुकाव या जूम-इन, एक वीडियो जो दृश्य 3 डी प्रभावों को फ्रोजेन मूवमेंट आदि रूप में दिखाता है, भी एहसास होना चाहिए। यह भी वांछनीय है कि उपयोगकर्ताओं को 3 डी विजन का आनंद बढ़ने के साथ मिलना चाहिए बड़े देखने के क्षेत्र।

केवल वास्तविक कैमरों की सीमित संख्या का उपयोग करते हुए इन आवश्यकताओं का समर्थन करना प्रमुख मुद्दों में से एक है। मल्टी कैमरा रिंग्स का उपयोग करके मनमाने ढंग से कैमरे कैप्चर करना कठिन है (और महंगा है)। यह प्राकृतिक पर्यावरण के एकीकरण में इन्फ्लेक्सिबिलिटी का कारण है। हालांकिमानकीकृत बहु-दृश्य वीडियो प्लस गहराई (एमवीडी) डेटा प्रारूप, वर्चुअल व्यूज बनाने की अनुमति देता है कैलिब्रेटेड सेटिंग में और डेप्थ इमेज बेस्ड रेंडरिंग (डीआईबीआर) का उपयोग करते हुए। लेकिन उच्च-गुणवत्ता वाले फ्री विएवपॉइंट वीडियो के लिए बड़ी संख्या में इंटरमीडिएट दृश्यों को संश्लेषित करना चुनौतीपूर्ण है। डीआईबीआर को इन्हेरेंट विजिबिलिटी और रेसंपलिंग समस्याओं से पीड़ित होने के लिए जाना जाता है। इसके परिणामस्वरूप कृत्रिम चित्रों में छेद, दरारें, कोरोना, आदि जैसे आर्टिफैक्ट्स की उपस्थिति होती है। मौजूदा तकनीके केवल नज़दीकी दृष्टिकोण से आर्टिफैक्ट्स का प्रतिपादन कम करती हैं, लेकिन मनमाने वर्चुअल कैमरा मूवमेंट्स को समर्थन देने के लिए अपर्याप्त हैं। इसके अलावा, डीआईबीआर मल्टीमीडिया कार्यों जैसे कि ग्राफिकल रेअलिस्म प्रदान करने में विफल रहता है।

यह शोध उपर्युक्त सुविधाओं का समर्थन करने वाले मल्टी-व्यू 3 डी टीवी सिस्टम के लिए अनकैलिब्रेटेड कैमरे पर आधारित आर्किटेक्चर प्रस्तुत करता है। हम सबसे पहले एक सिग्नल का प्रतिनिधित्व करते हैं जो सामग्री बनाने वाली पाइपलाइन और डीआईबीआर की अन्तरक्रियाशीलता में सुधार लाती है। यह एमवीडी पर आधारित है

जिसे एमपीईजी द्वारा मानकीकृत किया जा चुका है। हमने नवीन डीआईबीआर एल्गोरिदम की शुरुआत की जो एक दृश्य विभाजन के क्यू पर आधारित है और विशेष प्रभावों और व्यापक बेसलाइन एक्सटेंशन बनाने में हुई समस्याओं को संबोधित करते हैं। हमने डीआईबीआर रेंडरिंग की गणना लागत में कमी के लिए 3 डी वर्पिंग योजना का भी प्रस्ताव किया है।

हमने एक फुल पर्सपेक्टिव पैरामीटराइज्ड वैरायटी मॉडल के आधार पर मल्टी-व्यू 3 डी टीवी के लिए एक और छवि-आधारित प्रणाली तैयार की है। एक प्रभावी अलजेब्राइक योजना प्रस्तावित की गई है जो एक 3 डी दृश्य की छवियों के सेट को दर्शाने की समस्या को हल करती है। सिस्टम अनकैलिब्रेटेड हैंडहेल्ड कैमरे से इनपुट लेने में सक्षम है। दृश्य-स्थान का छवि-आधारित पैरामीटराइजेशन 3 डी मॉडल के बिना मनमाना दृष्टिकोण से उच्च-गुणवत्ता वाले आभासी दृश्य प्रस्तुत करने और कुछ नमूना चित्रों का उपयोग करने की अनुमति देता है। प्रस्तावित सिग्नल प्रस्तुतीकरण और फ्री विएवपॉइंट दृष्टिकोण को प्रस्तुत करने की पद्धति ज्योमेट्री-आधारित तरीकों की प्रमुख कमियों को दूर करती है, जहां दृश्य प्रदर्शन जटिल परिस्थितियों के कारण गंभीर रूप से प्रभावित होता है।

