Multiview 3D Content: Acquisition, Reconstruction, Immersive Visualization and Quality Assessment

Roopak Rajendra Tamboli

A Thesis Submitted to

Indian Institute of Technology Hyderabad

In Partial Fulfillment of the Requirements for

The Degree of Doctor of Philosophy

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Department of Electrical Engineering

January 2021

Abstract

Lifelike visual immersion has emerged as an important milestone for remote assistance and collaboration technology. Since previous decades, various technological solutions for remote collaboration have been suggested for a broad class of applications in healthcare, education, entertainment, research, defense and related areas. Majority of extant solutions are based on two-dimensional (2D) visualization, and remain in wide use on the strength of their affordability, ease of use and well-established infrastructure. Despite certain variations, the remote assistance and collaboration technology has essentially remained teleconferencing-like. Such systems, while achieving a degree of telepresence, generally leave much room for improvement towards achieving true immersion. To bridge this gap, three-dimensional (3D) visualization techniques have been explored since several decades. With the emergence of telepresence solutions based on real-time light field acquisition and multi-user glasses-free immersive visualization, the holy grail of true-to-scale lifelike immersion now appears within the technological reach.

A variety of applications have been adopting 3D visualization as an integral element. Consequently, in recent years, several research and development opportunities have opened up in the areas of 3D content acquisition, 3D reconstruction, processing, and visualization. Accordingly, a rapid progress has been observed in theory, algorithms, software, products and services pertaining to 3D applications. Taking into account such growing importance of 3D immersive visualization in the coming days, in this dissertation we explore various aspects of novel 3D media technologies, such as algorithmic reconstruction of 3D objects from multiple 2D views, rendering of such 3D content on 3D displays to create the experience of immersive visualization, and assessment of the resulting quality of experience (QoE). In this regard, we present a mathematical treatment of image acquisition and 3D object reconstruction from multiple views. Subsequently, we outline the principle behind 3D visualization on lightfield displays (LFDs). Furthermore, we envisage an end-to-end system that allows image acquisition and 3D reconstruction at remote location, transmission to user location, and eventual 3D visualization.

To elucidate the importance of the envisaged system, we illustrate its utility for a healthcare application involving teleophthalmology. Specifically, we propose a hybrid teleophthalmology framework in which the existing primary healthcare centres of eye-care networks, located in remote areas, would be equipped with a multi-camera-based imaging system in addition to the existing conventional ophthalmology equipment. In an emergency situation requiring facial surgery, the primary healthcare center would

capture facial multiview data and send it to the operation center to facilitate pre-operative planning while the subject is in transit. This approach would potentially reduce the time between admission of the subject and commencement of surgery. Furthermore, it can also enable remote post-operative assessment and as a consequence, reduce hospitalization time and the number of visits to the surgical facility. Towards realizing this goal, we present true-to-scale 3D reconstruction of a human face, and a preliminary surface fitting on recovered sparse point cloud. We also discuss its implications in the aforementioned hybrid teleophthalmology framework.

While various elements of the envisaged system have progressed significantly, relatively less efforts has been directed at understanding and quantifying the user experience. To fill this gap, our primary focus is on developing novel methods for assessing perceptual quality of the super-multiview (SMV) content. To this end, we create SMV content primarily for visualization on projection-based LFDs. These displays provide lifelike 3D visualization via virtually-continuous horizontal parallax without the need of specialized glasses. The spatial, angular and spatio-angular properties of SMV content are known to affect QoE on such displays. Therefore, we assess perceptual quality under various settings of spatial and angular resolution of a variety of static SMV contents; subjectively via conducting studies that involve human participation, and objectively using novel full-reference (FR) objective quality assessment methods. These objective algorithms strive to predict subjective opinions, thereby potentially mitigating the need of subjective studies in future. However, that depends on extensive testing and accordingly modifications in the proposed metrics for various test conditions. In addition to spatio-angular properties of the content, certain features of the content such as relative size of 3D objects, in real-life and on the screen of LFD, as well as their orientation are known to affect the perceptual experience. To investigate this aspect, we analyze subjective opinions obtained in the aforementioned studies vis-à-vis content features. We then report an interactive study where participants were asked to orient 3D objects rendered on an LFD such that the final view is the most appealing to them.

In summary, we envision a generic end-to-end remote collaboration system, describe currently available elements of such a system, and outline related opportunities and obstacles. We briefly review a variety of applications that currently employ 3D visualization and remote collaboration, and discuss how emerging 3D media technologies can enhance QoE. Specific to some of these applications, we identify and address a subset of problems in acquisition, 3D reconstruction, immersive visualization and quality assessment of SMV content.

Contents

Declaration ii
Approval Sheet iii
Acknowledgements iv
Abstract vii

1 Introduction **1**
1.1 Motivation 1
1.2 Technological Evolution 4
1.2.1 Evolution of 3D Displays 5
1.2.2 3D Content Generation 6
1.3 Envisioned System 7
1.4 Associated Technologies and Opportunities 8
1.4.1 Acquisition/Generation of Multiview Content 9
1.4.2 3D Reconstruction 9
1.4.3 Immersive 3D Visualization 10
1.4.4 Quality Assessment 12
1.5 Contributions and Thesis Organization 13
1.5.1 True-to-scale 3D Reconstruction and Immersive Visualization 14
1.5.2 Illustrative Healthcare Application of Envisaged End-to-end System 15
1.5.3 Perceptual Quality Assessment under Spatial Distortions . . . 17
1.5.4 Perceptual Quality Assessment under Angular Distortions . . 18
1.5.5 Preferred Views of 3D Objects: A Preliminary Study using Lightfield Displays 20
1.5.6 Summary and Future Directions 22
1.5.7 Other contributions not elaborated in this dissertation 23
1.6 Highlights 23

2 True-to-scale 3D Reconstruction and Immersive Visualization 25
2.1 Background 25
2.2 True-to-scale Auto-calibration of Multicamera Network 26
2.2.1 Image Formation 27
2.2.2 Multi-objective Optimization 29
2.2.3 Euclidean Auto-calibration 30
2.3 Effect of Imperfections in Stereo Camera on Calibration Accuracy . . 31
2.3.1 Methodology 31
2.3.2 Towards Calibration of a Camera Network 33
2.3.3 Results 35
2.4 Immersive Visualization 37
2.4.1 Light Field 37
2.4.2 Representation of Light Field 38
2.4.3 Projection-based Wide-baseline Lightfield Displays 38

3 Illustrative Healthcare Application of Envisaged End-to-end System 40
3.1 Motivation 41
3.2 3D Reconstruction of Head Model from Multiple 2D Views 42
3.2.1 Experimental Setup 42
3.2.2 Image Acquisition 43
3.2.3 3D Reconstruction 44
3.2.4 Results 45
3.2.5 Discussion 47
3.3 Novel Hybrid Teleophthalmology: Technological Case for Oculofacial Surgery 48
3.3.1 Conventional Teleophthalmology 50
3.3.2 Proposed Hybrid Teleophthalmology 52
3.3.3 3D Reconstruction of Human Face 54
3.3.4 Results and Discussion 57
3.4 Summary and Outlook 60

4 Perceptual Quality Assessment under Spatial Distortions 62
4.1 Prior Work on Quality Assessment in Related Applications 63
4.2 Contributions 64
4.3 Turntable-based Multiview Content Acquisition 66
4.3.1 Issues with Existing Turntable Datasets 66
4.3.2 Turntable Setup 67

4.3.3 Generation of Distorted Content 69
4.4 Subjective Quality Assessment 69
4.4.1 Generation of Test Stimuli 70
4.4.2 Subjective Evaluation Methodology 71
4.5 Objective Quality Assessment 73
4.5.1 Spatial Quality Assessment 74
4.5.2 Angular Quality Assessment 80
4.5.3 Spatio-Angular Quality Assessment 82
4.5.4 Performance Evaluation Criteria 82
4.6 Performance Evaluation and Validation of the Proposed Method . . . 84
4.6.1 2D Image Quality Assessment 84
4.6.2 Spatial Quality Assessment 85
4.6.3 Angular Quality Assessment 85
4.6.4 Spatio-Angular Quality Assessment 85
4.6.5 Summary 87
4.7 Generalizability of the Proposed Method 87
4.7.1 Synthetic Object Dataset 89
4.7.2 Skipped-image Distortion 90
4.7.3 Results 94
4.8 Objective Quality Assessment of Key Frames of Super-multiview Video 95
4.8.1 Methodology 95
4.8.2 Visual Stimuli 96
4.8.3 Results and Discussion 97
4.9 Objective Quality Evaluation of a Novel Lightfield Representation . . 98
4.9.1 Related Work 99
4.9.2 Experimental Details 100
4.9.3 Results and Discussion 102
4.10 Summary and Discussion 107

5 Perceptual Quality Assessment under Angular Distortions 109
5.1 Overview of Results 110
5.2 Related Work 111
5.3 Materials and Methods 112
5.3.1 Dataset 113
5.3.2 View Synthesis 113
5.3.3 Generation of 3D views 115

5.4 Subjective Evaluation 116
5.4.1 Qualitative Analysis of Subjective Scores 117
5.5 Objective Quality Assessment 119
5.5.1 Spatial Modeling 119
5.5.2 Angular Modeling 121
5.5.3 Spatio-Angular Modeling 128
5.6 Results 129
5.6.1 Spatial Quality Assessment 129
5.6.2 Angular Quality Assessment 130
5.6.3 Spatio-angular Quality Assessment 130
5.6.4 Validation of the Proposed Method 131
5.7 Summary 133
5.7.1 Subjective Quality Assessment 133
5.7.2 Objective Quality Assessment 133
5.7.3 Next Step: Experimentation using Real-life Datasets 134
5.8 A High-angular-resolution Turntable Dataset 134
5.8.1 Capture Configuration 135
5.8.2 Dataset Validation 137
5.8.3 Turntable Consistency 137
5.8.4 Camera Calibration 138
5.8.5 Discussion 138

6 Preferred Views of 3D Objects: A Preliminary Study using Light-field Displays 139
6.1 Effect of Primitive Features of Content on QoE 140
6.1.1 Qualitative Analysis of Subjective Scores 140
6.1.2 Discussion 142
6.2 Preferred Views of 3D Objects 143
6.2.1 Existing Work 144
6.2.2 Methodology 145
6.2.3 Results of Subjective Evaluation 147
6.2.4 Considerations in Developing Objective Metric 148
6.2.5 Discussion 149

7 Summary and Future Directions **151**
7.1 Summary 151
7.1.1 True-to-scale 3D Reconstruction and Immersive Visualization 152
7.1.2 Illustrative Healthcare Application of Envisaged End-to-end System 153
7.1.3 Perceptual Quality Assessment under Spatial Distortions . . . 153
7.1.4 Perceptual Quality Assessment under Angular Distortions . . 154
7.1.5 Preferred Views of 3D Objects: A Preliminary Study using Lightfield Displays 155
7.2 Future Directions 155
7.2.1 Acquisition and Reconstruction 157
7.2.2 Immersive Visualization and Quality Assessment 157

Acronyms **158**

References **162**

Chapter 1

Introduction

1.1 Motivation

Lifelike visual immersion has emerged as an important milestone for remote assistance and collaboration technology [1]. Since previous decades, various technological solutions for remote collaboration have been suggested for a broad class of applications in healthcare, education, entertainment, research, defense and related areas [2–8]. Majority of extant solutions are based on two-dimensional (2D) visualization, and remain in wide use on the strength of their affordability, ease of use and well-established infrastructure. Specifically, in multi-participant teleconference, where each participant uses his/her personal device, often either a smartphone or a computer, each participant's 2D screen is usually divided into small tiles, each containing the video stream corresponding to one attendee, and possibly one bigger tile, depicting the content under discussion. Within this framework, significant effort has been directed at improving immersive experience by enhancing resolution, frame rate and other parameters. Indeed, in modern educational, medical, government and business establishments, immersive multi-user experience is typically created using either a large 2D display or an array of 2D displays with advanced features such as very high frame rates, ultra-high-definition (UHD) resolution, wide color gamut (WCG), high dynamic range (HDR) and sometimes even curved screens. In short, despite certain variations, the remote assistance and collaboration technology has essentially remained teleconferencing-like [2, 9]. Such systems, while achieving a degree of telepresence, generally leave much room for improvement towards achieving true immersion.

To bridge this gap, three-dimensional (3D) visualization techniques have been explored since several decades. Particularly, generation of 3D models of room-sized environments has been attempted. Early approaches involved deploying multiple

cameras around the scene and using stereo reconstruction techniques to construct and continually update the 3D model of subjects, objects and the scene. An attempt in this direction produced a virtual space teleconferencing system, requiring eleven cameras, and head-mounted display (HMD) for 3D visualization [10]. An improved system, using a set of 51 cameras fixed on a dome, could subsequently capture dynamic scenes and render those from new viewpoints [11]. In the course, a bidirectional 3D telepresence system, called 'blue-c', was proposed that scanned as well as displayed a participant at each of two locations. There, the displays provide a CAVE (computer-assisted virtual environment)-like experience with head-tracked stereo visualization for single local user [12]. Towards achieving multi-user co-located collaboration in shared environments, a stereoscopic six-user display was later proposed [13]. Relatively recently, Holoportation reconstructs high-quality 3D models of people, and compresses and transmits those in real time [14]. However, users are required to wear mixed-reality-based HMD. While such stereoscopic display and HMD-based solutions enable co-presence in shared virtual or mixed environments, the user can only partially see his/her real self and his real environment. In response, a unidirectional telepresence solution based on real-time light field acquisition and multi-user glasses-free immersive visualization has been proposed in the past two years [8, 15]. In such a setting, the holy grail of true-to-scale lifelike immersion now appears within the technological reach.

Unsurprisingly, there has been a growing anticipation among experts that 3D visualization would become an integral part of various real-life applications over time [16, 17]. In this context, the entertainment industry has particularly been a major user of 3D media since several decades, and accordingly, has significant influence on related research and development. In this connection, note that the so-called 3D movies presently shown in theater and at home are usually stereoscopic, and require specialized glasses (and other headgears) for viewing. This arrangement, while acceptable in the darkness of the theater, tends to interfere with users' social interaction and other functions, which are desirable at home. Clearly, the prevalent 2D entertainment, requiring no such special equipment, does not so interfere. Therefore, for a majority of home-based consumers, the upgrade in quality of experience (QoE) from 2D to 3D entertainment must accompany a glasses-free viewing experience [18, 19]. Against this backdrop, we consider true 3D content rendered on lightfield displays (LFDs), to enjoy which one requires no specialized equipment, and which potentially fulfills the aforementioned technological demand on future home entertainment technology.

In addition, an upgrade from 2D to 3D visualization is expected to also improve healthcare delivery. Specifically, 3D representation and stereoscopic visualization of organs have already proven helpful in training, remote assistance, diagnosis, as well as pre-, peri- and post-operative surgical planning [20]. Glasses-free visualization of true 3D models is expected to advance the present state of the art. In addition, modern medical imaging equipment, including computed tomography (CT) and magnetic resonance imaging (MRI) scanners, sometimes offers volumetric representations of internal organs, so that clinicians do no longer need to go through 2D slices and mentally synthesize the underlying 3D structure. Unfortunately, such 3D information is traditionally visualized on 2D screens, and, in recent years, on stereoscopic displays. The experience of visualizing such information on a true 3D display is significantly richer and natural, leading potentially to improved diagnostic outcome. Furthermore, in unprecedented situations such as the raging Covid-19 pandemic [21], remote assistance and collaboration have come to be recognized as ever more important in ensuring healthcare delivery in challenging circumstances [22–24]. Of course, existing telemedicine technology already serves a number of related needs such as regular medical consultations. In certain circumstances, advanced 3D modeling and display should improve medical outcomes for patients needing to travel long distances for availing surgical treatment [25]. In fact, one anticipates that healthcare specialities essentially dependent on human presence – such as psychiatry [26] – could one day be practised remotely, when the telepresence technology would advance to such an extent that an immersive lifelike experience is created at both ends.

In the same vein, 3D visualization could lead to improvement in various other walks of life including education. Indeed, 3D demonstration, even if virtual, within classrooms potentially improves grasping of complex 3D concepts, such as molecular structures, when compared to the usual 2D illustrations [27]. Further, by upgrading from 2D to 3D visualization, a significant advancement is anticipated in the remote classroom experience, which is emerging as a useful modality for disseminating knowledge in circumstances, where trained educators are not available locally. At the same time, use of head-up displays (HUDs) has been proposed in vehicles to enhance driving experience and reduce decision fatigue by augmenting the windshield with the information regarding route, road signs, various warnings, pedestrians and other vehicles. In this space, turning the entire windshield into an LFD has been proposed recently [28]. Numerous uses of 3D visualization are also emerging for strategic and security applications [29]. For instance, municipal bodies often deploy a city-wide multicamera network for surveillance. Data acquired from such networks can be used

to reconstruct lifelike 3D representation of face and gait of persons of interest. Similarly, fusion of multimodal images could help create 3D walk-through of terrain and structures of strategic importance.

