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

MANSI SHARMA

DEPARTMENT OF ELECTRICAL ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY DELHI
MAY 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 विएवपॉइन्ट 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 रिग्स का उपयोग करके मनमाने ढंग से कैमरे कैप्चर करना कठिन है (और महंगा है)। यह प्राकृतिक पर्यावरण के एक्वीजीशन में इन्फ़्लेक्सिबिलिटी 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 विएवपॉइन्ट 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 चूका 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 विएवपॉइन्ट 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 मे 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

# LIST OF FIGURES

# LIST OF TABLES

# LIST OF ACRONYMS

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

**FDF** *Frame differential flicker*

**MQ** *Multi-variate quadratic*

**Cayley-Dixon-KSY** *Cayley-Dixon-Kapur-Saxena-Yang*

**Res** *Resultant*

**GCD** *Greatest common divisor*

**RSC** *Rank submatrix computation*

**monom** *Monomial*

**MVC** *Mean value coordinates*

**MRF** *Markov random field*

**BRISQUE** *Blind/Reference Image Spatial QUality Evaluator*

**NSS** *Natural scene statistic*

**SVR** *Support vector regressor*

**DC** *Discrete cosine*

**MSCN** *Mean subtracted contrast normalized*

**GGD** *Generalized gaussian distribution*

**AGGD** *Asymmetric generalized gaussian distribution*

**DCT** *Discrete cosine transform*

**VQA** *Video quality evaluation*

**ToF** *Time-of-flight*

**SURF** *Speeded up robust features*

**KPCA** *Kernel principal component analysis*

**SVD** *Singular value decomposition*

**GSVD** *Generalised singular value decomposition*

**OPG** *Oriented projective geometry*

**PBR** *Probability-based rendering*

**SSPM** *Steady-state matching likelihood*

**RWR** *Random walk with restart*

**VSRS** *View synthesis reference software*

**IQA** *Image quality assessment*

**DMOS** *Differential mean opinion score*

**ITU** *International Telecommunication Union*

**SROCC** *Spearman's rank ordered correlation coefficient*

**LCC** *Linear (Pearson's) rank ordered correlation coefficient*

**SfM** *Structure from motion*

**PSG** *Parameterized scene graph*

**PTZ** *Pan-tilt-zoom*

**GD** *Gradient descent*

**MRF** *Markov random field*

# LIST OF MATH SYMBOLS

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

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