इसके अलावा, हमने पैरामीटराइज्ड विविधता मॉडल के साथ केनक्ट को एकीकृत करके चश्मे से मुक्त 3 डी टीवी के लिए एक नई उच्च-गुणवत्ता वाली बहु-दृश्य रेंडरिंग योजना प्रस्तावित की। इस योजना को गहन गहराई, किसी भी हार्डवेयर संशोधन या सक्रिय-निष्क्रिय गहराई इमेजिंग विधियों की आवश्यकता नहीं है ताकि लापता गहराई ठीक हो सके। वास्तव में, यह डीआईबीआर और गहराई के अनुमानों में दोष से निहित चुनौतियों से निपटने का एक नया तरीका प्रस्तुत करती है।

अंत में, हमने 3 डी छवियों के लिए एक नया दृष्टिकोण प्रस्तावित किया है जो चश्मे से मुक्त बहु-परिप्रेक्ष्य 3 डी का वादा करता है। हमने एक महत्वपूर्ण प्रश्न का उत्तर दिया " छवि-स्थान का प्रतिनिधित्व कैसे करें एक व्यापक, एकीकृत रूपरेखा के भीतर परिप्रेक्ष्य और गैर-परिप्रेक्ष्य वाली स्टीरियो किस्मों के एक व्यापक श्रेणी का? " हमने एक सिग्नल प्रस्तुत किया है जो वैश्विक स्तर पर अनुकूलित तरीके से मैनिफोल्ड के अपीयरेंस-स्पेस के साथ कई अनकैलिब्रेटेड परिप्रेक्ष्य दृश्यों की जियोमेट्रिक-स्पेस को जोड़ता है। हमारी विधि अनकैलिब्रेटेड स्थिर/गतिशील दृश्यों के लिए काम करती है और

बनाती हैं बहु-परिप्रेक्ष्य 3 डी टीवी के लिए उच्च गुणवत्ता वाली सामग्री। अंत में, हमने पोस्ट उत्पादन दृश्य-प्रभावों के लिए अच्छी गुणवत्ता वाला 3 डी बनाने के लिए चुनौतियों का समाधान किया। हमारे प्रस्तावित डीआईबीआर कंपोजिटिंग तकनीक को बड़े वार्प्स, ऊर्ध्वाधर असमानताओं और स्टीरियो बेसलाइन परिवर्तनों को संभालने के लिए नियमित किया जाता है। कुल मिलाकर, सभी प्रस्तावित योजनाएं पीछे-के-अत्याधुनिक एमपीईजी उपकरण और मौजूदा छवि-आधारित रेंडरिंग सिस्टम के साथ संगत हैं।

TABLE OF CONTENTS

CERTIFICATE	i
ACKNOWLEDGMENTS	iv
ABSTRACT	vi
LIST OF FIGURES	xiii
LIST OF TABLES	xix
LIST OF ACRONYMS	xxi
LIST OF MATH SYMBOLS	xxiii
CHAPTER I: INTRODUCTION	1
CHAPTER II: 3D TV CONTENT REPRESENTATION AND PROCESS- ING: A REVIEW	9
2.1 Conventional stereo video (CVS)	9
2.2 Multi-view video (MVV)	10
2.3 Surface-based representations	10
2.4 Image-based rendering	11
2.5 Dense depth-based representation	12
2.6 Uncalibrated Camera Set-ups: Motivation and Challenges	20
CHAPTER III: UNCALIBRATED CAMERA-BASED FLEXIBLE ARCHI- TECTURE FOR MULTI-VIEW 3D TV	25
3.1 Flexible architecture for multi-view 3D TV and free-viewpoint video	26
3.2 Signal generation	28
3.3 Virtual view synthesis	39
3.4 Proposed algorithms for free-viewpoint rendering	43
3.5 Selective warping	47
3.6 Experimental evaluations	49
3.7 Conclusion	63
CHAPTER IV: PARAMETERIZED VARIETY-BASED ARCHITECTURE FOR MULTI-VIEW 3D TV SYSTEMS	71
4.1 Parameterized variety-based versatile architecture for 3D TV	72
4.2 Signal generation	74
4.3 Stage I – Full-perspective scene parameterization	74
4.4 Stage II – Scene classification and surface labeling	92
4.5 Multiview 3D content creation	93
4.6 Virtual stereoscopic 3D video composition	102
4.7 Experimental evaluations	111
4.8 Quality assessment of image-based rendering	122
4.9 Conclusion	128