Taking into account the growing importance of 3D immersive visualization in the coming days [30], in this dissertation we explore various aspects of such 3D technology, such as algorithmic reconstruction of 3D objects from multiple 2D views acquired using standard cameras, rendering of such 3D content on 3D displays to create the experience of immersive visualization, and assessment of the resulting QoE. In this regard, we present a mathematical treatment of image acquisition and 3D object reconstruction from multiple views. Subsequently, we outline the principle behind 3D visualization on LFDs. Further, we envisage an end-to-end system that allows image acquisition and 3D reconstruction at remote location, transmission to user location, and eventual 3D visualization. To elucidate the importance of the envisaged system, we illustrate its utility for a healthcare application involving teleophthalmology. While various aforementioned aspects of such a system have received significant attention, relatively little effort has been directed at understanding and quantifying the user experience. To fill this gap, here our primary focus is on developing novel methods for assessing perceptual quality of the super-multiview (SMV) content, both subjectively and objectively. First, we briefly review the technological evolution of 3D displays and 3D content generation methods.

1.2 Technological Evolution

Several 3D visualization technologies, such as stereoscopic television and cinema, autostereoscopic displays, multiview displays, volumetric displays, virtual reality (VR) headsets and LFDs, strive to achieve lifelike realism [31, 32]. Each of these technologies have certain potentials as well as inherent limitations and challenges [33]. Among those, the recently introduced LFDs, which we describe in greater details in Chapter 2, appear promising as the future of 3D visualization. With the rapid progress in various display technologies, extensive reviews have regularly appeared in the related literature over the last two decades [31, 34–47]. At this point, we briefly review the technological evolution involving 3D display and 3D content generation.

1.2.1 Evolution of 3D Displays

The first attempt at developing stereoscopic 3D displays dates back to 1838 when Wheatstone constructed a stereoscope using a pair of mirrors and two drawings depicting the same object from the left and right perspective [48]. The Brewster Stereoscope from 1849 was a more compact unit that allowed the creation of compact hand-held devices, and was demonstrated to a wide audience during the Great Exhibition in 1851. Between 1860 and 1930, stereoscopic photographs were in certain degree of use. In parallel, Friese-Greene demonstrated a stereoscope which fused the images played back from two films in the late 1890s. A more practical approach for exhibiting 3D movies involved anaglyph, which was first demonstrated in 1915 by Porter. The first publicly shown anaglyph movie was presented in 1922 by Fairall. The anaglyph method has since been used, albeit mostly in low-cost use cases. Also, in 1922, Hammond demonstrated the Teleview system, which used alternating left and right frames projected from a pair of projectors, and an individual shutter device for each viewer which was mounted on an adjustable gooseneck on each seat [49]. In this connection, parallax barrier was perhaps the oldest feasible technique to implement autosteroscopic imagery. The left and right (or multiple) images are interleaved in a striped image, positioned behind a very dense set of equidistant vertical barriers. The first known implementation dates back to 1692 by G. A. Bois-Clair. The first known photograph-based parallax barrier method appeared in 1903 [50]. Later, attempts to achieve 3D imaging using integral lenses lead to the development of lenticular lens-based displays. The first well-known implementation by Lippmann was reported in 1908 [51], where a fly-eye lens array was employed to capture and reproduce imagery including both horizontal and vertical parallax. The integral lens method was later simplified to a lenticular lens array in the 1920's that can be used to reproduce images with horizontal parallax only. Lenticular lenses can also be used in combination with flat 2D screens to generate animated 3D imagery, which was first reported in 1996 by van Berkel [52].

On the other hand, the origins of holography can be traced back to the work of Gabor from 1947 [53]. Leith and Upatnieks presented the first laser-based transmission holograms in 1964 [54]. The mass production of holograms was made possible by Dr. Stephen A. Benton, who invented holography with white light transmission (rainbow hologram), which enabled holograms to be seen in white light [55]. Recently, VR HMDs have been used to enable remote collaboration in immersive virtual environments. In 1968 Ivan Sutherland and Bob Sproull created the first VR/AR head mounted display that was connected to a computer [56]. It was too heavy for any

user to comfortably wear and was suspended from the ceiling.

According to the 'Gartner Hype Cycle for Emerging Technologies, 2019' report, immersive workspaces are in their so-called 'innovation trigger' phase and the 'plateau of productivity' is expected to reach within 5–10 years [30]. These would most likely be backed by advancements in the VR technology which is currently ahead of LFDs in terms of acceptance in the consumer market, due to comparatively lower cost, smaller form factor, and higher ease of use and content generation. However, VR technology shares many drawbacks with its predecessor, 3D (stereoscopic) television, which is practically non-existent in the consumer market [57, 58]. On the other hand, the current generation of LFDs offer lifelike immersive experience simultaneously to multiple viewers over a large field of view (FOV) via virtually-continuous (presently horizontal-only) parallax. We believe that LFDs would become a ubiquitous technology, if certain theoretical and technological challenges are overcome. On the other hand, lightfield (LF) acquisition/generation technology has progressed significantly over the past few years and LF cameras are already available in the consumer market. In the next section, we briefly review some of the techniques to acquire/generate 3D content.

1.2.2 3D Content Generation

While stereoscopic content can directly be captured using stereoscopic cameras, generation of true 3D content is generally more involved. The latter content is either generated using a computer, or acquired using range scanners, or reconstructed from multiple 2D views. In this dissertation, we primarily focus on SMV content for LF visualization. Accordingly, we next review the evolution of multiview content acquisition methods, and include LF cameras for the sake of completeness.

The work on first single-lens plenoptic camera was published in 1992 by Edward Adelson and John Wang. In 1996, Marc Levoy and Pat Hanrahan presented a simplified 4-dimensional mathematical representation of LF data [59]. Subsequently, they proposed methods for recording light fields, including computer-assisted camera gantries where a single moving camera took pictures at predetermined positions, and/or objects were rotated for recording. Camera arrays for capturing light field became popular in 2000s with the 64- and 100-camera setups at MIT and Stanford University, respectively [60]. Hand-held LF camera was conceptualized in 2005, the first commercial plenoptic camera (Raytrix R11) appeared in 2010, and the first consumer LF camera by Lytro appeared in 2012 [61]. Since then, several multicamera

setups have come to realization to acquire both narrow-baseline and wide-baseline light fields [62, 63]. Turntable datasets, although not aimed at LF visualization, have also been reported in recent years [64–67]

Against this backdrop, we envisage a futuristic remote collaboration system, which, we believe, will be achievable in the next decade(s). Specific technical challenges and contributions of this dissertation are discussed in Sections 1.4 and 1.5, respectively.

1.3 Envisioned System

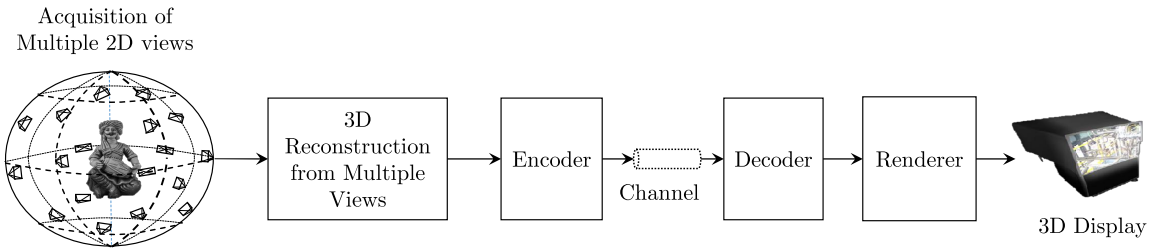

Figure 1.1: A schematic of the envisioned system.

The envisioned system would capture information from a real-life scene — such as a human participant in a (telepresence) meeting and his/her environment, an organ of a patient being collaboratively examined by geographically separated doctors, a hard-to-reach or hostile territory and so on — using a mechanism that offers instantaneous data acquisition and real-time communication, does not intrude into the scene and thereby does not require any behavioural changes in the subjects (if any), and calibrates itself automatically. One would then generate a parameterized 3D representation from the captured information, where parameters encode the information of shape, scale, geometry, colour, texture, reflectance and other surface properties of subjects, objects and scene. A parameterized representation would potentially eliminate the need of sending complete 3D data. Since the parameters do not change drastically with time, those may be quantized based on certain perceptual and other criteria dictated by the application at hand, and sent to the 3D display. The receiver would then impart these parameters onto generic deformable 3D models and recreate the reality. Subsequently, the content would be visualized on a device that provides lifelike immersive audio-visual experience to multiple viewers, simultaneously without any headgear, along with the possibility of remote interaction, haptic feedback and annotation. Alternatively, in certain use cases, instead of explicit 3D geometry reconstruction, one may encode the acquired SMV content by exploiting

intraview, interview and temporal redundancies. At the received end, image-based rendering techniques can be used to generate immersive visualization. Finally, the system should ideally be able to operate in full-duplex manner.

In order to realize the remote collaboration system envisioned above, several challenges in content acquisition, 3D reconstruction, representation, rendering, immersive visualization, user interface, user experience and interactivity need to be addressed. Many components of such a system are already available. With the emergence of multiplayer low-latency server-based gaming, high-frame-rate 8K television and 5G networks, the supporting communication infrastructure with high quality of service (QoS) would likely be available in a decade. In fact, a complete remote collaboration system has already been achieved – examples include Holopresence [15] and Holoportation [14] – in terms of functionality. Of course, one needs to presently overlook the various requirements of economic and market viability, including low cost (initial, operational and maintenance), small form factor, low power consumption, scalability and upgradability, durability, ease of deployment and use, and excellent QoE [5]. Encouragingly, the aforementioned continuously evolving systems already offer near real-time performance in certain circumstances. Against this backdrop, we review state of some of the associated technologies and seek to identify opportunities for improvement.

1.4 Associated Technologies and Opportunities

In this section, we discuss technical preliminaries, describe theoretical and technological challenges, and state the problems addressed in this dissertation. The organization of this section roughly follows the title of this dissertation, namely acquisition of multiview content (Section 1.4.1), 3D reconstruction from multiple views (Section 1.4.2), immersive visualization (Section 1.4.3) and quality assessment (Section 1.4.4). The rationale behind such organization stems from the schematic of the end-to-end 3D media system described in Section 1.3. As shown in Figure 1.1, such a system comprises of data acquisition/generation, recovering 3D geometry (if the use case demands) or alternatively encoding of acquired multiview data for efficient storage and transmission, decoding, rendering and visualization.

1.4.1 Acquisition/Generation of Multiview Content

In this section, we limit the scope to multiview imaging. Multiview content can be generated synthetically or acquired by physical cameras. Computer-generated (CG) multiview content can have arbitrary spatial and angular resolution, only limited by the storage and computing capacities of the rendering machine. Complex camera arrangements and movements can easily be scripted in a typical 3D modelling and rendering software. However, real-life content needs to be captured by physical camera arrays. Such arrays can have grid, linear, arc and circular (looking outward or looking inward) configurations. For a large static scene, one can deploy a camera gantry system to achieve dense sampling. For small static scenes, even a turntable suffices in several applications. Recently introduced LF cameras can achieve dense sampling of a small-sized scene with very high angular resolution. However, such content usually has low spatial resolution and does not cover a wide baseline. Therefore, such content is not suitable for wide-baseline projection-based LFDs. For such displays, real-life content needs to be captured using multicamera network.

Multicamera systems pose several challenges. First, the number of cameras (and consequently the cost) required for SMV content generation is very high. Next, cameras cannot be arranged too closely with each other due to their physical size, resulting in low angular resolution. Calibration (geometry and color), registration and synchronization of a large number of cameras is often a tedious task. Furthermore, acquiring SMV video generates large data volumes at very high rates. While storage is inexpensive, a high bandwidth interface and storage with fast read/write capability increases the cost of camera network. Not surprisingly, only a handful of datasets that can be used on wide-baseline LFDs are available [62, 63, 68, 69], and most of those are of indoor scenes.

1.4.2 3D Reconstruction

Multi-camera networks are becoming ubiquitous in a variety of applications related to healthcare [70], education [9], entertainment [71, 72], autonomous vehicles, civil security and defense [73, 74]. The foremost task in deploying a multi-camera network is camera calibration, which usually involves introducing an object with known geometry, such as chessboard patterns, checkered cubes and laser pointers into the scene. Traditionally, extraneous objects with known geometry have been used for calibrating single camera [75–77] as well as network of cameras [9, 78, 79]. In wide-area camera networks, additional information, such as global positioning system (GPS) data, is

sometimes used to obtain extrinsic parameters [80]. However, most of the aforementioned applications necessitate non-intrusive automatic camera calibration and reconstruction of the 3D likeness of objects. In these scenarios, such as telepresence and remote classroom, where an adaptive camera network dynamically maintains the image quality, pre-calibrated cameras could be restrictive, and extraneous objects may not be introduced [81]. Further, applications such as oculofacial measurements and multi-party telepresence, requiring fusion of multiple scenes, additionally demand preservation of scale [82]. To this end, a class of camera auto-calibration methods imposes constraints on the camera network rather than on the scene [83]. As discussed in the previous section, calibrating a large network of cameras is a difficult task. Assisted calibration of individual camera followed by registration remains valid only until the scene does not change significantly. Auto-calibration of a large camera network appears attractive. However, several issues need to be addressed in this regard, such as modelling of camera parameters with respect to time, inter-camera variations, imperfections in cameras and lens, choice of 3D co-ordinate system(s) for localizing cameras, error propagation and so on.

1.4.3 Immersive 3D Visualization

The human visual system (HVS) uses several cues to understand the spatial relations of real world objects, resulting in the perception of depth. The most important visual depth cues used by the HVS are binocular disparity, motion parallax, vergence and accommodation, and pictorial cues [39, 44].

Binocular disparity measures the difference of the position of object points when projected onto the left and right retinas. The amount of horizontal difference is inversely proportional to the distance of the object, and thereby depth. Further, motion parallax, a monocular depth cue, refers to the effect that objects closer to a moving viewer appear to move faster than objects further away. Vergence and accommodation are two corresponding oculomotor functions related to depth perception. Vergence is the simultaneous movement of eyes in opposite directions, so that the object of interest is projected to the center of the retina in both eyes. When looking at close objects, the eyes converge and when looking at an object further away, the eyes diverge. Accommodation involves the focusing of the eye's lens to the object of interest, so that is appears sharp on the retina. Retinal blur is the primary effect that controls the eye's accommodation. Vergence and accommodation typically work together to provide sharp and high-resolution images, as well as depth information

to the HVS.

On the other hand, pictorial cues, such as occlusion, perspective, relative size, depth from defocus, pattern scaling, shadows, and atmospheric effects are monocular depth cues that are learned from experiences. Most of the pictorial cues work for any distance, and are relied upon by the HVS when the usual depth cues loose their significance due to large distance. Several display techniques exist, with varying levels of implementation complexity and information content, aimed at reproducing the binocular visual cue. Such displays address at least the stereopsis, i.e. the ability of the human visual system to fuse two different perspective images, causing retinal disparity. The most well-known types of 3D displays include stereoscopic displays and multiview autostereoscopic displays. However, these displays fail to reproduce other visual cues such as focus and continuous parallax. Furthermore, such displays suffer from vergence-accommodation conflict due to the display screen not being at the depth level where the visualized object appears to be.

In recent years, advanced 3D displays, such as volumetric displays, LFDs and holographic displays have been proposed. Among those, LFDs are the most promising ones since they can emit a direction-dependent dense set of rays resulting in high image quality over a large FOV and depth of field (DoF) without 3D glasses. They also support virtually-continuous motion parallax. The current generation of projection-based wide-baseline LFDs support only horizontal parallax. Since binocular disparity and motion parallax occur mostly in horizontal direction, these displays can cater to most of the existing and emerging use cases. The scope of this dissertation is limited to light fields captured using camera arrays and turntable setups, and subsequently visualization of projection-based LFDs. We briefly discuss the working principle of such displays in the following section.

Projection-based wide-baseline lightfield displays

These displays attempt to recreate the light field from the underlying multiview data input and thus to provide a virtually-continuous horizontal parallax, over a large FOV. The display screen uses a sheet of holographic optical elements, illuminated by a series of projectors from different angles. The input to the projectors is determined by the geometry of the display system. With proper software control, the light rays leaving the various pixels of the screen are made to propagate in multiple directions, recreating the experience of observing a natural scene through a glass window. Thus, the LFD does not show discrete individual views. Instead, each pixel has its own set of angular sectors, determined by its location relative to the projectors. Therefore, the

3D content seems more continuous, without the inter-view discontinuity or crosstalk encountered in the other types of 3D displays [31, 84]. LFDs reproduce all depth cues to some extent. Binocular disparity is reproduced for objects in the depth range. Motion parallax is reproduced in the horizontal direction, within the FOV of the display, with a given angular resolution. The vergence-accommodation conflict observed in stereoscopic displays is much reduced with LFDs.

While the display technology itself is not in the scope of this dissertation, we identify some of the challenges that current-generation LFDs need to overcome. Firstly, present projection-based LFDs are bulky and require huge amount of power for operation. The initial, operational and maintenance costs are prohibitively high. As opposed to flat-panel displays, periodic calibration of the equipment is required. Since the LFD technology is under development, various implementations of such displays are available. Consequently, the conversion/mapping of captured content to the light field needs to be performed separately for each variant of such displays. Furthermore, in order to provide virtually-continuous parallax, glasses-free displays require a large number of input views [31]. Unfortunately, acquisition of such SMV content with high angular resolution requires special hardware/software setup, and significant economic investment.