CHAPTER V: HYBRID KINECT-VARIETY BASED MULTI-VIEW RENDERING SCHEME FOR GLASSES-FREE 3D DISPLAYS	131
5.1 HYBRID SCHEME FOR MULTI-VIEW 3D TV	133
5.2 STAGE I - Image points selection and associated view-dependent depth value extraction	134
5.3 STAGE II - Kinect-variety fusion	137
5.4 Degenerate configurations	144
5.5 Multi-view rendering	149
5.6 Experimental results	156
5.7 View synthesis quality evaluation	161
5.8 Non-linear FP-PIV vs Hybrid linear FP-PIV formulation	165
5.9 Conclusion	166
CHAPTER VI: VARIETY-BASED 3D TV CONTENT GENERATION SCHEME FOR CASUALLY CAPTURED SPARSE PHOTO COLLECTIONS	169
6.1 Parameterized scene graph construction	170
6.2 Consistent parameterization of scene graph	172
6.3 Computing final tour	175
6.4 Results and experiments	175
6.5 Assessment of rendering quality	176
6.6 Limitations of scheme	177
6.7 Conclusion	178
CHAPTER VII: SPACE-TIME PARAMETERIZED VARIETY MANIFOLDS	181
7.1 Overview of proposed approach	184
7.2 Stage I-II: Parameterization variety representation and virtual video rendering	186
7.3 Stage III: Space-time parameterized variety manifolds	187
7.4 Graph construction: optimal parameterized variety manifold rendering	200
7.5 Results	203
7.6 Conclusion	206
CHAPTER VIII: CONTENT AWARE FREE-VIEWPOINT STEREOSCOPIC 3D COMPOSITING SCHEME	209
8.1 System overview	212
8.2 Proposed content-aware 3D compositing	215
8.3 Adjusting source object perspective	225
8.4 Content-aware consistent local warps for free-viewpoint compositing	226
8.5 Shadow Synthesis	235
8.6 Results	237
8.7 Conclusion	242
CONCLUSIONS & FUTURE RESEARCH DIRECTIONS	245
BIBLIOGRAPHY	255
LIST OF THESIS PUBLICATIONS	275
APPENDIX A: DEPTH-IMAGE BASED RENDERING	279
APPENDIX B: FP-PIV MULTIVARIATE QUADRATIC EQUATIONS	283
APPENDIX C: EUCLIDEAN CONSTRAINTS ASSOCIATED WITH FULL PERSPECTIVE CAMERAS	287
BIO-DATA OF THE AUTHOR	291

LIST OF FIGURES

2.1	(a) Holes, cracks caused by visibility, disocclusion and resampling problems associated with 3D warping. The size of artifacts increases in wide baseline extensions (b) and arbitrary virtual camera movements such as 3D rotation, translation and zoom in/out effect (c,d).	14
3.2	A flexible uncalibrated camera based multi-view 3D TV system.	27
3.3	Scene classification of “Charlotte-Potsdam I”: (b) Assigned surface labels of the original view (a). Brown corresponds to ground (horizontal). Cyan, dark blue, light blue corresponds to left, right, and center facing. (c-g) Assigned confidence measures with each surface label. (h) Assigned geometric class labels (green (support region), red (vertical), blue (sky)), surface orientations (arrows), non-planar solid (‘x’) and porous (‘o’).	28
3.4	Example scenes and their classification geometric labels	31
3.5	Depth estimation steps: (b) Plane family labels before sweeping and graph cut (GC). (c) Plane family labels after sweeping and graph cut minimization. (d) Estimated depth of original view (a). (e) Depth map provided by [41].	36
3.6	Multiple objects in a scene.	39
3.7	Warping technique for scene content based virtual view synthesis when original cameras are in small baseline multi-view stereo setup.	40
3.8	Warping technique for scene content based virtual view synthesis when multiple input views are captured by arbitrary placed cameras.	41
3.9	Novel view synthesis of “Charlotte-Potsdam I” using method 3.3.2: (a-b) Warped virtual view and plane labels (<i>notice the crack and holes</i>). (c) After nearest neighbour filling of labels. (d) Filled virtual image depth map. (e) Final virtual image after inverse warping the cracks and holes. (f-h) Rendered novel views of “Charlotte-Potsdam I” using method 3.3.2 when the left wall of the scene is selected.	42
3.10	Features of the proposed architecture.	43
3.11	3D scene augmentation: (a,b) Original frames of “Warrior” datasets. (c,d) Extracted object masks. (e) Warp extracted object (a) at a novel viewpoint with respect to the reference camera viewpoint (a). (f) Artifacts removed by hole-filling. (g) Warp original view (a) at another virtual viewpoint. (h) Final rendered view. (i) Augment object (a) in a novel scene (h) through α -blending and Z-buffering. (j,k) Final augmented views.	43
3.12	Proposed segment-based warping technique for free-viewpoint rendering.	44
3.13	Example of virtual view synthesis using <i>algorithm II</i> 3.4.2 for rendering a virtual zoomed view of “Charlotte-Potsdam I” : Three virtual image clusters after warping (1^{st} , 2^{nd} and 3^{rd} image). Clusters after hole-filling, inverse warping and merging (4^{th} image). Final rendered view after disocclusion filling (5^{th} image).	45