A note on the usage of the term 'light field'

In the related literature, many variations of the term 'light field' are found, namely, 'light-field', 'light field', 'lightfield' and even 'field-of-light' by certain groups [85]. In this dissertation, we use the term 'light field' to denote the plenoptic function and it is not abbreviated. The term 'lightfield' is used as an adjective, e.g. lightfield content, lightfield display etc., and is abbreviated as 'LF'. Furthermore, we use the abbreviation 'LFD' for 'lightfield display'.

1.4.4 Quality Assessment

In most of the applications of 3D visualization discussed so far, humans are the ultimate consumers of the content. Therefore QoE of the end user is of utmost importance. According to the Qualinet White Paper on Definitions of Quality of Experience, "Quality of experiencing is the degree of delight or annoyance of a person during the process of experiencing. It results from the person's evaluation of the fulfillment of his or her expectations and needs with respect to the utility (pragmatic and hedonic) in the light of the person's context, personality and current state" [86–

88]. Various processing stages from content generation/acquisition to visualization affect the QoE. Therefore, it is necessary to be able to assess and quantify the QoE, which in turn would provide an objective cost that should ideally be minimized by the aforementioned steps. An immediate way of determining the quality is to solicit opinion from human observers. To this end, subjective methods of quality assessment require human volunteers to evaluate content according to a well-defined protocol, using some pre-determined criteria, and sometimes also answer a questionnaire.

On the other hand, the goal of objective quality assessment is to develop quantitative measures that can automatically predict perceived quality. Such metrics can be used to dynamically monitor and adjust content quality. Also, they can be used to optimize algorithms and parameter settings of content acquisition, processing and visualization. Depending on the amount of information — complete, partial or zero — available for the reference/distortion-free content accessible by the objective metrics, they are classified as full-reference (FR), reduced-reference (RR) and no-reference (NR) metrics, respectively. The performance of the objective metrics is evaluated with respect to the prediction accuracy, the prediction monotonicity, and the prediction consistency in relation to predicting the subjective assessment of content quality over the range of the considered test stimuli.

Subjective and objective quality assessment of 2D and stereoscopic images and videos has been studied for several decades. Consequently, benchmark datasets, well-defined guidelines of conducting subjective studies as well as objective metrics are available. Comparatively, quality assessment of glasses-free LF visualization is an emerging area. Accordingly, there are no standard guidelines for subjective quality assessment and established metrics for objective quality assessment, to the best of the author's knowledge. As alluded to earlier, LFDs need to overcome several challenges in order to be adopted widely. Furthermore, content acquisition for such wide-baseline displays is also not easy. Consequently, the existing studies on LF quality assessment in the literature make use of narrow-basline datasets, captured using either plenoptic cameras or small camera arrays. The subjective quality assessment is conducted using 2D or stereoscopic displays. Recently, simulators for LF visualization have also been proposed [89–91]

1.5 Contributions and Thesis Organization

The contents of this dissertation are organized into seven chapters. Figure 1.2 depicts the key theme of each chapter. Chapter 2 presents theoretical underpinnings of

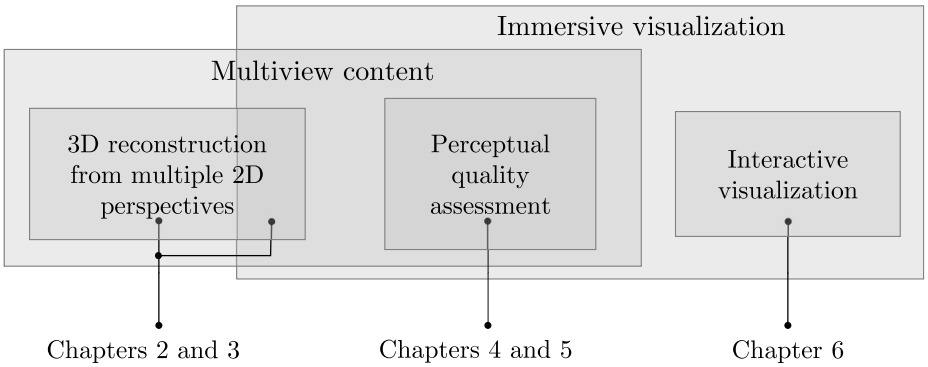

Figure 1.2: Scope and organization of the thesis.

true-to-scale 3D reconstruction from multiple views, as well as 3D visualization. In Chapter 3, we use this method to reconstruct 3D face and subsequently propose a hybrid teleophthalmology paradigm. Chapters 4 and 5 discuss novel subjective and objective methodologies to assess perceptual quality of SMV content visualized on projection-based LFDs. In Chapter 6, we seek preferred orientations of various 3D objects via interactive visualization. Finally, we provide a summary of the dissertation and discuss future directions in Chapter 7. In the following, we provide a brief overview of each chapter.

1.5.1 Chapter 2: True-to-scale 3D Reconstruction and Immersive Visualization

In many use cases of 3D visualization, such as telepresence, scene contents do not change drastically and the scene span is limited. Therefore, a very dense sampling of light field can be achieved. Existing LFDs can convert SMV video streams to display-specific light fields and provide lifelike immersive experience. However, several applications, such as medical visualization, do require a complete 3D representation with accurate and true-to-scale measurements. To this end, in Chapter 2, from the various of techniques of acquiring and reconstructing 3D representation, we focus at true-to-scale 3D reconstruction from multiple 2D perspectives. We describe a previously proposed method that calibrates a network of cameras and achieves true-to-scale 3D reconstruction without requiring an extraneous known object. In this method, preservation of scale is ensured by having at least one stereo camera pair in the network. We also discuss basics of light field, various representations of light field, projection-based LFDs and rendering of light fields on such displays.

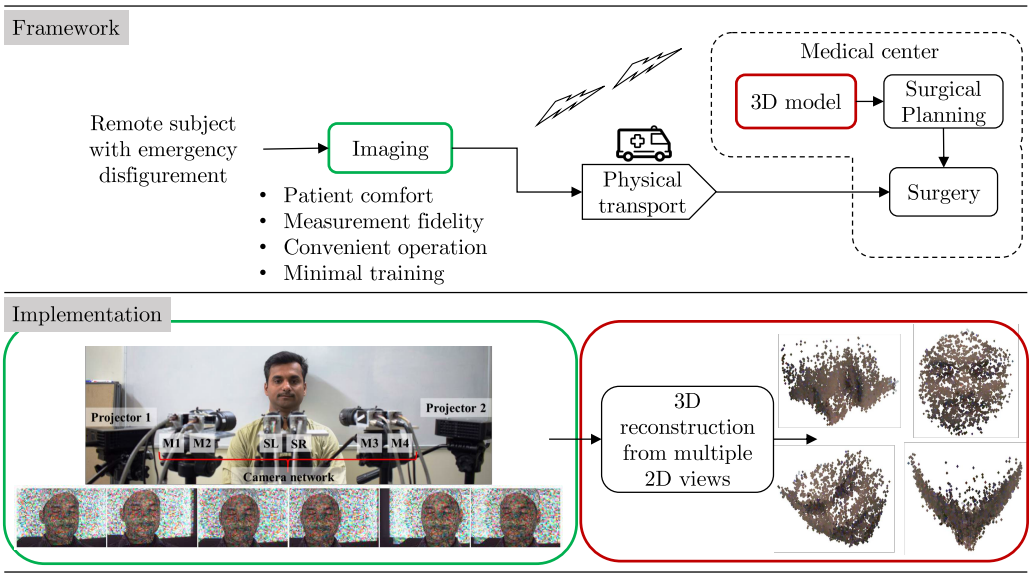

Figure 1.3: Proposed hybrid teleophthalmology paradigm.

Contributions

The method described in this chapter has been proposed and published earlier in conference proceedings [92, 93]. The contents of this section have been published originally in the PhD thesis of Dr. Kiran Kumar Vupparaboina [94]. They are reproduced here for a comprehensive discussion. Subsequently, since presence of stereoscopic camera(s) in the network is the key requirement for non-intrusive true-to-scale auto-calibration, we investigate effects of imperfections in stereo cameras on calibration accuracy.

Publications

C01 R. R. Tamboli, K. K. Vupparaboina, M. S. Reddy, P. A. Kara, A. Cserkaszky, M. G. Martini, A. Richhariya, and S. Jana, "Towards euclidean auto-calibration of stereo camera arrays," in *Optical System Alignment, Tolerancing, and Verification XII*, vol. 10747, 2018, pp. 135–144.

1.5.2 Chapter 3: Illustrative Healthcare Application of Envisaged End-to-end System

The auto-calibration method described in Chapter 2 has been applied previously for 3D reconstruction of inanimate objects from their multiple 2D views [92]. Motivated by the accuracy of the reconstructed point clouds, we investigated if this method

can be used for surface imaging and 3D reconstruction in an healthcare application. We demonstrate reconstruction of 3D face from its multiple views towards facilitating oculofacial surgical planning and assessment. To this end, we first reconstruct a dense point cloud from multiple images of a 3D-printed head model. Since the ground truth 3D information of head model is available, this approach serves as an experimental corroboration of the method described in Chapter 2. The imaging platform used in the process fulfilled many desired attributes of a facial imaging system such as instantaneous and contact-free acquisition, ease of deployment and operation, and accurate true-to-scale measurement. We then provide preliminary results of 3D reconstitution of human face and discuss its potential in teleophthalmology. Specifically, we propose a hybrid teleophthalmology framework. A graphical summary of Chapter 3 is given in Figure 1.3.

Contributions

The proposed hybrid teleophthlmology framework potentially reduces the time between admission of the subject and commencement of surgery. Furthermore, it can also enable remote post-operative assessment and as a consequence, reduce hospitalization time and the number of visits to the surgical facility. Towards realizing this goal, we present true-to-scale 3D reconstruction of a human face, and a preliminary surface fitting on recovered sparse point cloud. In this chapter, Section 3.2 is based on a joint publication (C04 listed below) [96] and has been published previously in the PhD thesis of Dr. Kiran Kumar Vupparaboina [94]. It is reproduced here for the sake of completeness.

Publications

C02 R. R. Tamboli, V. Sandu, S. Nerasala, A. Richhariya, K. K. Vupparaboina, and S. Jana, "Novel hybrid teleophthalmology: Technological case for oculofacial surgery," in *IEEE Global Humanitarian Technology Conference*, 2019, pp. 1–8.

C03 K. K. Vupparaboina, R. R. Tamboli, S. Manne, P. A. Kara, M. G. Martini, A. Barsi, A. Richhariya, and S. Jana, "Towards true-to-scale 3d reconstruction of the human face using structured light projection and off-the-shelf cameras," in *International Conference on 3D Immersion*, 2018, pp. 1–7.

C04 K. K. Vupparaboina, R. R. Tamboli, S. Manne, A. Richhariya, M. Naik, and S. Jana, "Oculofacial surgical planning: true-to-scale 3d feature quantification using multicamera network," in *IEEE Annual India Conference*, 2016, pp. 1–6.

1.5.3 Chapter 4: Perceptual Quality Assessment under Spatial Distortions

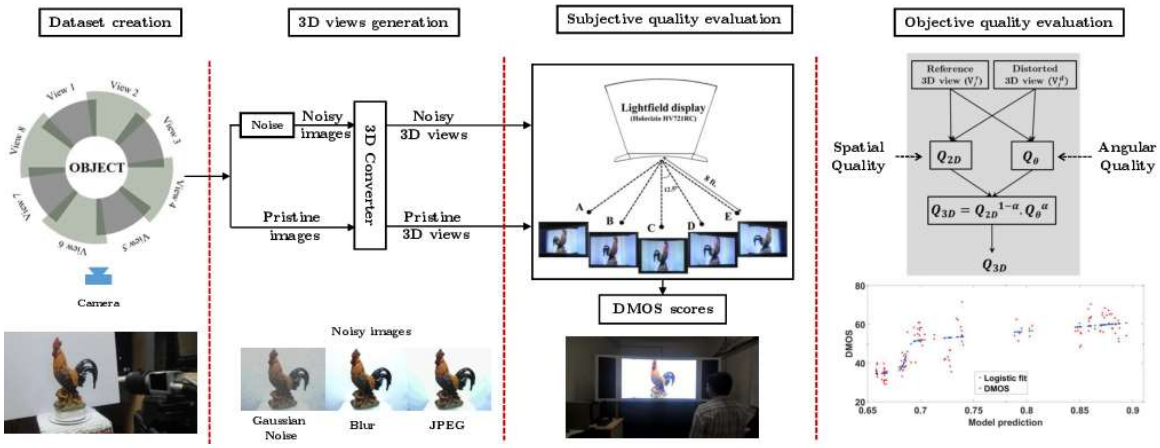

Figure 1.4: Perceptual quality assessment of super-multiview content under spatial distortions.

In this chapter, we present the results of subjective and objective quality assessment of static SMV content, conducted with twenty subjects on a glasses-free horizontal-parallax (HP) LFD. First, we created a multiview image dataset using three real objects, with high angular resolution. The display generates 3D views from fixed subsets of consecutive images from the aforementioned dataset. Next, we extended the standard guidelines of subjective assessment to our evaluation environment. In the course, we made recommendations on acquiring SMV data as well as conducting subjective assessment. Subsequently, we observed that the existing 2D FR objective quality metrics, applied to individual 3D views, are inadequate for 3D quality assessment. To fill the gap, we proposed a 3D FR objective quality metric. For every 3D view, the proposed metric combines spatial information from each constituent image and angular information extracted from pairs of consecutive images. Finally, we showed that the proposed metric correlates significantly with subjective scores, outperforming existing 2D metrics. The pooling of spatial and angular information takes into account the fact that angular information plays a crucial role in 3D perception. Figure 1.4 provides a graphical summary of this work. The proposed metric was validated and also applied in various test cases.

Contributions

In this work, we made the following contributions. We first created a high-angular-resolution turntable dataset of three real objects. We then extended of the standard

guidelines on subjective quality evaluation of 2D content to subjective quality evaluation of SMV content visualized on an LFD. Finally, a 3D FR objective quality metric which takes into account the spatio-angular nature of the 3D content was presented. We used the proposed metric to assess the quality of LF content rendered in a novel display-independent representation, and also key frames of SMV video.

Publications

J01 R. R. Tamboli, B. Appina, S. Channappayya, and S. Jana, "Super-multiview content with high angular resolution: 3D quality assessment on horizontal-parallax lightfield display," *Signal Processing: Image Communication*, vol. 47, pp. 42–55, 2016.

J02 A. Cserkaszky, P. A. Kara, R. R. Tamboli, A. Barsi, M. G. Martini, L. Bokor, and T. Balogh, "Angularly continuous light-field format: Concept, implementation, and evaluation," *Journal of the Society for Information Display*, vol. 27, no. 7, pp. 442–461, 2019.

C05 R. Tamboli, K. K. Vupparaboina, J. Ready, S. Jana, and S. Channappayya, "A subjective evaluation of true 3d images," in *International Conference on 3D Imaging*, 2014, pp. 1–8.

C06 R. R. Tamboli, A. Cserkaszky, P. A. Kara, A. Barsi, and M. G. Martini, "Objective quality evaluation of an angularly-continuous light-field format," in *International Conference on 3D Immersion*, 2018, pp. 1–8.

C07 R. R. Tamboli, P. A. Kara, A. Cserkaszky, A. Barsi, M. G. Martini, B. Appina, S. S. Channappayya, and S. Jana, "3d objective quality assessment of light field video frames," in *3DTV-Conference: The True Vision-Capture, Transmission and Display of 3D Video*, 2018, pp. 1–4.

1.5.4 Chapter 5: Perceptual Quality Assessment under Angular Distortions

SMV LFDs provide glasses-free immersive 3D experience via virtually-continuous motion parallax. Such displays require input views acquired at high angular resolution, which could be economically and technologically prohibitive. To achieve high view density, additional views are often synthesized using the captured ones. However, a critical proportion of the views should be captured by actual cameras to ensure

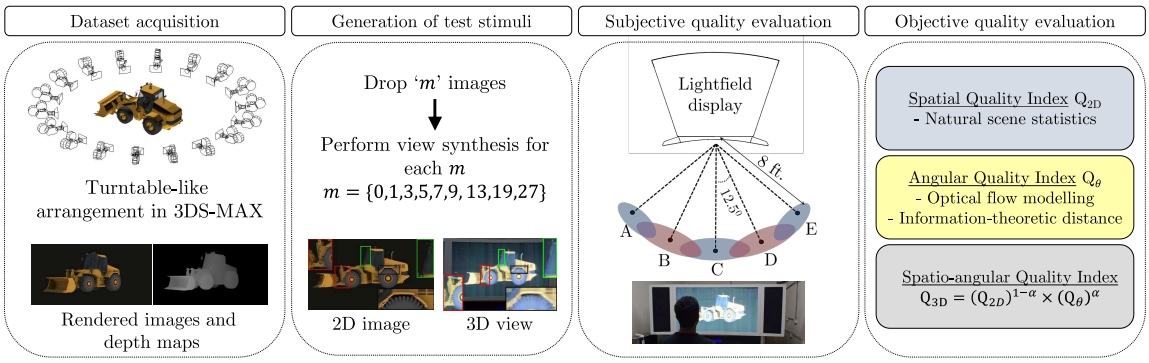

Figure 1.5: Perceptual quality assessment of super-multiview lightfield content under angular distortions.

acceptable quality of the 3D content. In order to determine the said proportion, we subjectively assessed the degradation in perceptual quality vis-à-vis the proportion of synthesized views. A qualitative analysis of the subjective scores ascertained that the quality of 3D perception is mainly affected by discontinuities in motion parallax introduced by view synthesis, along with spatial quality of the individual views. The subjective scores collected in the process exhibited a three-regime behaviour, corresponding to small, moderate and large proportion of synthesized views. Subsequently, we proposed a FR spatio-angular objective quality index to assess quality of such content. The spatial quality index was based on the existing natural scene statistics (NSS) framework. Specifically, we modified the spatial quality index proposed earlier by introducing spatial weights to the metric. The weights accounted for the distance of a synthesized image from the two reference images used for view synthesis. The angular quality index, instead of computing structural differences between optical flow arrays as in the previous metric, modelled the optical flow using bivariate generalized Gaussian distribution (BGGD) and computed information-theoretic distance between the fitted models. A high agreement, in terms of linear correlation coefficient (LCC), Spearman's rank order correlation coefficient (SROCC) and root mean squared error (RMSE), was found between the subjective scores and the objective scores obtained using the proposed metric. A graphical abstract of this chapter is provided in Figure 1.5.

Contributions

In this work, we made the following contributions. We created a synthetic high-angular-resolution dataset along with ground truth depth maps using a turntable-like setting. For the objective quality assessment, we proposed a FR metric guided by

information theory. Finally, to validate our hypothesis on real-life data, we created a high-angular-resolution tristereo-based dataset of several real-life objects using automated turntable setup.

Publications

C08 R. R. Tamboli, P. A. Kara, N. Bisht, A. Barsi, M. G. Martini, and S. Jana, "Objective quality assessment of 2d synthesized views for light-field visualization," in *International Conference on 3D Immersion*, 2018, pp. 1–7.

C09 R. R. Tamboli, B. Appina, S. S. Channappayya, and S. Jana, "Achieving high angular resolution via view synthesis: quality assessment of 3d content on super multiview lightfield display," in *International Conference on 3D Immersion*, 2017, pp. 1–8.

C10 R. R. Tamboli, M. S. Reddy, P. A. Kara, M. G. Martini, S. S. Channappayya, and S. Jana, "A high-angular-resolution turntable data-set for experiments on light field visualization quality," in *Tenth International Conference on Quality of Multimedia Experience*, 2018, pp. 1–3.

J03 B. Appina, R. R. Tamboli, S. Channappayya, and S. Jana, "Super-Multiview Lightfield Content: Objective Quality Assessment via Spatio-Angular Modeling", *Under preparation.*

1.5.5 Chapter 6: Preferred Views of 3D Objects: A Preliminary Study using Lightfield Displays

As alluded to earlier, wide-baseline projection-based LFDs need content with large FOV, and with high spatial and angular quality. Accordingly, spatial, angular and spatio-angular aspects of LF visualization as well as the effect of various zooming levels of the displayed content and region of interest (ROI) on QoE has been explored recently [106]. However, there has been no systematic attempt to see how the features of the content itself affect the visual experience. In this backdrop, we studied the effects of some primitive features of the LF content on QoE (summarized in Figure 1.6b). Subsequently, we investigated if the content rendered on LFDs has any preferred orientation and what determines such preference. LF visualization allows the users to freely choose a preferred location for observation within the display's valid FOV. If we consider static objects and scenes, then the user's location determines the perceived orientation of the visualized content. In case of interactive LF

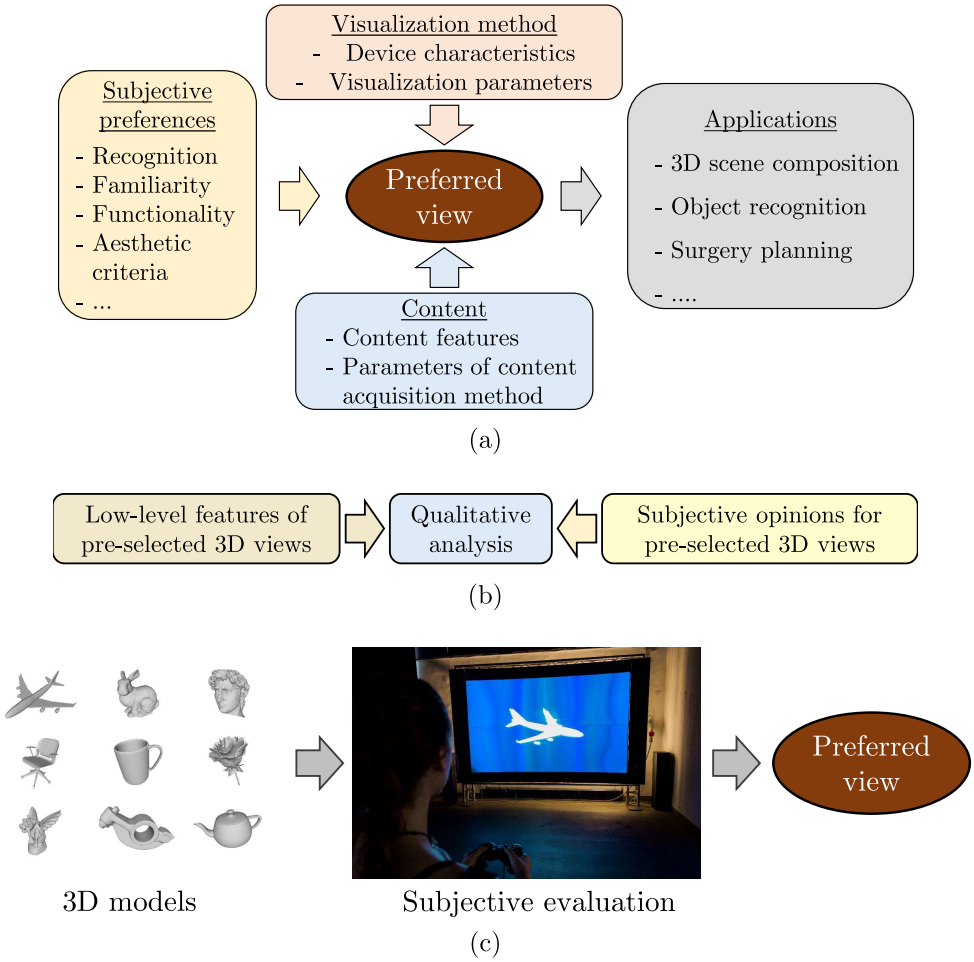

Figure 1.6: Preferred/canonical views: (a) determining factors and applications. (b) and (c) methodologies adopted to seek preferred views.

visualization, the arbitrary rotation of content enables efficient orientation changes without the need for actual user movement. However, the preference of content orientation is a subjective matter, yet it is possible to be objectively assessed as well. In this work, we present a the results of the subjective test that we carried out on an LFD (shown in Figure 1.6c) that addresses static content orientation preference. The state-of-the-art objective methodologies were used to evaluate the experimental setup and the content.

Contributions

Several studies have addressed the topic of preferred orientation of 3D objects visualized on 2D displays, and its implications. In contrast, the work presented in this dissertation, to the best of the authors' knowledge, is the first ever that is conducted

on a glasses-free LFD. In this pilot study, we proposed a protocol to seek preferred views subjectively, which can guide similar studies in future. We also discussed the difficulties involved in designing an objective method to predict preferred orientations, when visualized on an LFD.

Publications

C11 R. R. Tamboli, B. Appina, P. A. Kara, M. G. Martini, S. S. Channappayya, and S. Jana, "Effect of primitive features of content on perceived quality of light field visualization," in *Tenth International Conference on Quality of Multimedia Experience*, 2018, pp. 1–3.

C12 R. R. Tamboli, P. A. Kara, A. Cserkaszky, A. Barsi, M. G. Martini, and S. Jana, "Canonical 3d object orientation for interactive light-field visualization," in *Applications of Digital Image Processing XLI*, vol. 10752, 2018, pp. 77–83.

1.5.6 Chapter 7: Summary and Future Directions

In this chapter, we provide a summary of this dissertation. We also discuss future research directions in acquisition, reconstruction, immersive visualization and quality assessment of 3D content. As a part of future work, we discuss SMV video quality assessment. Some of the early works in this regard resorted to seated subjects for video quality assessment [109], which indeed is sufficient for several use cases. However, in futuristic use cases, viewers may need to move while consuming dynamic content visualized on an LFD. To this end, in an early investigation, we proposed a subjective assessment protocol where the subjects were allowed to move on a predetermined trajectory during SMV video playback [110, 111].

Publications

C13 P. A. Kara, R. R. Tamboli, A. Cserkaszky, M. G. Martini, A. Barsi, and L. Bokor, "The viewing conditions of light-field video for subjective quality assessment," in *International Conference on 3D Immersion*, 2018, pp. 1–8.

C14 A. Cserkaszky, P. A. Kara, R. R. Tamboli, A. Barsi, M. G. Martini, and T. Balogh, "Light-field capture and display systems: limitations, challenges, and potentials," in *Novel Optical Systems Design and Optimization XXI*, vol. 10746, 2018, pp. 10–18.

C15 ——, "The perceived quality of light-field video services," in *Applications of Digital Image Processing XLI*, vol. 10752, 2018, pp. 138–147.

1.5.7 Other contributions not elaborated in this dissertation

The widespread adaptation of LFDs ultimately depends on associated QoE, which is achieved through excellence regarding the key performance indicators (KPIs) of the display system and the visualized content. To this end, we reviewed certain KPIs of LF display and content, as well as discussed their implications on various stages of LF media pipeline [112]. One of the key KPI of projection-based LFDs is angular resolution, which determines the smoothness of the parallax and also shapes the valid viewing area of LF visualization. Higher angular resolutions support greater viewing distances. Accordingly, we explored the effect of distance between the subject and the display screen on binocular disparity [113]. Such a study could help manufacturers create guidelines and recommendations regarding operational distance.

Additional publications

C16 P. A. Kara, R. R. Tamboli, A. Cserkaszky, A. Barsi, A. Simon, A. Kusz, L. Bokor, and M. G. Martini, "Objective and subjective assessment of binocular disparity for projection-based light field displays," in *International Conference on 3D Immersion*, 2019, pp. 1–8.

J04 P. A. Kara, R. R. Tamboli, O. Doronin, A. Cserkaszky, A. Barsi, Z. Nagy, M. G. Martini, and A. Simon, "The key performance indicators of projection-based light field visualization," *Journal of Information Display*, vol. 20, no. 2, pp. 81–93, 2019.

1.6 Highlights

- ***Multiview datasets***:
  - Multiview datasets of a 3D head model [96] and human face [25, 97] captured using multi-camera network. (Section 3.3.3)
  - Turntable datasets of real-life objects [98, 105]. (Sections 4.3 and 5.8)
  - Synthetic dataset generated using a turntable-like arrangement [98, 104]. (Sections 4.7.1 and 5.3.1)

- ***Methodologies for subjective quality assessment of SMV content*** visualized on projection-based LFDs for the following test cases:
  - Static content evaluated from predetermined viewing locations [98, 104]. (Sections 4.4 and 5.4)
  - Interactive static content with free movement of observers [108]. (Section 6.2)
- ***Full-reference metrics for objective quality evaluation of SMV content*** captured for visualization on projection-based LFDs.
  - A metric for content with spatial distortions [98]. (Section 4.5)
  - A metric for content with angular distortions. (Section 5.5)

Chapter 7

Summary and Future Directions

In this dissertation, we envisioned a generic end-to-end remote collaboration system, described currently available elements of such a system, and outlined the opportunities and challenges towards consumer adaptation. We briefly reviewed a variety of applications that currently employ three-dimensional (3D) visualization and remote collaboration, and discussed how emerging 3D media technologies can enhance quality of experience (QoE). Specific to some of these applications, we identified and addressed a subset of problems in acquisition, 3D reconstruction, immersive visualization and quality assessment of super-multiview (SMV) content. In Table 7.1, we summarize this dissertation under a common framework.

7.1 Summary

We now provide a brief summary of each chapter in the following. In chapter 2, we presented theoretical underpinnings of true-to-scale 3D reconstruction from multiple views, as well as 3D visualization. In Chapter 3, we used this method to reconstruct 3D face and subsequently proposed a hybrid teleophthalmology paradigm. In chapters 4 and 5, we discussed novel subjective and objective methodologies to assess perceptual quality of SMV content visualized on projection-based lightfield displays (LFDs). In Chapter 6, we sought preferred orientations of various 3D objects via interactive visualization.

Table 7.1: A summary of the dissertation.

	Chapter 3	**Chapter 4**	**Chapter 5**	**Chapter 6**
Content	Real-life	Real-life, Synthetic	Synthetic	Synthetic, 3D Models
Method of acquisition	Sparse network of multiple cameras	Turntable	Turntable	–
Method of 3D generation	Reconstruction of 3D geometry via auto-calibration	Conversion of SMV content to display-specific light field		Display-specific ray tracing
Visualization	2D and lightfield display	Lightfield display		
Subjective evaluation	–	Perceptual quality		Preferred orientation(s)
Objective evaluation	Reconstruction accuracy	Signal fidelity and perceptual quality		–

7.1.1 True-to-scale 3D Reconstruction and Immersive Visualization

In many applications of 3D visualization, such as telepresence, scene contents do not change drastically and the scene span is limited. Therefore, very dense sampling of light field can be achieved. Existing LFDs can convert SMV video streams to display-specific light fields and provide lifelike immersive visual experience. However, several applications, such as medical visualization, do require a complete 3D representation with accurate and true-to-scale measurements. To this end, in Chapter 2, from the wide variety of techniques of acquiring and reconstructing 3D representation [73, 74, 287,288], we focused at true-to-scale 3D reconstruction from multiple two-dimensional (2D) perspectives.

We described a previously proposed method that calibrates a network of cameras and achieves true-to-scale 3D reconstruction. In comparison to the widely-used assisted methods of camera calibration, the said method does not require any known objects or markers to be introduced into the scene. Instead, it only requires presence of a stereoscopic camera with known baseline separation in the camera network. This constraint preserves the scale of 3D reconstruction without any extraneous assistance and therefore is an auto-calibration method. Given the set of corresponding points across the cameras in the network, the algorithm first obtains X and Y coordinates of the 3D points by stereo triangulation. It then solves a multi-objective optimization problem to remove scale ambiguity and estimate Z coordinates by minimizing re-projection errors for monocular cameras. Since a stereo camera is a key element

in this approach, we also investigated effect of imperfections in stereo camera(s) on calibration accuracy. We then described the working principle of projection-based LFDs and explained how 3D, multiview video plus depth (MVD) and SMV contents are visualized on such displays.

7.1.2 Illustrative Healthcare Application of Envisaged End-to-end System

The auto-calibration method described in Chapter 2 has been applied previously for 3D reconstruction of inanimate objects from their multiple 2D views [92]. Motivated by the accuracy of the reconstructed point clouds, we investigated if this method can be used for surface imaging and 3D reconstruction in an healthcare application. Specifically, a dense 3D reconstruction of a 3D head model from its multiple 2D views was obtained. The imaging platform used in the process fulfilled many desired attributes of a facial imaging system such as instantaneous and contact-free acquisition, ease of deployment and operation, and accurate true-to-scale measurement. Subsequently, we proposed a hybrid teleophthalmology paradigm. In this framework, the existing primary healthcare centres of eyecare networks, located in remote areas, would be equipped with the aforementioned imaging system, in addition to the already existing conventional ophthalmology equipment. In an emergency situation requiring facial surgery, the primary healthcare center would capture facial multiview data and send it to the operation center to facilitate pre-operative planning while the subject is in transit. This approach would potentially reduce the time between admission of the subject and commencement of surgery. Furthermore, it can also enable remote post-operative assessment and as a consequence, reduce hospitalization time and the number of visits to the surgical facility. Towards realizing this goal, we presented true-to-scale 3D reconstruction of a human face, and a preliminary surface fitting on recovered sparse point cloud. We then discussed its implications in the aforementioned hybrid teleophthalmology framework.

7.1.3 Perceptual Quality Assessment under Spatial Distortions

In Chapter 4, we presented subjective and objective quality assessment of static SMV content, conducted with twenty subjects on a glasses-free horizontal-parallax (HP) LFD. First, we created a high-angular-resolution static SMV dataset of three real

objects in a turntable setting. The conversion of SMV content to display-specific light field was achieved using a proprietary software. However, the principle behind such conversion process has been described in Chapter 2. A few spatial distortions were introduced to the 2D images before such conversion to generate multiple test conditions. We then conducted subjective assessment, by adapting the standard guidelines of such studies for our evaluation environment. We then proposed a full-reference (FR) spatio-angular objective quality index. For every 3D view, the proposed method combines spatial information from each constituent image and angular information extracted from pairs of consecutive images. The pooling of spatial and angular information captures the interplay between spatial and angular components. The proposed method correlates significantly with subjective scores in terms of various performance criteria. We then validated the proposed method and tested it against an unseen synthetic turntable dataset and an unseen distortion. Furthermore, we studied if the proposed method can assess the quality of key frames of SMV videos. Next, we evaluated a novel display-agnostic angularly-continuous light field representation against the conventional perspective camera representation, using the proposed objective quality index. We observed that the advantage of display-independence comes without any compromise in objective quality.