3.14	Virtual view rendering using <i>alg. II</i> 3.4.2 (1 st & 2 nd row) and <i>alg. III</i> 3.4.3 (3 rd row) for synthesizing camera 4 th using cameras 3 rd and 5 th respectively: (a-c) Three virtual image (VIMG) clusters (out of five) after warping right reference camera. (d-f) Clusters after hole-filling, inverse warping and merging left-right segments (<i>alg. II</i>). (g) Final rendered view by <i>alg. II</i> . (h) Extracted local portion of (g). (i,j) Two virtual right image clusters (out of seven) after over-segmentation of depth map, plane fitting, hole-filling and merging left-right segments (<i>alg. III</i>). (k) Final rendered virtual view after Z-buffering all segments and disocclusion filling using <i>alg. III</i> . (l) Extracted local portion of final view (k). Over segmentation and plane fitting improves the quality of virtual synthesized view as seen in (l) around the legs and hands of the dancer.	47
3.15	Reprojection error in pixels after calibration of different datasets, with (WSC) or without (WoSC) imposing scene constraints.	50
3.16	Depth estimation results: Classified frames of “Stone”, “Alt-Moabit”, “Charlotte-Potsdam II” and their estimated depth maps (first six views). Compared to the depth maps provided by [41] (last three views in second row), recovered depth using the classified scene data 3.2.1 is much better. Recovered depth of frame 8 th of camera 0 (“Breakdancers”) and frame 5 th of camera 7 (“Ballet”) (row third).	51
3.17	Estimated depth analysis results	53
3.18	Rendered virtual zoomed views of “Merton II” using <i>algorithm II</i> 3.4.2 when the viewer is intended to see the center of the scene (row first). Rendered free-viewpoint images of “IkB” using <i>algorithm III</i> . Experiment on a piecewise planar scene which does not meet the scene assumptions as described in section 3.1 (second row). Classified view and its estimated depth (2 nd and 3 rd view in second row). Warped zoomed view (notice the large holes and cracks) and the final rendered view using <i>algorithm III</i> (last two views in second row).	54
3.19	(a, c) Mild ($\sigma_h, \sigma_v, w_h, w_v$) = (4, 4, 13, 13) and (b, d) strong ($\sigma_h, \sigma_v, w_h, w_v$) = (30, 30, 61, 61), ($\sigma_h, \sigma_v, w_h, w_v$) = (20, 20, 61, 61) symmetric smoothing. (e, g) Mild ($\sigma_h, \sigma_v, w_h, w_v$) = (4, 12, 13, 41) and (f, h) strong ($\sigma_h, \sigma_v, w_h, w_v$) = (20, 60, 61, 193) asymmetric smoothing. (i) Discontinuity-preserving and gradient direction-based smoothing [35]. (j) Edge-oriented smoothing [28]. (k-m) Synthesis results of different interpolation and inpainting methods. (n, o) Synthesis results of the proposed algorithms <i>II</i> and <i>III</i> respectively.	55
3.20	Comparative rendering results at varying camera baselines and under arbitrary virtual camera motion: (a, f, k, p) Current baseline (3 → 4 ← 5). (b, g, l, q) Baseline doubled (2 → 4 ← 6). (c, h, m, r) Baseline tripled (1 → 4 ← 7). (d, i, n, s) Arbitrary virtual zoomed views at a novel viewpoint. (e, j, o, t) Frozen moment of camera 4 th by arbitrary zooming and rotating camera view direction.	56
3.21	Comparative rendering results at varying camera baselines and under arbitrary virtual camera motion: (a, f, k, p) Current baseline (3 → 4 ← 5). (b, g, l, q) Baseline doubled (2 → 4 ← 6). (c, h, m, r) Baseline (1 → 4 ← 5). (d, i, n, s) Arbitrary virtual zoomed views at a novel viewpoint. (e, j, o, t) Frozen moment of camera 4 th by arbitrary zooming and rotating camera view direction.	57
3.22	Performance of the proposed view synthesis algorithms in comparison with different rendering methods.	62
3.23	Comparison with time-consistency based algorithm [47].	63
3.24	Selective warping: Rendering camera 1 using cameras 0 and 2 for two consecutive frames 7 th and 8 th . (a,b) Warped virtual image obtained from left & right camera (for reconstructing frame 7 th of camera 1). (c) Resynthesized frame 7 th . (d,e) Changed pixels in the left and right frame of camera 2 w.r.t camera 0. (f,g) Unwarped virtual images (a) and (b). (h,i) Warp only the changed macroblocks onto the virtual viewpoint (for reconstructing frame 8 th). (j) Final image (resynthesized frame 8 th) after adding the remaining macroblocks and merging (h) and (i).	64

3.25	Results of selective warping for synthesizing virtual views using a single input video sequence. (a,b) Original 6 th and 7 th frame of camera 1. (c) Virtual view obtained after warping input view (a). (d) Synthesized virtual view obtained from original camera (a). (e) Changed pixels in frame 7 th w.r.t frame 6 th of camera 1. (f) Pixels in the previous virtual image (c) to be unwarped. (g) Unwarped virtual image (c). (h) Warp only the changed macro-blocks onto the virtual viewpoint from reference view (b) (for synthesizing a novel view from frame 7 th). (i) Virtual view obtained after copying other macroblocks. (j) Synthesized virtual view (using original view (b)).	64
3.26	Selective warping results on “Breakdancers” and “Redskirt” sequence: (a-d) Rendering of camera 1 using cameras 0 and 2 for 35 consecutive frames. (e-h) Rendering of novel views at viewpoint of camera 2 using frames of camera 1 for 35 consecutive frames. The warping procedures are shown in fig. 3.24 and fig. 3.25 respectively.	65
3.27	Rendered views of “Lab” data using proposed over-segmentation based DIBR algorithm 3.4.3.	66
3.28	Segmentation-based view rendering (3.4.2) increases overhead cost for high quality output.	66
4.29	Overview of proposed full-perspective variety-based generalised flexible architecture for multi-view 3D displays.	73
4.30	Classification output of “Poznan Street”.	74
4.31	Geometric setup of full-perspective variety model.	75
4.32	Degenerate scene and camera configuration.	85
4.33	Examples of classified geometric labels.	92
4.34	Rendered novel translatory views of “Midd1” data along x direction.	94
4.35	Rendered novel views of “Lab” sequence along $x - y$ direction.	94
4.36	Synthesized novel views of “Garden” sequence along different $x - y - z$ direction.	94
4.37	Synthesized novel arbitrary views of “Statue” sequence.	96
4.38	Resolving occlusion or visibility estimates through point ordering in projective oriented geometry.	100
4.39	Proposed parameterized virtual view-based novel stereo video compositing or editing pipeline.	103
4.40	Geometric relation between the supporting ground surface and the capturing camera viewpoint.	105
4.41	Support (ground) surface orientation correction. (a) Input source (left), (b) Input target (left), (c) Without alignment, (d) Novel background, (e) With alignment.	106
4.42	Composited object scales, size and depth correction.	106
4.43	3D transform for object positioning.	107
4.44	Proposed content-sensitive mean-value compositing vs mean-value coordinates (MVC) technique [131].	110
4.45	Comparision with Fusiello et al. [160, 164] and Colombari et al. [48]	113
4.46	3D reconstruction results of “Yellow house” and “Aquarium”.	115
4.47	Rendering results of Silhouette aware warping [167].	115
4.48	Rendering results of Silhouette aware warping [167].	116
4.49	Proposed FP-PIV synthesized high quality views of “Street” dataset.	116
4.50	View synthesis output from “Cottage Tree” dataset (d) Hsu et al. [43] (left); Proposed FP-PIV (right).	118
4.51	View synthesis output from “Angkorwat” dataset: Hsu et al. [43] (left); Proposed FP-PIV (right).	118
4.52	View synthesis output from “Statue” dataset: Hsu et al. [43] (left); Proposed FP-PIV (right).	118
4.53	Comparision with Zhao et al. [169] view rendering method.	120
4.54	Comparision with Cheng et al. [47] and Mobile 3DTV [41].	121
4.55	Full reference quality assessment in comparision with time-consistency based algorithm [47] and Mobile 3DTV [41].	123
4.56	Histogram of MSCN coefficients.	124