7.1.4 Perceptual Quality Assessment under Spatial Distortions

SMV LFDs provide glasses-free immersive 3D experience via virtually-continuous motion parallax. Such displays require input views acquired at high angular resolution, which could be economically and technologically prohibitive. To achieve high view density, additional views are often synthesized using the captured ones. However, a critical proportion of the views should be captured by actual cameras in order to ensure acceptable quality of the 3D content. In order to determine the said proportion, we subjectively assessed the degradation in perceptual quality vis-à-vis the proportion of synthesized views. A qualitative analysis of the subjective scores ascertained that the quality of 3D perception is mainly affected by discontinuities in motion parallax introduced by view synthesis, along with spatial quality of the individual views. The subjective scores collected in the process exhibited a three-regime behaviour, corresponding to small, moderate and large proportion of synthesized views. Subsequently, we proposed a FR spatio-angular objective quality index to assess the quality of such content. The spatial quality index was based on the existing natural scene statistics

(NSS) framework. Specifically, we modified the spatial quality index proposed earlier by introducing spatial weights to the metric. The weights accounted for the distance of a synthesized image from the two reference images used for view synthesis. The angular quality index, instead of computing structural differences between optical flow arrays as in the previous method, modelled the optical flow as bivariate generalized Gaussian distribution (BGGD) and computed dissimilarity, guided by information-theoretic distances, between the fitted models. A high agreement, in terms of various performance metrics, was found between the subjective opinions and the objective scores.

7.1.5 Preferred Views of 3D Objects: A Preliminary Study using Lightfield Displays

During the subjective evaluations reported in Chapter 4, it was observed that certain 3D views were preferred over the others, irrespective of the spatial distortions added to the content. Against this backdrop, we investigated how certain features of the content, other than spatio-angular properties, affect the experience of lightfield (LF) visualization. Based on the insights gained from a qualitative analysis of DMOS vis-à-vis primitive features of the content, we designed an interactive study to seek preferred views of 3D objects. In this experiment, participants were asked to orient selected 3D models visualized on an LFD in a way that the obtained view was most appealing to them. The results of this study, which was the first ever to be conducted on an LFD to the best of our knowledge, were found to corroborate with many previous studies conducted on 2D displays. We then discussed its possible implications on content acquisition, reconstruction and rendering. Finally, we stated difficulties involved in designing an objective method for prediction of preferred orientations of 3D objects, when visualized on a glasses-free LFD.

7.2 Future Directions

As alluded to several times in this dissertation, 3D visualization has become an indispensable component of many applications in education, entertainment, scientific research, telepresence, healthcare, defense, surveillance etc. While these areas have vastly different use cases, requirements and constraints, the quality of perceptual experience remains the single most important factor. Accordingly, the entire pipeline of 3D media comprising of acquisition, coding, reconstruction/representation, rendering

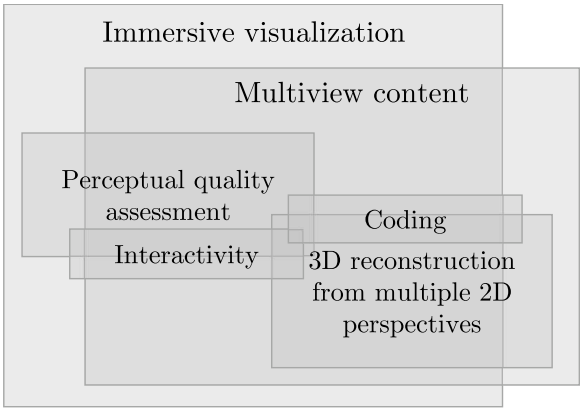

Figure 7.1: Outlook and future directions.

and visualization should ideally be optimized for the best QoE. While QoE can be assessed subjectively, computer algorithms that predict subjective scores, possibly without any reference, are instrumental in accelerating the overall progress of 3D media adaptation. Such methods eliminate/minimize the need of tedious subjective evaluations, the duration of which scales up quickly with the number of test cases.

Among the multitude of existing and emerging display technologies, full-parallax (FP) LFDs can offer a true multi-user immersive 3D experience. However, the current generation of LFDs are horizontal-parallax-only (HPO), which is sufficient for several use cases. Due to the large form factor, high equipment cost, and training required for operation and maintenance etc. (see Table 2.2), their entry into the consumer market may take a few decades. Efforts towards solving these issues are already being undertaken. Anticipating the future proliferation of light field ecosystem, LF content acquisition, representation, compression and processing has already progressed significantly leading to standardization initiatives [39, 289, 290]. This has opened several avenues of research and development in theory, algorithms, software and hardware pertaining to 3D media in general and light field in particular. These future directions are summarized graphically in Figure 7.1. In this dissertation, we discussed 3D reconstruction (Chapter 2 and 3), quality assessment (Chapters 4 and 5) and interaction (Chapter 6) separately with an overview of the interplay among them. However, in a futuristic end-to-end remote collaboration system that we envisaged earlier in Chapter 1, typically 2D images would be captured using multiple cameras, compressed and then communicated. The received images, depending on the use case, would either be used to reconstruct 3D model of the scene or converted to device-specific light field representation. Signal fidelity, perceptual quality and capability to interact with resultant 3D content is affected by various processing stages. Thus, it

is necessary to optimize entire 3D media pipeline for desired quality of service (QoS) and QoE. In other words, in a practical use case, the aforementioned components are strongly connected. We have attempted to indicate such interrelation among various blocks via overlaps in the Figure 7.1. In this backdrop, we plan to further our research in existing and new directions, as explained next.

7.2.1 Acquisition and Reconstruction

We plan to improve/upgrade our multiview capture systems and generate a wide variety of datasets using turntable and multicamera arrangements, mainly aimed at achieving high-quality 3D reconstruction from multiple views and perceptual quality assessment. To begin with, the manually-operated turntable setup presented in Chapter 4 was upgraded to an automated setup. We presented a high-angular-resolution turntable dataset of real-life objects captured using this setup in Section 5.8. We shall incorporate controlled diffused light sources in our imaging systems in near future. Finally, a SMV video capture, storage and processing system is planned to be set up.

We presented two implementations of facial imaging platform Chapter 3, one for imaging a 3D head model and another for imaging a human test subject. We plan to make it robust and economical. To this end, instead of the present near-instantaneous acquisition, we shall incorporate hardware-trigger-based simultaneous capture, and also make the entire imaging operation described in Section 3.3.3 automatic. We plan to use densely packed cameras. As discussed previously, we shall investigate if inexpensive universal serial bus (USB) cameras can be used in such a system, and if the associated shortcomings can be mitigated in software, without compromising the quality of the outcome. We aim to study the quality of our 3D reconstruction against lens distortions, imperfect stereo arrangement, incorrect point correspondences, sub-optimal lighting conditions, arrays of non-identical cameras and so on. In near future, we also plan to modify/improve our method in order to calibrate a larger network of cameras. Finally, we aim to achieve a dense, true-to-scale, parameterized, deformable and photorealistic 3D representation of the human face.

7.2.2 Immersive Visualization and Quality Assessment

In the area of perceptual quality assessment of SMV content, we plan to introduce a wider variety of objects/scenes and test cases (spatial and angular distortions/distortions due to source and channel coding) while keeping our subjective studies tractable. Consequently, we will need improved objective quality assessment

methods, model-based as well as learning-based, to predict perceptual quality, preferred orientations and so on.

In this work, we limited the scope of our investigation to turntable datasets in order to study spatio-angular aspects of perceptual quality on an LFD. However, real-life applications such as telepresence require wide-baseline content acquired using an array of cameras [63, 109]. Therefore, we plan to extend our work to SMV video compression and quality assessment [103]. LF video quality assessment is very challenging compared to 2D and stereoscopic video quality assessment due to significantly higher volumes of information present in an LF video. Some of the early works resorted to seated subjects for video quality assessment [109], which indeed is sufficient for several use cases. However, completely new use cases may emerge in future, such as LF gaming, where the content is dynamic and subjects are free to/need to move in front of an LFD. Every subject, depending on his/her position in from of an LFD, may receive a completely different stimulus at a given time. This issue has been considered recently [18]. Consequently, a few early subjective assessment protocols for video quality assessment have proposed predetermined trajectory for observer movement [89, 90, 110, 111, 256, 291].

Summa summarum, we explored various aspects of novel 3D media technologies in this dissertation, such as reconstruction of 3D objects from multiple 2D views acquired using standard cameras, rendering of such 3D content on 3D displays to create the experience of immersive visualization, and assessment of the resulting quality of experience. We envisaged an end-to-end system that allows image acquisition and 3D reconstruction at remote location, transmission to user location, and eventual 3D visualization. While various aforementioned aspects of such a system have received significant attention, relatively little effort has been directed at understanding and quantifying the user experience. To this end, we primarily focused on developing novel methods for subjective and objective assessment of perceptual quality of SMV content. Finally, in view of various existing technological gaps and in accordance with trends in emerging 3D media technologies, we outlined a few short-term as well as long-term plans for future explorations.

Acronyms

2D two-dimensional.
3D three-dimensional.
3DSwIM 3D Synthesized view Image quality Metric.

ABS acrylonitrile butadiene styrene.
ACR absolute category rating.
ASIFT affine scale-invariant feature transform.

BGGD bivariate generalized Gaussian distribution.

CG computer-generated.

DERS depth estimation reference software.
DIBR depth image-based rendering.
DMOS difference mean opinion score.
DoF depth of field.
DOG difference of Gaussians.
DSLR digital single-lens reflex.

FOV field of view.
FP full-parallax.
FR full-reference.
FSIM Feature SIMilarity index.

GPS global positioning system.

HDR high dynamic range.

HEVC high efficiency video coding.
HMD head-mounted display.
HP horizontal-parallax.
HPO horizontal-parallax-only.
HUD head-up display.
HVS human visual system.

IFC information fidelity criterion.
IITH Indian Institute of Technology Hyderabad.
IQA image quality assessment.
IR infrared.

KLD Kullback–Leibler divergence.
KPI key performance indicator.

LCC linear correlation coefficient.
LF lightfield.
LFD lightfield display.

MCCV Monte-Carlo cross validation.
MGGD multivariate Generalized Gaussian distribution.
MOS mean opinion score.
MPEG Moving Picture Experts Group.
MS-SSIM Multi-Scale Structural SIMilarity.
MSE mean squared error.
MVD multiview video plus depth.

NCC normalized cross-correlation.
NR no-reference.
NSS natural scene statistics.

PLCC Pearson linear correlation coefficient.
PSNR peak signal-to-noise ratio.

QoE quality of experience.
QoS quality of service.

RANSAC random sample consensus.
RMSE root mean squared error.
ROI region of interest.
RR reduced-reference.

SBA sparse bundle adjustment.
SIFT scale-invariant feature transform.
SMV super-multiview.
SROCC Spearman's rank order correlation coefficient.
SSIM Structural SIMilarity.
SURF speeded-up robust features.

UGGD univariate generalized Gaussian distribution.
UHD ultra-high-definition.
USB universal serial bus.

VR virtual reality.
VSRS view synthesis reference software.

WCG wide color gamut.

References

[1] A. Steed and R. Schroeder, "Collaboration in immersive and non-immersive virtual environments," in *Immersed in Media.* Springer, 2015, pp. 263–282.

[2] H. Fuchs, A. State, and J.-C. Bazin, "Immersive 3d telepresence," *Computer*, vol. 47, no. 7, pp. 46–52, 2014.

[3] C. C. Bracken and P. Skalski, *Immersed in media: Telepresence in everyday life.* Routledge, 2010.

[4] M. Lombard, F. Biocca, J. Freeman, W. IJsselsteijn, and R. J. Schaevitz, *Immersed in media: Telepresence theory, measurement & technology.* Springer, 2015.

[5] G. Kurillo, R. Vasudevan, E. Lobaton, and R. Bajcsy, "A framework for collaborative real-time 3d teleimmersion in a geographically distributed environment," in *Tenth IEEE International Symposium on Multimedia*, 2008, pp. 111–118.

[6] D. Gotsch, X. Zhang, T. Merritt, and R. Vertegaal, "Telehuman2: A cylindrical light field teleconferencing system for life-size 3d human telepresence," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, vol. 18, 2018, pp. 1–10.

[7] A. Maimone and H. Fuchs, "Encumbrance-free telepresence system with real-time 3d capture and display using commodity depth cameras," in *International Symposium on Mixed and Augmented Reality*, 2011, pp. 137–146.

[8] A. Cserkaszky, A. Barsi, Z. Nagy, G. Puhr, T. Balogh, and P. A. Kara, "Real-time light-field 3d telepresence," in *7th European Workshop on Visual Information Processing*, 2018, pp. 1–5.

[9] H. Baker, D. Tanguay, I. Sobel, D. Gelb, M. E. Goss, W. B. Culbertson, and T. Malzbender, "The coliseum immersive teleconferencing system," in *International Workshop on Immersive Telepresence*, vol. 6, 2002.

[10] H. Fuchs, G. Bishop, K. Arthur, L. McMillan, R. Bajcsy, S. Lee, H. Farid, and T. Kanade, "Virtual space teleconferencing using a sea of cameras," in *Proc. First International Conference on Medical Robotics and Computer Assisted Surgery*, vol. 26, 1994.

[11] T. Kanade, P. Rander, and P. Narayanan, "Virtualized reality: Constructing virtual worlds from real scenes," *IEEE multimedia*, vol. 4, no. 1, pp. 34–47, 1997.

[12] M. Gross, S. Würmlin, M. Naef, E. Lamboray, C. Spagno, A. Kunz, E. Koller-Meier, T. Svoboda, L. Van Gool, S. Lang *et al.*, "blue-c: a spatially immersive display and 3d video portal for telepresence," *ACM Transactions on Graphics (TOG)*, vol. 22, no. 3, pp. 819–827, 2003.

[13] A. Kulik, A. Kunert, S. Beck, R. Reichel, R. Blach, A. Zink, and B. Froehlich, "C1x6: A stereoscopic six-user display for co-located collaboration in shared virtual environments," *ACM Trans. Graph.*, vol. 30, no. 6, p. 1–12, 2011.

[14] "Holoportation," https://www.microsoft.com/en-us/research/project/holoportation-3/, Accessed: June 2020.

[15] "Holopresence," http://holografika.com/holopresence/, Accessed: June 2020.

[16] F. Dufaux, B. Pesquet-Popescu, and M. Cagnazzo, *Emerging technologies for 3D video: creation, coding, transmission and rendering.* John Wiley & Sons, 2013.

[17] A. M. Kondoz and T. Dagiuklas, *3D Future Internet Media.* Springer, 2014.

[18] P. A. Kara, M. G. Martini, Z. Nagy, and A. Barsi, "Cinema as large as life: Large-scale light field cinema system," in *International Conference on 3D Immersion*, 2017, pp. 1–8.

[19] N. Efrat, P. Didyk, M. Foshey, W. Matusik, and A. Levin, "Cinema 3d: Large scale automultiscopic display," *ACM Trans. Graph.*, vol. 35, no. 4, 2016.

[20] A. Ashraf, D. Collins, M. Whelan, R. O'Sullivan, and P. Balfe, "Three-dimensional (3d) simulation versus two-dimensional (2d) enhances surgical skills acquisition in standardised laparoscopic tasks: a before and after study," *International Journal of Surgery*, vol. 14, pp. 12–16, 2015.

[21] "Coronavirus disease (COVID-19) pandemic ," https://www.who.int/emergencies/diseases/novel-coronavirus-2019, Accessed: June 2020.

[22] A. Kapoor, S. Guha, M. K. Das, K. C. Goswami, and R. Yadav, "Digital healthcare: The only solution for better healthcare during covid-19 pandemic?" *Indian Heart Journal*, 2020.

[23] H. O. Al-Shamsi, W. Alhazzani, A. Alhuraiji, E. A. Coomes, R. F. Chemaly, M. Almuhanna, R. A. Wolff, N. K. Ibrahim, M. L. Chua, S. J. Hotte *et al.*, "A practical approach to the management of cancer patients during the novel coronavirus disease 2019 (covid-19) pandemic: an international collaborative group," *The oncologist*, vol. 25, no. 6, 2020.

[24] R. Ohannessian, T. A. Duong, and A. Odone, "Global telemedicine implementation and integration within health systems to fight the covid-19 pandemic: a call to action," *JMIR public health and surveillance*, vol. 6, no. 2, 2020.

[25] R. R. Tamboli, V. Sandu, S. Nerasala, A. Richhariya, K. K. Vupparaboina, and S. Jana, "Novel hybrid teleophthalmology: Technological case for oculofacial surgery," in *IEEE Global Humanitarian Technology Conference*, 2019, pp. 1–8.

[26] A. R. Kavoor, K. Chakravarthy, and T. John, "Remote consultations in the era of covid-19 pandemic: Preliminary experience in a regional australian public acute mental health care setting," *Asian Journal of Psychiatry*, vol. 51, 2020.

[27] H. Tüzün and F. Özdinç, "The effects of 3d multi-user virtual environments on freshmen university students' conceptual and spatial learning and presence in departmental orientation," *Computers & Education*, vol. 94, pp. 228–240, 2016.

[28] "3D Light Field Head-up Display for Automotive ," https://ec.europa.eu/eipp/desktop/en/projects/project-11269.html, Accessed: August 2020.

[29] T. Balogh and P. T. Kovács, "Holovizio: The next generation of 3D oil & gas visualization," in *70th EAGE Conference and Exhibition-Workshops and Fieldtrips*, 2008.

[30] "Gartner Hype Cycle for Emerging Technologies, 2019," https://www.gartner.com/smarterwithgartner/5-trends-appear-on-the-gartner-hype-cycle-for-emerging-technologies-2019/, Accessed: June 2020.

[31] N. S. Holliman, N. A. Dodgson, G. E. Favalora, and L. Pockett, "Three-dimensional displays: a review and applications analysis," *IEEE Transactions on Broadcasting*, vol. 57, no. 2, pp. 362–371, 2011.

[32] Y. Takaki, "Super multi-view display with 128 viewpoints and viewpoint formation," in *Stereoscopic Displays and Applications XX*, vol. 7237, 2009, pp. 581–588.

[33] A. Cserkaszky, P. A. Kara, R. R. Tamboli, A. Barsi, M. G. Martini, and T. Balogh, "Light-field capture and display systems: limitations, challenges, and potentials," in *Novel Optical Systems Design and Optimization XXI*, vol. 10746, 2018, pp. 10–18.

[34] R. R. Hainich and O. Bimber, *Displays: fundamentals & applications.* CRC press, 2016.

[35] J. A. Castellano, *Handbook of display technology.* Elsevier, 2012.

[36] J. Chen, W. Cranton, and M. Fihn, *Handbook of visual display technology.* Springer, 2016.

[37] B. G. Blundell and M. Fihn, "Introduction to 3d displays." Springer, 2016.

[38] M. Martínez-Corral and B. Javidi, "Fundamentals of 3d imaging and displays: a tutorial on integral imaging, light-field, and plenoptic systems," *Advances in Optics and Photonics*, vol. 10, no. 3, pp. 512–566, 2018.

[39] P. T. Kovács, "Methods for Light Field Display Profiling and Scalable Super-Multiview Video Coding," Ph.D. dissertation, Tampere University of Technology, 2018.

[40] J. Hong, Y. Kim, H.-J. Choi, J. Hahn, J.-H. Park, H. Kim, S.-W. Min, N. Chen, and B. Lee, "Three-dimensional display technologies of recent interest: principles, status, and issues [invited]," *Appl. Opt*, vol. 50, no. 34, pp. 87–115, 2011.

[41] X. Xiao, B. Javidi, M. Martinez-Corral, and A. Stern, "Advances in three-dimensional integral imaging: sensing, display, and applications," *Applied optics*, vol. 52, no. 4, pp. 546–560, 2013.

[42] M. Yamaguchi, "Light-field and holographic three-dimensional displays," *JOSA A*, vol. 33, no. 12, pp. 2348–2364, 2016.

[43] B. Lee, "Three-dimensional displays, past and present," *Physics Today*, vol. 66, no. 4, pp. 36–41, 2013.

[44] J. Geng, "Three-dimensional display technologies," *Advances in optics and photonics*, vol. 5, no. 4, pp. 456–535, 2013.

[45] L. Yang, H. Dong, A. Alelaiwi, and A. El Saddik, "See in 3d: state of the art of 3d display technologies," *Multimedia Tools and Applications*, vol. 75, no. 24, pp. 17 121–17 155, 2016.

[46] B. Javidi and F. Okano, *Three-dimensional television, video, and display technologies*. Springer Science & Business Media, 2002.

[47] G. Wetzstein, "Why people should care about light-field displays," *Information Display*, vol. 31, no. 2, pp. 22–28, 2015.

[48] C. Wheatstone, "Contributions to the physiology of vision.–part the first. on some remarkable, and hitherto unobserved, phenomena of binocular vision," *Philosophical transactions of the Royal Society of London*, pp. 371–394, 1838.

[49] L. Hammond, "Stereoscopic motion picture," *Patent, US*, vol. 1, 1922.

[50] F. E. Ives, "Parallax stereogram and process of making same." Apr. 14 1903, uS Patent 725,567.

[51] G. Lippmann, "Epreuves reversibles. photographies integrals," *Comptes-Rendus Academie des Sciences*, vol. 146, pp. 446–451, 1908.

[52] C. Van Berkel, D. W. Parker, and A. R. Franklin, "Multiview 3d lcd," in *Stereoscopic Displays and Virtual Reality Systems III*, vol. 2653. International Society for Optics and Photonics, 1996, pp. 32–39.

[53] D. Gabor, "A new microscopic principle," 1948.

[54] E. N. Leith and J. Upatnieks, "Holographic imagery through diffusing media," *JOSA*, vol. 56, no. 4, pp. 523–523, 1966.

[55] P. St-Hilaire, S. A. Benton, M. E. Lucente, M. L. Jepsen, J. Kollin, H. Yoshikawa, and J. S. Underkoffler, "Electronic display system for computational holography," in *Practical Holography IV*, vol. 1212. International Society for Optics and Photonics, 1990, pp. 174–182.

[56] I. E. Sutherland, "The ultimate display," *Multimedia: From Wagner to virtual reality*, 1965.

[57] "Shambling corpse of 3D TV finally falls down dead," https://www.cnet.com/news/shambling-corpse-of-3d-tv-finally-falls-down-dead/, Accessed: June 2020.

[58] T. Balogh and P. A. Kara, "Vr versus lf: towards the limitation-free 3d," in *Digital Optical Technologies*, vol. 10335, 2017, pp. 223–231.

[59] M. Levoy and P. Hanrahan, "Light field rendering," in *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*. ACM, 1996, pp. 31–42.

[60] "The Stanford 3D Scanning Repository," http://graphics.stanford.edu/data/3Dscanrep/, accessed: May 2020.

[61] T. Georgiev, Z. Yu, A. Lumsdaine, and S. Goma, "Lytro camera technology: theory, algorithms, performance analysis," in *Multimedia Content and Mobile Devices*, vol. 8667. International Society for Optics and Photonics, 2013, p. 86671J.

[62] T. Fujii, K. Mori, K. Takeda, K. Mase, M. Tanimoto, and Y. Suenaga, "Multipoint measuring system for video and sound-100-camera and microphone system," in *International Conference on Multimedia and Expo*, 2006, pp. 437–440.

[63] T. Herfet, T. Lange, and H. P. Hariharan, "Enabling multiview-and light field-video for veridical visual experiences," in *4th International Conference on Computer and Communications*, 2018, pp. 1705–1709.

[64] A. W. Fitzgibbon, G. Cross, and A. Zisserman, "Automatic 3d model construction for turn-table sequences," in *European Workshop on 3D Structure from Multiple Images of Large-Scale Environments*. Springer, 1998, pp. 155–170.

[65] P. Moreels and P. Perona, "Evaluation of features detectors and descriptors based on 3d objects," *International Journal Of Computer Vision*, vol. 73, no. 3, pp. 263–284, 2007.

[66] A. Kasper, Z. Xue, and R. Dillmann, "The kit object models database: An object model database for object recognition, localization and manipulation in service robotics," *The International Journal of Robotics Research*, vol. 31, no. 8, pp. 927–934, 2012.

[67] A. Singh, J. Sha, K. S. Narayan, T. Achim, and P. Abbeel, "Bigbird: A large-scale 3d database of object instances," in *IEEE International Conference on Robotics and Automation*, 2014, pp. 509–516.

[68] B. S. Wilburn, M. Smulski, H.-H. K. Lee, and M. A. Horowitz, "Light field video camera," in *Media Processors 2002*, vol. 4674. International Society for Optics and Photonics, 2001, pp. 29–36.

[69] P. Debevec, "The light stages and their applications to photoreal digital actors," *SIGGRAPH Asia*, vol. 2, no. 4, 2012.

[70] R. Satava, "3-d vision technology applied to advanced minimally invasive surgery systems," *Surgical Endoscopy*, vol. 7, no. 5, pp. 429–431, 1993.

[71] G. Jones and M. Christal, "The future of virtual museums: on-line, immersive, 3d environments," *Created Realities Group*, vol. 4, pp. 1–12, 2002.

[72] Y. Shi, W. Xie, and G. Xu, "Smart remote classroom: Creating a revolutionary real-time interactive distance learning system," in *Advances in Web-Based Learning*. Springer, 2002, pp. 130–141.

[73] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2003.

[74] H. Aghajan and A. Cavallaro, *Multi-camera Networks: principles and applications*. Academic Press, 2009.

[75] R. Tsai, "A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf tv cameras and lenses," *IEEE Journal on Robotics and Automation*, vol. 3, no. 4, pp. 323–344, 1987.

[76] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 11, pp. 1330–1334, 2000.

[77] J. Heikkila and O. Silven, "A four-step camera calibration procedure with implicit image correction," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1997, pp. 1106–1112.

[78] G. Kurillo, Z. Li, and R. Bajcsy, "Wide-area external multi-camera calibration using vision graphs and virtual calibration object," in *Second ACM/IEEE International Conference on Distributed Smart Cameras*, 2008, pp. 1–9.

[79] T. Svoboda, D. Martinec, and T. Pajdla, "A convenient multicamera self-calibration for virtual environments," *Presence: Teleoperators & Virtual Environments*, vol. 14, no. 4, pp. 407–422, 2005.

[80] T. Kuo, Z. Ni, S. Sunderrajan, and B. Manjunath, "Calibrating a wide-area camera network with non-overlapping views using mobile devices," *ACM Transactions on Sensor Networks*, vol. 10, no. 2, p. 26, 2014.

[81] E. E. Hemayed, "A survey of camera self-calibration," in *IEEE Conference on Advanced Video and Signal Based Surveillance*, 2003, pp. 351–357.

[82] E. Langbehn, G. Bruder, and F. Steinicke, "Scale matters! analysis of dominant scale estimation in the presence of conflicting cues in multi-scale collaborative virtual environments," in *IEEE Symposium on 3D User Interfaces*, 2016, pp. 211–220.

[83] J. L. Lerma, S. Navarro, M. Cabrelles, and A. E. Seguí, "Camera calibration with baseline distance constraints," *The Photogrammetric Record*, vol. 25, no. 130, pp. 140–158, 2010.

[84] T. Balogh, T. Forgács, T. Agács, O. Balet, E. Bouvier, F. Bettio, E. Gobbetti, and G. Zanetti, "A scalable hardware and software system for the holographic display of interactive graphics applications." in *Eurographics (Short Presentations)*, 2005, pp. 109–112.

[85] "Streaming Model for Field-of-Light Displays (SMFoLD)," http://www.smfold.org/blog/, Accessed: June 2020.

[86] "European Network on Quality of Experience in Multimedia Systems and Services," http://www.qualinet.eu/, Accessed: June 2020.

[87] P. Le Callet, S. Möller, A. Perkis *et al.*, "Qualinet white paper on definitions of quality of experience," *European network on quality of experience in multimedia systems and services (COST Action IC 1003)*, vol. 3, no. 2012, 2012.

[88] K. Brunnström, S. A. Beker, K. De Moor, A. Dooms, S. Egger, M.-N. Garcia, T. Hossfeld, S. Jumisko-Pyykkö, C. Keimel, M.-C. Larabi *et al.*, "Qualinet White Paper on Definitions of Quality of Experience," 2013.

[89] M. Hamilton, M. Troke, C. Rumbolt, R. Lockyer, D. Benoit, and T. Butyn, "High-performance light field display simulator," 2020, uS Patent App. 16/156,421.

[90] M. Hamilton, C. Rumbolt, T. Butyn, D. Benoit, R. Lockyer, and M. Troke, "P-91: Light field display simulator for experience and quality evaluation," in *SID Symposium Digest of Technical Papers*, vol. 49, no. 1. Wiley Online Library, 2018, pp. 1523–1526.

[91] P. Carballeira, J. Gutiérrez, F. Morán, J. Cabrera, F. Jaureguizar, and N. García, "MultiView Perceptual Disparity Model for Super MultiView Video," *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 1, pp. 113–124, Feb 2017.

[92] K. K. Vupparaboina, K. Raghavan, and S. Jana, "Smart camera networks: an analytical framework for auto calibration without ambiguity," in *IEEE Recent Advances in Intelligent Computational Systems*, 2013, pp. 310–315.

[93] ——, "Euclidean auto calibration of camera networks: baseline constraint removes scale ambiguity," in *Twenty Second National Conference on Communication*, 2016, pp. 1–6.

[94] K. K. Vupparaboina, "Perspectives on imaging and image analytics in ophthalmology," Ph.D. dissertation, Indian Institute of Technology Hyderabad, 2017.

[95] R. R. Tamboli, K. K. Vupparaboina, M. S. Reddy, P. A. Kara, A. Cserkaszky, M. G. Martini, A. Richhariya, and S. Jana, "Towards euclidean auto-calibration of stereo camera arrays," in *Optical System Alignment, Tolerancing, and Verification XII*, vol. 10747, 2018, pp. 135–144.

[96] K. K. Vupparaboina, R. R. Tamboli, S. Manne, A. Richhariya, M. Naik, and S. Jana, "Oculofacial surgical planning: true-to-scale 3d feature quantification using multicamera network," in *IEEE Annual India Conference*, 2016, pp. 1–6.

[97] K. K. Vupparaboina, R. R. Tamboli, S. Manne, P. A. Kara, M. G. Martini, A. Barsi, A. Richhariya, and S. Jana, "Towards true-to-scale 3d reconstruction

of the human face using structured light projection and off-the-shelf cameras," in *International Conference on 3D Immersion*, 2018, pp. 1–7.

[98] R. R. Tamboli, B. Appina, S. Channappayya, and S. Jana, "Super-multiview content with high angular resolution: 3D quality assessment on horizontal-parallax lightfield display," *Signal Processing: Image Communication*, vol. 47, pp. 42–55, 2016.

[99] A. Cserkaszky, P. A. Kara, R. R. Tamboli, A. Barsi, M. G. Martini, L. Bokor, and T. Balogh, "Angularly continuous light-field format: Concept, implementation, and evaluation," *Journal of the Society for Information Display*, vol. 27, no. 7, pp. 442–461, 2019.

[100] R. Tamboli, K. K. Vupparaboina, J. Ready, S. Jana, and S. Channappayya, "A subjective evaluation of true 3d images," in *International Conference on 3D Imaging*, 2014, pp. 1–8.

[101] R. R. Tamboli, A. Cserkaszky, P. A. Kara, A. Barsi, and M. G. Martini, "Objective quality evaluation of an angularly-continuous light-field format," in *International Conference on 3D Immersion*, 2018, pp. 1–8.

[102] R. R. Tamboli, P. A. Kara, A. Cserkaszky, A. Barsi, M. G. Martini, B. Appina, S. S. Channappayya, and S. Jana, "3d objective quality assessment of light field video frames," in *3DTV-Conference: The True Vision-Capture, Transmission and Display of 3D Video*, 2018, pp. 1–4.

[103] R. R. Tamboli, P. A. Kara, N. Bisht, A. Barsi, M. G. Martini, and S. Jana, "Objective quality assessment of 2d synthesized views for light-field visualization," in *International Conference on 3D Immersion*, 2018, pp. 1–7.

[104] R. R. Tamboli, B. Appina, S. S. Channappayya, and S. Jana, "Achieving high angular resolution via view synthesis: quality assessment of 3d content on super multiview lightfield display," in *International Conference on 3D Immersion*, 2017, pp. 1–8.

[105] R. R. Tamboli, M. S. Reddy, P. A. Kara, M. G. Martini, S. S. Channappayya, and S. Jana, "A high-angular-resolution turntable data-set for experiments on light field visualization quality," in *Tenth International Conference on Quality of Multimedia Experience*, 2018, pp. 1–3.

[106] S. Darukumalli, P. A. Kara, A. Barsi, M. G. Martini, and T. Balogh, "Subjective quality assessment of zooming levels and image reconstructions based on region of interest for light field displays," in *International Conference on 3D Imaging*, 2016, pp. 1–6.

[107] R. R. Tamboli, B. Appina, P. A. Kara, M. G. Martini, S. S. Channappayya, and S. Jana, "Effect of primitive features of content on perceived quality of light field visualization," in *Tenth International Conference on Quality of Multimedia Experience*, 2018, pp. 1–3.

[108] R. R. Tamboli, P. A. Kara, A. Cserkaszky, A. Barsi, M. G. Martini, and S. Jana, "Canonical 3d object orientation for interactive light-field visualization," in *Applications of Digital Image Processing XLI*, vol. 10752, 2018, pp. 77–83.

[109] A. Dricot, J. Jung, M. Cagnazzo, B. Pesquet, F. Dufaux, P. T. Kovács, and V. K. Adhikarla, "Subjective evaluation of super multi-view compressed contents on high-end light-field 3d displays," *Signal Processing: Image Communication*, vol. 39, pp. 369–385, 2015.

[110] P. A. Kara, R. R. Tamboli, A. Cserkaszky, M. G. Martini, A. Barsi, and L. Bokor, "The viewing conditions of light-field video for subjective quality assessment," in *International Conference on 3D Immersion*, 2018, pp. 1–8.

[111] ——, "The perceived quality of light-field video services," in *Applications of Digital Image Processing XLI*, vol. 10752, 2018, pp. 138–147.

[112] P. A. Kara, R. R. Tamboli, O. Doronin, A. Cserkaszky, A. Barsi, Z. Nagy, M. G. Martini, and A. Simon, "The key performance indicators of projection-based light field visualization," *Journal of Information Display*, vol. 20, no. 2, pp. 81–93, 2019.