4.57	3-D scatter plot between shape, left scale and right scale obtained by fitting AGGD to horizontal paired products of original and synthesized views.	125
4.58	Natural scene statistic-based distortion-generic blind image quality assessment (BRISQUE) scores.	126
4.59	Natural scene statistic-based blind spatio-temporal BLIINDS scores.	127
5.60	Kinect depth estimation problems: Interference of projected dot patterns, transparency, specularity, and natural sunlight.	132
5.61	Workflow of hybrid Kinect-variety based high quality multi-view rendering scheme.	134
5.62	Corresponding image points depth value estimation using local processing of depth maps.	135
5.63	Average run times obtained for a number of calibration images.	136
5.64	Degenerate scene configuration.	144
5.65	Resolving occlusion or visibility estimates during arbitrary view generation through point ordering in projective oriented geometry.	153
5.66	Geometry of four views for FP-PIV rendering of a new viewpoint.	154
5.67	Kinect depth estimation - surface reflection, dark objects under flat viewing angle, scene high detail and complex depth structure, shadow, environmental clutter causes problem.	157
5.68	Rendered novel views of “Lab” data: (I^{st} row) proposed hybrid Kinect-variety approach; (II^{nd} row) Ham et al. [192]; (III^{rd} row) Algorithm III 3.4.3, Cluster size 2; (IV^{th} row) Algorithm III 3.4.3, Cluster size 5; (V^{th} row) MPEG VSRS [193].	159
5.69	Arbitrary novel views of “Patio” dataset: (I^{st} row) Hybrid Kinect-variety approach and Ham et al. [192] (2^{nd} row).	160
5.70	Arbitrary novel views of “Lounge” dataset using Hybrid Kinect-variety approach (row I^{st}) and using Alg. II 3.4.2 (row II^{nd}).	160
5.71	Histogram of MSCN coefficients.	162
5.72	(a,b) BRISQUE scores, (c,d) 3D scatter plot between shape, left scale and right scale obtained by fitting AGGD to horizontal paired products of original and synthesized views, (e,f) BLIINDS scores.	164
5.73	(a,b) Correlation analysis results of BRISQUE and BLINDS predicted scores with human opinions; (c) full-reference quality assessment.	165
5.74	Qualitative comparative analysis of non-linear FP-PIV vs Hybrid linear FP-PIV formulation.	166
6.75	3D reconstruction result of Achyuta DevaRaya Gate from a sparse collection of 25 high resolution images using structure from motion [223].	170
6.76	Parameterized variety based 3D exploration model	171
6.77	Context-appearance refinement process for best portal identification	172
6.78	Parameterized photo tours: Different novel intermediate and arbitrary 3D views in various tours rendered for mandapas inside “Vittala” complex. Movies show smooth transitions with substantial parallax.	174
6.79	3-D scatter plot between shape, left scale and right scale obtained by fitting AGGD to horizontal paired products of original and synthesized views.	177
6.80	Natural scene statistic-based distortion-generic blind image quality assessment (BRISQUE) scores.	177
6.81	Examples of bad photos.	178
7.82	Multi-view rendering; blue: base trajectory and base camera viewpoints; orange and green: virtual camera paths; pink and green: synthesized virtual camera viewpoints.	182
7.83	Overview of parameterized variety manifold scheme for multi-perspective 3D content generation.	184
7.84	Blue lines denote the strips and dots (circles) represent the ‘variety points’ obtained by solving the multi-variate quadratic polynomials for novel cameras.	188

7.85	N_p^* : region inside pink area; it is satisfactory for the algorithm for constructing variety-appearance manifold to converge to this region. $N_p^\dagger$ is the region outside cream area; it is expected that the manifold should not be initialized in this region. N_p is the cream region around N_p^* where the algorithm enforces variety-appearance constraint on the desired manifold.	192
7.86	Seamlessly interconnecting substantially different views of unstructured long videos.	200
7.87	(a-d) Sample images of “head” sequence [14] captured by moving a camera on a circular path. (e) Cyclograph generated by Seitz and Kim [14] using original video. (f) Multi-perspective view synthesized using the proposed method by cutting the x - y - t volume of virtual synthesized video. (g) Artistic neocubism style multi-perspective view rendered using our framework.	202
7.88	Synthesized novel stereo views (a-f)(g-l) of the scene from arbitrary viewpoints. Occlusion and perspective are correctly handled in PIV rendered views. (m-n) Novel multi-perspective stereo views obtained using the proposed variety-appearance manifolds by stacking rendered stereo sequence. Unwanted objects are removed from the scene, preserving the background geometry correctly.	202
7.89	(1 st row) Some frames of original video with horizontal and vertical parallax. (2 nd) Distortion free multi-perspective view generated using proposed technique. Minute details are preserved, perspective is correctly computed between regions with depth variation. No duplication or truncation of objects is observed.	203
7.90	(1 st row): Video synopsis (induced by optimal variety manifolds) of famous movie “Elephants Dream” at different time instants. (2 nd , 3 rd , 4 th row): Rendered seamless multiscale (at different virtual zoom levels) tapestry of famous movie “Sintel”. The movies contain highly dynamic changing scenes with unstructured motion and complex non-planar geometry. (<i>Courtesy of the Blender Foundation</i>)	203
7.91	A composite scene of “Tree Garden” sequence is created using variety-appearance manifold. This scene is portrayed by cutting the space-time volume of video frames with multiple overlapping regions. Our algorithm could interpret the “tree bark” in front of the scene in captured video as moving object. Essence of the scene is preserved here. However, some artifacts are visible in rendered composite where local smoothness of minute details is lost due to mix of content while retaining information from neighboring images.	204
8.92	Conventional 2D composition challenges vs proposed content-aware 3D compositing.	210
8.93	(a) Vague boundaries (b) Absence of partial occlusions due to lack of fine depth structures leads the copied object appears to float.	212
8.94	(c) Effect of warping with planar proxy geometry. The man is distorted due to large orientation changes during 3D warping. (d) To create a realistic looking view, camera mismatch is corrected with both new background synthesis and foreground object surface alignment. The compositing is performed to compensate depth-inaccuracies.	212
8.95	Content aware free-viewpoint stereoscopic 3D compositing	213
8.96	Example of classified scenes used in compositing.	214
8.97	Computed stereo saliency aware objectness prior.	216
8.98	Misclassified geometric labels due to dominant surrounding surfaces, small object or shadow.	218
8.99	Considerable disparity variation in between a standing man and around his surrounding surfaces. But disparity changes within the ground floor or walls are smooth (3 rd image). There is no specific treatment is needed in our proposed scheme to treat the abruptness change and smooth disparity variations differently (4 th image: recovered 3D saliency map).	220
8.100	Proposed stereo saliency analysis results.	220
8.101	Screen parallax and comfort disparity zones.	222
8.102	Scale and orientation adjustment: (a,b) User marked object support surfaces, (c) Rendering without background adjustment, (d) slope is compensated with the new synthesized target background. The pasted object in (c) is floating in air. In (d) it is well accommodated with target scene geometry.	226