[113] P. A. Kara, R. R. Tamboli, A. Cserkaszky, A. Barsi, A. Simon, A. Kusz, L. Bokor, and M. G. Martini, "Objective and subjective assessment of binocular disparity for projection-based light field displays," in *International Conference on 3D Immersion*, 2019, pp. 1–8.

[114] A. Kondoz and T. Dagiuklas, *Novel 3D Media Technologies.* Springer, 2015.

[115] T. S. Huang, *Calibration and Orientation of Cameras in Computer Vision.* Springer, 2001.

[116] D. G. Lowe, "Object recognition from local scale-invariant features," in *Seventh IEEE International Conference on Computer Vision*, vol. 2, 1999, pp. 1150–1157.

[117] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (surf)," *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346–359, 2008.

[118] P. Sturm and B. Triggs, "A factorization based algorithm for multi-image projective structure and motion," in *European Conference on Computer Vision*. Springer, 1996, pp. 709–720.

[119] S. J. Maybank and O. D. Faugeras, "A theory of self-calibration of a moving camera," *International Journal of Computer Vision*, vol. 8, no. 2, pp. 123–151, 1992.

[120] M. Han and T. Kanade, "Creating 3D models with uncalibrated cameras," in *Fifth IEEE Workshop on Applications of Computer Vision*, 2000, pp. 178–185.

[121] R. Szeliski and S. B. Kang, "Shape ambiguities in structure from motion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 5, pp. 506–512, 1997.

[122] J. Knight, A. Zisserman, and I. Reid, "Linear auto-calibration for ground plane motion," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, 2003.

[123] H. Zhong and Y. Hung, "Self-calibration from one circular motion sequence and two images," *Pattern Recognition*, vol. 39, no. 9, pp. 1672–1678, 2006.

[124] Z. Zhao, Y. Liu, and Z. Zhang, "Camera calibration with three noncollinear points under special motions," *IEEE Transactions on Image Processing*, vol. 17, no. 12, pp. 2393–2402, 2008.

[125] E. Malis and R. Cipolla, "Camera self-calibration from unknown planar structures enforcing the multiview constraints between collineations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 9, pp. 1268–1272, 2002.

[126] G. Wang, Q. J. Wu, and W. Zhang, "Kruppa equation based camera calibration from homography induced by remote plane," *Pattern Recognition Letters*, vol. 29, no. 16, pp. 2137–2144, 2008.

[127] D. Devarajan, R. J. Radke, and H. Chung, "Distributed metric calibration of ad hoc camera networks," *ACM Transactions on Sensor Networks*, vol. 2, no. 3, pp. 380–403, 2006.

[128] M. I. Lourakis and A. A. Argyros, "Sba: A software package for generic sparse bundle adjustment," *ACM Transactions on Mathematical Software*, vol. 36, no. 1, pp. 1–30, 2009.

[129] Y. Furukawa and J. Ponce, "Accurate camera calibration from multi-view stereo and bundle adjustment," *International Journal of Computer Vision*, vol. 84, no. 3, pp. 257–268, 2009.

[130] G. Qian and R. Chellappa, "Bayesian self-calibration of a moving camera," *Computer Vision and Image Understanding*, vol. 95, no. 3, pp. 287–316, 2004.

[131] J. Civera, D. R. Bueno, A. J. Davison, and J. Montiel, "Camera self-calibration for sequential bayesian structure from motion," in *IEEE International Conference on Robotics and Automation*, 2009, pp. 403–408.

[132] A. Choudhary, G. Sharma, S. Chaudhury, and S. Banerjee, "Distributed calibration of pan-tilt camera network using multi-layered belief propagation," in *IEEE Computer Society Conference on computer Vision and Pattern Recognition Workshops*, 2010, pp. 33–40.

[133] "Holovizio 721RC light field display," http://www.archive.holografika.com/Products/HoloVizio-721RC.html, accessed: May 2020.

[134] J. Chen, W. Mustafa, A. Siddig, and W. Kulesza, "Applying dithering to improve depth measurement using a sensor-shifted stereo camera," *Metrology and Measurement Systems*, vol. 17, no. 3, pp. 335–347, 2010.

[135] C. Fehn, "Depth-image-based rendering (dibr), compression, and transmission for a new approach on 3d-tv," in *Stereoscopic Displays and Virtual Reality Systems XI*, vol. 5291, International Society for Optics and Photonics. SPIE, 2004, pp. 93–104.

[136] C.-H. J. Tzou, N. M. Artner, I. Pona, A. Hold, E. Placheta, W. G. Kropatsch, and M. Frey, "Comparison of three-dimensional surface-imaging systems," *Journal of Plastic, Reconstructive & Aesthetic Surgery*, vol. 67, no. 4, pp. 489–497, 2014.

[137] "3D head model," https://grabcad.com/library/face-13, Accessed: June 2020.

[138] K. Deb, *Multi-objective optimization using evolutionary algorithms.* John Wiley & Sons, 2001, vol. 16.

[139] MATLAB, *version 8.2.0.701 (R2013b).* The MathWorks Inc., 2013.

[140] J.-Y. Son, H. Lee, B.-R. Lee, and K.-H. Lee, "Holographic and light-field imaging as future 3-d displays," *Proceedings of the IEEE*, vol. 105, no. 5, pp. 789–804, 2017.

[141] E. H. Adelson and J. R. Bergen, "The plenoptic function and the elements of early vision," in *Computational Models of Visual Processing.* Cambridge, MA: MIT Press, 1991, pp. 3–20.

[142] "HoloVizio C80 light field cinema system," http://holografika.com/c80-glasses-free-3d-cinema/, Accessed: May 2020.

[143] J. Pallanch, *3-D imaging technologies in facial plastic surgery, an issue of facial plastic surgery clinics-e-book.* Elsevier Health Sciences, 2012, vol. 19, no. 4.

[144] G. Lekakis, P. Claes, G. S. Hamilton III, and P. Hellings, "Three-dimensional surface imaging and the continuous evolution of preoperative and postoperative assessment in rhinoplasty," *Facial Plastic Surgery*, vol. 32, no. 01, pp. 088–094, 2016.

[145] J. L. Henderson, W. F. Larrabee, and B. D. Krieger, "Photographic standards for facial plastic surgery," *Archives of Facial Plastic Surgery*, vol. 7, no. 5, pp. 331–333, 2005.

[146] "acA1300-30gc - Basler ace," https://www.baslerweb.com/en/products/cameras/area-scan-cameras/ace/aca1300-30gc/, Accessed: Jun. 2019.

[147] "Goyo Optical GMDN24012C Lens," http://www.goyooptical.com/products/cctv/manual/GMDN24012C.pdf, Accessed: Jun. 2019.

[148] "Dell M110 Ultra-Mobile Projector," https://www.dell.com/is/business/p/dell-m110/pd, Accessed: Jun. 2019.

[149] "uPrint SE 3D printer, Stratasys," http://www.stratasys.com/3d-printers/idea-series/uprint-se, Accessed: December 2016.

[150] R. C. Gonzalez, R. E. Woods *et al.*, "Digital image processing," 2002.

[151] "CloudCompare," https://www.danielgm.net/cc/, Accessed: July 2020.

[152] "Blender," https://www.blender.org/, Accessed: July 2020.

[153] "The LVPEI Eye Care Network," http://www.lvpei.org/assets/images/services/icare/ICARE_Training_Brochure.pdf, Accessed: Jun. 2019.

[154] B. S. Chandra, "Towards next-generation cardiac care in resource-constrained and high-risk scenarios," Ph.D. dissertation, Indian Institute of Technology Hyderabad, 2019.

[155] World Health Organization, *Telemedicine: opportunities and developments in member states. Report on the second global survey on eHealth.*, 2010.

[156] M. Chokshi, B. Patil, R. Khanna, S. B. Neogi, J. Sharma, V. Paul, and S. Zodpey, "Health systems in india," *Journal of Perinatology*, vol. 36, no. 3, pp. 9–12, 2016.

[157] J. Chuang, C. Barnes, and B. J. Wong, "Overview of facial plastic surgery and current developments," *The Surgery Journal*, vol. 2, no. 1, pp. e17–e28, 2016.

[158] J. B. A. Martel, U. M. Anders, and V. Kravchuk, "Comparative study of teleophthalmology devices: smartphone adapted ophthalmoscope, robotic ophthalmoscope, and traditional fundus camera-the recent advancements in telemedicine," *New Front Ophthalmol*, vol. 1, no. 1, pp. 2–5, 2015.

[159] R. P. Abhinav, K. Selvarasu, G. U. Maheswari, A. A. Taltia *et al.*, "The patterns and etiology of maxillofacial trauma in South India," *Annals of Maxillofacial Surgery*, vol. 9, no. 1, pp. 114–117, 2019.

[160] I. P. S. E. Farias, Í. D. Bernardino, L. Nóbrega, R. G. Grempel, and S. D'avila, "Maxillofacial trauma, etiology and profile of patients: an exploratory study," *Acta Ortopedica Brasileira*, vol. 25, no. 6, pp. 258–261, 2017.

[161] A. Shah, V. Nautiyal, A. Gupta, and V. Ramola, "Trends of maxillofacial fractures in the Garhwal Himalayas at Government Medical College, Srinagar, Uttarakhand," *National Journal of Maxillofacial Surgery*, vol. 7, no. 1, pp. 80–85, 2016.

[162] *Performance appraisal and district economic scenario, 2017-18.* Directorate of Economics and Statistics, Planning Department, Government of Andhra Pradesh, 2018.

[163] *Economic survey of Karnataka 2017-18.* Department of Planning, Programme Monitoring and Statistics, Government of Karnataka, 2018.

[164] *Odisha economic survey 2014-15.* Planning and Coordination Department, Directorate of Economics and Statistics, Government of Odisha, 2015.

[165] *Socio economic outlook 2018.* Planning Department, Government of Telangana, 2018.

[166] S. Malik, G. Singh, G. Kaur, S. Yadav, and H. C. Mittal, "Orofacial trauma in rural india: a clinical study," *Chinese Journal of Traumatology*, vol. 20, no. 4, pp. 216–221, 2017.

[167] C. Boehnen and P. Flynn, "Accuracy of 3d scanning technologies in a face scanning scenario," in *Fifth International Conference on 3-D Digital Imaging and Modeling*, 2005, pp. 310–317.

[168] "DES-1016D 16-Port Fast Ethernet Unmanaged Desktop Switch," https://eu.dlink.com/uk/en/products/des-1016d-16-port-10-100mbps-desktop-switch, Accessed: Jun. 2019.

[169] A. C. Bovik, *The essential guide to image processing.* Academic Press, 2009.

[170] G. Yu and J.-M. Morel, "Asift: An algorithm for fully affine invariant comparison," *Image Processing On Line*, vol. 1, pp. 11–38, 2011.

[171] M. Kazhdan and H. Hoppe, "Screened poisson surface reconstruction," *ACM Transactions on Graphics*, vol. 32, no. 3, p. 29, 2013.

[172] P. Cignoni, M. Callieri, M. Corsini, M. Dellepiane, F. Ganovelli, and G. Ranzuglia, "MeshLab: an Open-Source Mesh Processing Tool," in *Eurographics Italian Chapter Conference*, 2008, pp. 129–136.

[173] B. E. Jackson, D. J. Evangelista, D. D. Ray, and T. L. Hedrick, "3d for the people: multi-camera motion capture in the field with consumer-grade cameras and open source software," *Biology Open*, vol. 5, no. 9, pp. 1334–1342, 2016.

[174] "Mobile Eye Surgical Unit," https://www.hticiitm.org/mesu, Accessed: June 2020.

[175] R. P. Sangameswaran, G. K. Verma, N. Raghavan, J. Joseph, and M. Sivaprakasam, "Cataract surgery in mobile eye surgical unit: Safe and viable alternative," *Indian Journal of Ophthalmology*, vol. 64, no. 11, pp. 835–839, 2016.

[176] L. Huang, Q. Zhang, and A. Asundi, "Flexible camera calibration using not-measured imperfect target," *Applied Optics*, vol. 52, no. 25, pp. 6278–6286, 2013.

[177] A. S. Jackson, A. Bulat, V. Argyriou, and G. Tzimiropoulos, "Large pose 3d face reconstruction from a single image via direct volumetric cnn regression," in *IEEE International Conference on Computer Vision*, 2017, pp. 1031–1039.

[178] A. Brunton, A. Salazar, T. Bolkart, and S. Wuhrer, "Review of statistical shape spaces for 3d data with comparative analysis for human faces," *Computer Vision and Image Understanding*, vol. 128, pp. 1–17, 2014.

[179] "Bellus3d: High-quality 3d face scanning," https://www.bellus3d.com/, Accessed: Jun. 2019.

[180] A. Nouri, C. Charrier, and O. Lézoray, "Full-reference saliency-based 3d mesh quality assessment index," in *International Conference on Image Processing*, 2016, pp. 1007–1011.

[181] W. Chen, J. Fournier, M. Barkowsky, and P. Le Callet, "New requirements of subjective video quality assessment methodologies for 3dtv," in *Video Processing and Quality Metrics*, Jan. 2010.

[182] R. Shacked and D. Lischinski, "Automatic lighting design using a perceptual quality metric," in *Computer Graphics Forum*, vol. 20, no. 3. Wiley Online Library, 2001, pp. 215–227.

[183] K. Seshadrinathan, R. Soundararajan, A. C. Bovik, and L. K. Cormack, "Study of subjective and objective quality assessment of video," *IEEE Transactions on Image Processing*, vol. 19, no. 6, pp. 1427–1441, 2010.

[184] S. Chikkerur, V. Sundaram, M. Reisslein, and L. J. Karam, "Objective video quality assessment methods: A classification, review, and performance comparison," *IEEE Transactions on Broadcasting*, vol. 57, no. 2, pp. 165–182, 2011.

[185] C. Deng, L. Ma, W. Lin, K. N. Ngan *et al.*, *Visual signal quality assessment.* Springer, 2016.

[186] A. Benoit, P. Le Callet, P. Campisi, and R. Cousseau, "Quality assessment of stereoscopic images," *EURASIP Journal On Image And Video Processing*, vol. 2008, pp. 1–13, 2009.

[187] A. K. Moorthy, C.-C. Su, A. Mittal, and A. C. Bovik, "Subjective evaluation of stereoscopic image quality," *Signal Processing: Image Communication*, vol. 28, no. 8, pp. 870–883, 2013.

[188] G. Wu, B. Masia, A. Jarabo, Y. Zhang, L. Wang, Q. Dai, T. Chai, and Y. Liu, "Light field image processing: An overview," *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 7, pp. 926–954, 2017.

[189] M. Levoy, "Light fields and computational imaging," *Computer*, vol. 39, no. 8, pp. 46–55, 2006.

[190] A. Mustafa, M. Volino, J.-y. Guillemaut, and A. Hilton, "4D Temporally Coherent Light-field Video," *arXiv preprint arXiv:1804.11276*, 2018.

[191] F. Battisti, E. Bosc, M. Carli, P. Le Callet, and S. Perugia, "Objective image quality assessment of 3d synthesized views," *Signal Processing: Image Communication*, vol. 30, pp. 78–88, 2015.

[192] X. Liu, Y. Zhang, S. Hu, S. Kwong, C.-C. J. Kuo, and Q. Peng, "Subjective and objective video quality assessment of 3d synthesized views with texture/depth compression distortion," *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 4847–4861, 2015.

[193] E. Ekmekcioglu, S. Worrall, D. De Silva, A. Fernando, and A. M. Kondoz, "Depth based perceptual quality assessment for synthesised camera viewpoints," in *International Conference on User Centric Media.* Springer, 2010, pp. 76–83.

[194] A. Chappuis, M. Rerabek, P. Hanhart, and T. Ebrahimi, "Subjective evaluation of an active crosstalk reduction system for mobile autostereoscopic displays," in *Stereoscopic Displays and Applications XXV*, vol. 9011. International Society for Optics and Photonics, 2014, pp. 241–253.

[195] A. Tikanmaki, A. Gotchev, A. Smolic, and K. Miller, "Quality assessment of 3d video in rate allocation experiments," in *IEEE International Symposium on Consumer Electronics*, 2008, pp. 1–4.

[196] ITU-R Rec., "BT.2020-2: Parameter values for ultra-high definition television systems for production and international programme exchange," 2015.

[197] I. Viola, M. Rerabek, and T. Ebrahimi, "Comparison and evaluation of light field image coding approaches," *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 7, pp. 1092–1106, 2017.

[198] V. K. Adhikarla, M. Vinkler, D. Sumin, R. K. Mantiuk, K. Myszkowski, H.-P. Seidel, and P. Didyk, "Towards a quality metric for dense light fields," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 3720–3729.

[199] ITU-R Rec., "BT.500-13: Methodology for the subjective assessment of the quality of television pictures," 2012.

[200] Y. Liu, L. K. Cormack, and A. C. Bovik, "Statistical modeling of 3-d natural scenes with application to bayesian stereopsis," *IEEE Transactions on Image Processing*, vol. 20, no. 9, pp. 2515–2530, 2011.

[201] B. Potetz and T. S. Lee, "Statistical correlations between two-dimensional images and three-dimensional structures in natural scenes," *Journal of Optical Society of America A*, vol. 20, no. 7, pp. 1292–1303, 2003.

[202] "MPEG-I Visual Test Material Summary," https://mpeg.chiariglione.org/standards/exploration/immersive-video/mpeg-i-visual-test-material-summary, Accessed: May 2020.

[203] A. Borji, S. Izadi, and L. Itti, "ilab-20m: A large-scale controlled object dataset to investigate deep learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 2221–2230.

[204] E. Bosc, "Compression of multi-view-plus-depth (mvd) data: from perceived quality analysis to mvd coding tools designing," Ph.D. dissertation, INSA de Rennes, 2012.

[205] A. M. Van Dijk, J.-B. Martens, and A. B. Watson, "Quality asessment of coded images using numerical category scaling," in *Advanced Image and Video Communications and Storage Technologies*, vol. 2451, 1995, pp. 90–101.

[206] M. J. Wainwright and E. P. Simoncelli, "Scale mixtures of gaussians and the statistics of natural images," in *Advances in neural information processing systems*, 2000, pp. 855–861.

[207] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Transactions on Image Processing*, vol. 15, no. 2, pp. 430–444, 2006.

[208] D. J. Field, "Relations between the statistics of natural images and the response properties of cortical cells," *Journal of the Optical Society of America*, vol. 4, no. 12, pp. 2379–2394, 1987.

[209] E. P. Simoncelli and W. T. Freeman, "The steerable pyramid: A flexible architecture for multi-scale derivative computation," in *International Conference on Image Processing*, vol. 3, 1995, pp. 444–447.

[210] A. Karasaridis and E. Simoncelli, "A filter design technique for steerable pyramid image transforms," in *International Conference on Acoustics, Speech, and Signal Processing Conference Proceedings*, vol. 4, 1996, pp. 2387–2390.

[211] D. J. Field, "What is the goal of sensory coding?" *Neural Computation*, vol. 6, no. 4, pp. 559–601, 1994.

[212] S. K. Md, B. Appina, and S. S. Channappayya, "Full-reference stereo image quality assessment using natural stereo scene statistics," *IEEE Signal Processing Letters*, vol. 22, no. 11, pp. 1985–1989, 2015.

[213] M. M. Deza and E. Deza, *Encyclopedia of distances.* Springer, 2009.

[214] S.-H. Cha, "Comprehensive survey on distance/similarity measures between probability density functions," *International Journal of Mathematical models and Methods in Applied Sciences*, vol. 1, no. 4, pp. 300–307, 2007.

[215] S.-S. Choi, S.-H. Cha, and C. C. Tappert, "A survey of binary similarity and distance measures," *Journal of Systemics, Cybernetics and Informatics*, vol. 8, no. 1, pp. 43–48, 2010.

[216] J. J. Koenderink and A. J. van Doorn, "Facts on optic flow," *Biological cybernetics*, vol. 56, no. 4, pp. 247–254, 1987.

[217] A. Wedel, T. Brox, T. Vaudrey, C. Rabe, U. Franke, and D. Cremers, "Stereoscopic scene flow computation for 3d motion understanding," *International Journal of Computer Vision*, vol. 95, no. 1, pp. 29–51, 2011.

[218] M. J. Black and P. Anandan, "A framework for the robust estimation of optical flow," in *4th International Conference on Computer Vision*, 1993, pp. 231–236.

[219] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[220] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers*, vol. 2, 2003, pp. 1398–1402.

[221] H. R. Sheikh, A. C. Bovik, and G. De Veciana, "An information fidelity criterion for image quality assessment using natural scene statistics," *IEEE Transactions on Image Processing*, vol. 14, no. 12, pp. 2117–2128, 2005.

[222] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "Fsim: A feature similarity index for image quality assessment," *IEEE Transactions on Image Processing*, vol. 20, no. 8, pp. 2378–2386, 2011.

[223] A. K. Moorthy and A. C. Bovik, "Visual importance pooling for image quality assessment," *Journal of Selected Topics in Signal Processing*, vol. 3, no. 2, pp. 193–201, 2009.

[224] ITU-T, "Objective perceptual assessment of video quality: Full reference television," 2004.

[225] "JCB 435b 3D Model," https://www.3dcadbrowser.com/3d-model/jcb-435b, Accessed: May. 2020.

[226] "3ds Max," https://www.autodesk.com/education/free-software/3ds-max, Accessed: May. 2020.

[227] P. A. Kara, A. Cserkaszky, A. Barsi, T. Papp, M. G. Martini, and L. Bokor, "The interdependence of spatial and angular resolution in the quality of experience of light field visualization," in *International Conference on 3D Immersion*, 2017, pp. 1–8.

[228] M. Domański, T. Grajek, C. Conti, C. J. Debono, S. M. de Faria, P. Kovacs, L. F. Lucas, P. Nunes, C. Perra, N. M. Rodrigues *et al.*, "Emerging imaging technologies: trends and challenges," in *3D Visual Content Creation, Coding and Delivery.* Springer, 2019, pp. 5–39.

[229] P. A. Kara, A. Cserkaszky, M. G. Martini, A. Barsi, L. Bokor, and T. Balogh, "Evaluation of the concept of dynamic adaptive streaming of light field video," *IEEE Transactions on Broadcasting*, vol. 64, no. 2, pp. 407–421, 2018.

[230] A. Cserkaszky, A. Barsi, P. A. Kara, and M. G. Martini, "Towards display-independent light-field formats," in *International Conference on 3D Immersion*, 2017, pp. 1–7.

[231] M. Damghanian, P. Kerbiriou, V. Drazic, D. Doyen, and L. Blondé, "Camera-agnostic format and processing for light-field data," in *IEEE International Conference on Multimedia Expo Workshops*, 2017, pp. 7–12.

[232] T. Ebrahimi, S. Foessel, F. Pereira, and P. Schelkens, "Jpeg pleno: Toward an efficient representation of visual reality," *IEEE Multimedia*, vol. 23, no. 4, pp. 14–20, 2016.

[233] I. Viola, M. Rerabek, T. Bruylants, P. Schelkens, F. Pereira, and T. Ebrahimi, "Objective and subjective evaluation of light field image compression algorithms," in *Picture Coding Symposium*, 2016, pp. 1–5.

[234] W. Ahmad, M. Sjöström, and R. Olsson, "Compression scheme for sparsely sampled light field data based on pseudo multi-view sequences," in *Optics, Photonics, and Digital Technologies for Imaging Applications V*, vol. 10679, 2018, pp. 147–154.

[235] B. Guo, J. Wen, and Y. Han, "Two-pass light field image compression for spatial quality and angular consistency," *arXiv:1808.00630*, 2018.

[236] A. Cserkaszky, P. A. Kara, A. Barsi, and M. G. Martini, "Expert evaluation of a novel light-field visualization format," in *3DTV-Conference: The True Vision-Capture, Transmission and Display of 3D Video*, 2018, pp. 1–4.

[237] P. A. Kara, A. Cserkaszky, S. Darukumalli, A. Barsi, and M. G. Martini, "On the edge of the seat: Reduced angular resolution of a light field cinema with fixed observer positions," in *Ninth International Conference on Quality of Multimedia Experience*, 2017, pp. 1–6.

[238] "George W. Hart's Rapid Prototyping Web Page," www.georgehart.com/rp/rp.html, Accessed: May 2020.

[239] "Jotero.com 3D-Scan and 3D Measurement," forum.jotero.com/viewtopic.php?t=3, Accessed: May 2020.

[240] "JPEG," https://jpeg.org/jpeg/index.html, Accessed: May 2020.

[241] "JPEG2000," https://jpeg.org/jpeg2000/index.html, Accessed: May 2020.

[242] "WebP: A new image format for the Web," https://developers.google.com/speed/webp/, Accessed: May 2020.

[243] A. Cserkaszky, A. Barsi, P. A. Kara, and M. G. Martini, "To interpolate or not to interpolate: Subjective assessment of interpolation performance on a light field display," in *IEEE International Conference on Multimedia & Expo Workshops*, 2017, pp. 55–60.

[244] F. Porikli, A. Bovik, C. Plack, G. AlRegib, J. Farrell, P. Le Callet, Q. Huynh-Thu, S. Möller, and S. Winkler, "Multimedia quality assessment [dsp forum]," *IEEE Signal Processing Magazine*, vol. 28, no. 6, pp. 164–177, 2011.

[245] Z. Pylyshyn, "Is vision continuous with cognition?: The case for cognitive impenetrability of visual perception," *Behavioral and Brain Sciences*, vol. 22, no. 3, pp. 341–365, 1999.

[246] U. Engelke and P. Le Callet, "Perceived interest and overt visual attention in natural images," *Signal Processing: Image Communication*, vol. 39, pp. 386–404, 2015.

[247] ITU-T Rec., "BT.2021: Subjective methods for the assessment of stereoscopic 3DTV systems," 2015.

[248] "Digital Reality and Multimedia (DReaM) Lab, IIT Hyderabad," http://www.iith.ac.in/DReaM/3dqa.html, Accessed: May 2018.

[249] V. K. Adhikarla, P. T. Kovács, A. Barsi, T. Balogh, and P. Szolgay, "View synthesis for lightfield displays using region based non-linear image warping," in *International Conference on 3D Imaging*, 2012, pp. 1–6.

[250] G. Lafruit, M. Domański, K. Wegner, T. Grajek, T. Senoh, J. Jung, P. T. Kovács, P. Goorts, L. Jorissen, A. Munteanu *et al.*, "New visual coding exploration in mpeg: Super-multiview and free navigation in free viewpoint tv," *Electronic Imaging*, vol. 2016, no. 5, pp. 1–9, 2016.

[251] Y. Zhou, L. Li, K. Gu, Y. Fang, and W. Lin, "Quality assessment of 3d synthesized images via disoccluded region discovery," in *International Conference on Image Processing*, 2016, pp. 1012–1016.

[252] M. Sharma, S. Chaudhury, and B. Lall, "A novel hybrid kinect-variety-based high-quality multiview rendering scheme for glass-free 3d displays," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 27, no. 10, pp. 2098–2117, 2016.

[253] K. Wegner, O. Stankiewicz, M. Tanimoto, and M. Domanski, "Enhanced view synthesis reference software (vsrs) for free-viewpoint television," *ISO/IEC JTC1/SC29/WG11 MPEG2013 M*, vol. 31520, 2013.

[254] "Porsche 911 GT2 3D Model," https://free3d.com/3d-model/porsche-911-gt-43465.html, Accessed: May. 2020.

[255] "Early Biplane 3D Model," http://www.cadnav.com/3d-models/model-37147.html, Accessed: May. 2020.

[256] S. Darukumalli, P. A. Kara, A. Barsi, and M. G. Martini, "How to rate the new glasses-free experience: Subjective quality assessment methodologies for experiments on light field displays," *International Young Researcher Summit on Quality of Experience in Emerging Multimedia Services*, pp. 1–4, 2017.

[257] T. M. Cover, *Elements of information theory.* John Wiley & Sons, 1999.

[258] N. Bouhlel and A. Dziri, "Kullback–leibler divergence between multivariate generalized gaussian distributions," *IEEE Signal Processing Letters*, vol. 26, no. 7, pp. 1021–1025, July 2019.

[259] P. A. Kara, M. G. Martini, P. T. Kovács, S. Imre, A. Barsi, K. Lackner, and T. Balogh, "Perceived quality of angular resolution for light field displays and the validy of subjective assessment," in *International Conference on 3D Imaging*, 2016, pp. 1–7.

[260] M. Domański, O. Stankiewicz, K. Wegner, and T. Grajek, "Immersive visual media—mpeg-i: 360 video, virtual navigation and beyond," in *International Conference on Systems, Signals and Image Processing*, 2017, pp. 1–9.

[261] "Nagoya university multi-view sequences download list," http://www.fujii.nuee.nagoya-u.ac.jp/multiview-data/, Accessed: Feb. 2018.

[262] P. A. Kara, P. T. Kovacs, S. Vagharshakyan, M. G. Martini, A. Barsi, T. Balogh, A. Chuchvara, and A. Chehaibi, "The effect of light field reconstruction and angular resolution reduction on the quality of experience," in *12th International Conference on Signal-Image Technology & Internet-Based Systems*, 2016, pp. 781–786.

[263] "PhotoCapture 360, Ortery Inc." https://www.ortery.com/photography-equipment/photocapture-series/photocapture-360/, Accessed: Mar. 2018.

[264] "Nikon D5300," http://imaging.nikon.com/lineup/dslr/d5300/, Accessed: Jun. 2019.

[265] "AF-S DX Micro NIKKOR 40mm f2.8G," https://www.nikon.co.in/en_IN/product/nikkor-lenses/f-mount/dx-format/special-purpose/af-s-dx-micro-nikkor-40mm-f-2-8g, Accessed: Jun. 2019.

[266] "DoF Simulator," https://dofsimulator.net/en/, Accessed: Jun. 2019.

[267] "Neewer Pro 4-Way Macro Focusing Focus Rail Slider," https://www.amazon.in/Neewer-Focusing-Close-up-Shooting-Digital/dp/B009SJ7UWU, Accessed: March 2018.

[268] K. Wegner and O. Stankiewicz, "Depth estimation using modified cost function for occlusion handling," *Signal, Image and Video Processing*, vol. 13, no. 8, pp. 1539–1547, 2019.

[269] "Camera Calibration Toolbox for Matlab," http://www.vision.caltech.edu/bouguetj/calib_doc/, Accessed: March 2018.

[270] Z. Zhang, "Flexible camera calibration by viewing a plane from unknown orientations," in *IEEE International Conference on Computer Vision*, vol. 1, 1999, pp. 666–673.

[271] M. J. Tarr and D. J. Kriegman, "What defines a view?" *Vision Research*, vol. 41, no. 15, pp. 1981–2004, 2001.

[272] H. Dutagaci, C. P. Cheung, and A. Godil, "A benchmark for best view selection of 3d objects," in *Proceedings of the ACM Workshop on 3D Object Retrieval*, 2010, pp. 45–50.

[273] L. M. Wong, C. Dumont, and M. A. Abidi, "Next best view system in a 3D object modeling task," in *IEEE International Symposium on Computational Intelligence in Robotics and Automation*, 1999, pp. 306–311.

[274] X. Bonaventura, M. Feixas, M. Sbert, L. Chuang, and C. Wallraven, "A survey of viewpoint selection methods for polygonal models," *Entropy*, vol. 20, no. 5, 2018.

[275] R. Niimi and K. Yokosawa, "Three-quarter views are subjectively good because object orientation is uncertain," *Psychonomic Bulletin & Review*, vol. 16, no. 2, pp. 289–294.

[276] A. Secord, J. Lu, A. Finkelstein, M. Singh, and A. Nealen, "Perceptual models of viewpoint preference," *ACM Transactions on Graphics*, vol. 30, no. 5, pp. 1–12, 2011.

[277] M. Carrasco, "Visual attention: The past 25 years," *Vision Research*, vol. 51, no. 13, pp. 1484–1525, 2011.

[278] M. Mortara and M. Spagnuolo, "Semantics-driven best view of 3D shapes," *Computers & Graphics*, vol. 33, no. 3, pp. 280–290, 2009.

[279] V. Blanz, M. J. Tarr, and H. H. Bülthoff, "What object attributes determine canonical views?" *Perception*, vol. 28, no. 5, pp. 575–599, 1999.

[280] N. A. Massios and R. B. Fisher, "A best next view selection algorithm incorporating a quality criterion," in *Proceedings of the Ninth British Machine Vision Conference*, 1998, pp. 780–789.

[281] G. Peters, *A View Based Approach to Three Dimensional Object Perception.* Shaker, 2002.

[282] B. Kwon, J. Huh, K. Lee, and S. Lee, "Optimal camera point selection toward the most preferable view of 3-d human pose," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2020.

[283] D. Giorgi, S. Biasotti, and L. Paraboschi, "Shape retrieval contest 2007: Watertight models track," *SHREC competition*, vol. 8, no. 7, 2007.

[284] "Logitech F310 Gamepad," https://www.logitechg.com/en-us/product/f310-gamepad, Accessed: May. 2020.

[285] V. K. Adhikarla, "Rendering and interaction on projection-based light field displays," Ph.D. dissertation, Pázmány Péter Katolikus Egyetem, 2015.

[286] S.-h. Kim, Y.-W. Tai, J.-Y. Lee, J. Park, and I. S. Kweon, "Category-specific salient view selection via deep convolutional neural networks," in *Computer Graphics Forum*, vol. 36, no. 8. Wiley Online Library, 2017, pp. 313–328.

[287] G. Medioni and S. B. Kang, *Emerging topics in computer vision.* Prentice Hall PTR, 2004.

[288] Y. Ma, S. Soatto, J. Kosecka, and S. S. Sastry, *An invitation to 3-d vision: from images to geometric models.* Springer Science & Business Media, 2012, vol. 26.

[289] P. Schelkens, T. Ebrahimi, A. Gilles, P. Gioia, K.-J. Oh, F. Pereira, C. Perra, and A. M. Pinheiro, "Jpeg pleno: Providing representation interoperability for holographic applications and devices," *ETRI Journal*, vol. 41, no. 1, pp. 93–108, 2019.

[290] IEEE-SA P3333.1.4_WG, "Standard for the Quality Assessment of Light Field Imaging," https://standards.ieee.org/project/3333_1_4.html, Accessed: July 2020.

[291] W. Hong and L. Yu, "Subjective assessment methodology for super multiview content," in *3DTV-Conference: The True Vision-Capture, Transmission and Display of 3D Video*, 2016, pp. 1–4.