8.103(a) Direct depth merging and 3D warping leads to visible discontinuities and resampling artifacts. (b) No rendering artifact in proposed consistent depth merging with local warping.	228
8.104 Prominent lines of toy’s legs in the view are not parallel to the main axes. There is no visual distortion appear in different degrees of squeezing of adjacent mesh triangles.	229
8.105 Proposed content-aware mean value compositing.	234
8.106 Proposed content-aware MVC vs instant MVC [131].	235
8.107(a) Depth map used for shadow synthesis. (b) Object and the underlying surfaces are rendered into the depth buffer.	237
8.108 Comparson of our proposed content-based stereo saliency analysis with other 2D saliency algorithms.	238
8.109 Comparson of our proposed content-based stereo saliency analysis with other 2D saliency algorithms.	239
8.110 Precision-recall curves of salient object detection problem.	239
8.111 Composite “Helicopter” scene.	240
8.112 Composite “Flying” scene. No color bleeding artifacts appear in final composite scene (e). In this example, we consider target camera movement on the z-axis. Therefore, the movement of composite objects or their source cameras is inferred in the direction of the z axis. We create a depth map that compensates the movement.	240
8.113 Content-aware compositing on “Teddy”, “Bunny”, “Toy”, “Playing dog” scenes. (c) Composite looks natural with proper alignment and rendered contact shadow of warped Teddy from source (a). (f) Partial occlusion due to grass leaves is rendered correctly. (g,h) Scenes look real with synthesized shadows along with objects.	241
8.114 Content-aware compositing on “Herz” and “Butterfly” data. In both scenes, (row 1 st , 2 nd , 3 rd), view interpolation is performed to achieve better results. In Herz data (row 1 st and 2 nd), novel intermediate views are synthesized to plausibly transfer the object region in wide baseline orientation. (4 th) Vertical disparity is reduced in 3D composite view (2 nd anaglyph view) with view interpolation.	242
8.115 Content-aware compositing on “Firework” data. Firework has intricate silhouettes with transparency. This case is difficult to handle with segmentation-based approaches.	243

LIST OF TABLES

3.1	Features computed on superpixels	32
3.2	Frame differential flicker measure (FDF)	61
5.3	Standard deviation of the reprojection errors	136
6.4	Best portal identification	176

LIST OF ACRONYMS

FTV	<i>Free-viewpoint television</i>
FVP	<i>Free-viewpoint interpolation</i>
FVV	<i>Free-viewpoint video</i>
DIBR	<i>Depth image-based rendering</i>
PIV	<i>Parameterized image variety</i>
FP-PIV	<i>Full-perspective parameterized image variety</i>
IBR	<i>Image-based rendering</i>
MVV	<i>Multi-view video</i>
nD	<i>n dimensional</i>
MVD	<i>Multi-view video plus depth</i>
LDV	<i>Layered depth video</i>
LDI	<i>Layered depth image</i>
AFX	<i>Animation Framework eXtension</i>
PGS	<i>Projective grid space</i>
MPEG	<i>Moving Pictures Expert Group</i>
3DV	<i>3D video</i>
RGB	<i>Red-Green-Blue</i>
HSV	<i>Hue-Saturation-Value</i>
SIFT	<i>Scale-invariant feature transform</i>
RANSAC	<i>Random sample consensus</i>
Slerp	<i>Spherical linear interpolation</i>
XOR	<i>Exclusive OR</i>
VIMG	<i>Virtual image</i>
Alg.	<i>Algorithm</i>
HHI	<i>Heinrich-Hertz-Institute</i>
MSR	<i>Microsoft Research</i>
PSNR	<i>Peak Signal-to-Noise Ratio</i>
MS-SSIM	<i>MultiScale structural similarity index</i>
VQM	<i>Video quality metric</i>
JPF	<i>Joint projection filling</i>
YUV	<i>Color space, Y (luminance or luma) , U, V (chroma)</i>

FDF	<i>Frame differential flicker</i>
MQ	<i>Multi-variate quadratic</i>
Cayley-Dixon-KSY	<i>Cayley-Dixon-Kapur-Saxena-Yang</i>
Res	<i>Resultant</i>
GCD	<i>Greatest common divisor</i>
RSC	<i>Rank submatrix computation</i>
monom	<i>Monomial</i>
MVC	<i>Mean value coordinates</i>
MRF	<i>Markov random field</i>
BRISQUE	<i>Blind/Reference Image Spatial QQuality Evaluator</i>
NSS	<i>Natural scene statistic</i>
SVR	<i>Support vector regressor</i>
DC	<i>Discrete cosine</i>
MSCN	<i>Mean subtracted contrast normalized</i>
GGD	<i>Generalized gaussian distribution</i>
AGGD	<i>Asymmetric generalized gaussian distribution</i>
DCT	<i>Discrete cosine transform</i>
VQA	<i>Video quality evaluation</i>
ToF	<i>Time-of-flight</i>
SURF	<i>Speeded up robust features</i>
KPCA	<i>Kernel principal component analysis</i>
SVD	<i>Singular value decomposition</i>
GSVD	<i>Generalised singular value decomposition</i>
OPG	<i>Oriented projective geometry</i>
PBR	<i>Probability-based rendering</i>
SSPM	<i>Steady-state matching likelihood</i>
RWR	<i>Random walk with restart</i>
VSRs	<i>View synthesis reference software</i>
IQA	<i>Image quality assessment</i>
DMOS	<i>Differential mean opinion score</i>
ITU	<i>International Telecommunication Union</i>
SROCC	<i>Spearman's rank ordered correlation coefficient</i>
LCC	<i>Linear (Pearson's) rank ordered correlation coefficient</i>
SfM	<i>Structure from motion</i>
PSG	<i>Parameterized scene graph</i>
PTZ	<i>Pan-tilt-zoom</i>
GD	<i>Gradient descent</i>
MRF	<i>Markov random field</i>

LIST OF MATH SYMBOLS

$P(\cdot)$	<i>Probability function</i>
$P(\cdot \mid \cdot)$	<i>Conditional probability function</i>
f_m	<i>Weak learner based on the naive density estimates</i>
L_m	<i>Likelihood function</i>
Q_∞^*	<i>Absolute dual quadric</i>
Ω_∞	<i>Absolute conic</i>
π_∞	<i>Plane at infinity</i>
P	<i>Camera projection matrix</i>
ω	<i>Image of absolute conic Ω_∞</i>
ω^*	<i>Dual image of absolute conic Ω_∞</i>
K	<i>Internal camera matrix</i>
R	<i>Camera rotation matrix</i>
$\mathfrak{R} = KR, \mathfrak{R}\mathfrak{R}^{-1}$	<i>Infinite homography</i>
R^n	<i>Real coordinate space of n dimensions</i>
θ	<i>Angle</i>
$H_{\Lambda_{kl}, \cdot}$	<i>Homography induced by the plane Λ_{kl}</i>
argmin	<i>Argument of the minimum</i>
$\alpha/\beta/\gamma$	<i>Parameter value</i>
I	<i>Camera view/Image</i>
Z_d	<i>Depth</i>
$Q/X/Y/V$	<i>3D scene point</i>
q_i	<i>Image point</i>
m_{ji}	<i>Orientation basis vectors</i>
ς_i	<i>FP-PIV structure parameters/coefficients</i>
ζ	<i>FP-PIV structure parameters/coefficients vector</i>
λ_i	<i>Projective depth associated with scene point</i>
$F_p = Z/pZ$	<i>Quotient of the ring of integers Z mod the ideal pZ, p is a prime</i>
$F_p[\cdot] / f(\cdot)$	<i>Finite quotient fields</i>
$\operatorname{rank}(A)$	<i>Rank of a matrix A</i>
$AS(\cdot, \cdot)$	<i>Appearance score</i>
$\max(\cdot, \cdot)$	<i>Maximum of two quantities</i>

$\{S_k\}_1^N$	<i>An ordered sequence</i>
Ξ	<i>Space-time volume</i>
$\frac{\partial F_{\cdot,p}}{\partial p} \big _p$	<i>Partial derivative with respect to p</i>
$\nu = (\cdot, \cdot, \cdot)^T$	<i>An arbitrary vector</i>
$GF(p^n)$	<i>Galois field of size p^n, p is a prime</i>
$\hat{I}$	<i>MSCN coefficients</i>
Δ	<i>3D plane</i>
$H_{L(\cdot, \zeta)}$	<i>Hessian of the Lagrangian with respect to ζ</i>
$MS(I_r, I_s)$	<i>Matched SIFT features between view I_r and I_s</i>
$S_{s_p}^o$	<i>Objectness score's for superpixels s_p</i>
E	<i>Energy term</i>