University of Wollongong
Research Online

University of Wollongong Thesis Collection

University of Wollongong Thesis Collections

2014

Effective sampling density for quality assessment and optimization of light field rendering and acquisition

Hooman Shidanshidi
University of Wollongong, hooman@uow.edu.au

Recommended Citation

Shidanshidi, Hooman, Effective sampling density for quality assessment and optimization of light field rendering and acquisition, Doctor of Philosophy thesis, School of Electrical, Computer and Telecommunications Engineering, University of Wollongong, 2014. http://ro.uow.edu.au/theses/4285

Research Online is the open access institutional repository for the University of Wollongong. For further information contact the UOW Library: research-pubs@uow.edu.au

UNIVERSITY OF WOLLONGONG

COPYRIGHT WARNING

You may print or download ONE copy of this document for the purpose of your own research or study. The University does not authorise you to copy, communicate or otherwise make available electronically to any other person any copyright material contained on this site. You are reminded of the following:

Copyright owners are entitled to take legal action against persons who infringe their copyright. A reproduction of material that is protected by copyright may be a copyright infringement. A court may impose penalties and award damages in relation to offences and infringements relating to copyright material. Higher penalties may apply, and higher damages may be awarded, for offences and infringements involving the conversion of material into digital or electronic form.

Effective Sampling Density for Quality Assessment and Optimization of Light Field Rendering and Acquisition

Hooman Shidanshidi

PhD in Computer Engineering

A thesis submitted in fulfillment of the requirements for the award of the Degree

Doctor of Philosophy

School of Electrical, Computer and Telecommunications Engineering

ICT Research Institute

Faculty of Engineering and Information Sciences

University of Wollongong

Australia

2014

Supervisors: Professor Farzad Safaei, Associate Professor Wanqing Li

“Knowledge is as wings to man's life, and a ladder for his ascent. Its acquisition is incumbent upon everyone. The knowledge of such sciences, however, should be acquired as can profit the peoples of the earth, and not those which begin with words and end with words. Great indeed is the claim of scientists and craftsmen on the peoples of the world.”

Baha'u'llah

To my Mum, Dad, Sister, and all Bahá'ís persecuted in Iran

CERTIFICATION

I, Hooman Shidanshidi, declare that this thesis, submitted in fulfillment of the requirements for the award of Doctor of Philosophy, in the School of Electrical, Computer and Telecommunications Engineering, Faculty of Engineering and Information Sciences, University of Wollongong, is wholly my own work unless otherwise acknowledged. The document has not previously been submitted for qualifications at any academic institution.

Hooman Shidanshidi

March 2014

Acknowledgments

This thesis could not have been pursued without the undeviating support from ICT Research Institute - University of Wollongong, SMART Services CRC, Bahá'í Institute for Higher Education (BIHE), and the sacrifices made by my family and friends.

First and foremost, I would like to thank my supervisor, Professor Farzad Safaei and my co-supervisor, Associate Professor Wanqing Li for their continued support and guidance from the very early stages of this research. They gave me the opportunity to learn from their wide knowledge and experiences to become an experienced researcher. I am indebted to my supervisors more they are aware. Farzad's vast knowledge, patience, and high level of professionalism and Wanqing's truly scientist intuition and unflinching encouragement made this research possible. They both kindly supervised me step by step in my PhD research and with utmost devotion and dedication. It is impossible for me to put in words my appreciation and gratitude here.

In addition, I would like to thank Kamran Mortezaei Farid, vice-chancellor of BIHE University to whom I owe my success in the life and also to all of my lecturers during my undergraduate study as well as my colleagues in BIHE Computer Engineering Department who have dedicated their life for education of Bahá'í youths in Iran. The Bahá'í Institute for Higher Education (BIHE) was founded in 1987 in response to the Iranian government's continuing campaign to deny Iranian Bahá'ís access to higher education. Bahá'ís are the largest religion minorities in Iran who have been systematically persecuted since the inception of the Bahá'í Faith a century and a half ago. Persecution worsened after the Islamic Revolution of 1979. Since the early 1980s, the Iranian government has imprisoned, harassed, and executed Bahá'ís. The unrelenting religious persecutions involve a wide range of human rights violations, including systematic denial of access to higher education. As such, the origin of BIHE is rooted in a spirit of purposefulness, dedication, and a belief in the power of true education. The unique

circumstances surrounding BIHE, whereby professors teach without compensation, and all staffs and students participate in the university at great personal risks, have unwittingly forced university to become a leader in combining on-line learning, traditional classroom instruction and preparation for a successful career.

I would like to thank my beloved mum for her moral support during my study and her sacrifices in bringing me up, especially after passing away of my father. She has sacrificed her life for my education and success. Also I would like to thank my sister for her continues empowerment and encouragement and also other family members and friends in Iran and Australia who have always encouraged and supported me. I also would like to remember all the Bahá'ís who have been persecuted, imprisoned, and martyred in Iran. Since my childhood, I have witnessed all of these persecutions on one hand and the steadfastness, sacrifice, devotion, and love of Bahá'ís on the other hand. Even now at the time that I am writing this thesis, there are more than 160 Bahá'ís, 18 to 86 years old, in prison just because of their Faith including some of the lecturers of BIHE university.

Finally, I would like to acknowledge my appreciation of academic, administrative, and general staff in ICT Research Institute, SMART Services CRC, School of Electrical, Computer and Telecommunications Engineering and anyone else at University of Wollongong who helped me during my PhD.

I am truly grateful to be given the opportunity to become involved in this interesting area of research. This degree and the capabilities I have acquired during last four and half years would enable me to serve more efficiently locally to my community and globally to the world. In particular, I hope I can teach my fellow Bahá'í brothers and sisters in Iran.

Hooman Shidanshidi

March 2014,

Wollongong, Australia

Abstract

Free Viewpoint Video (FVV) aims to provide users with the ability to select arbitrary views of a dynamic scene in real-time. FVV systems widely adopt simplified plenoptic signal representations, in particular light field (LF). This is referred as an LF-based FVV system in this thesis. An LF-based FVV system consists of three main components: *acquisition component*, *rendering component*, and *compression/transmission component*. The efficacies of these components directly affect the quality of the output video.

The main aim of this research is to propose a novel theory and mathematical framework for analytical comparison, evaluation, and optimization of the LF acquisition and rendering components for a realistic *under-sampled* LF and *approximated depth information* with errors in depth maps. In contrast, most of the current researches on LF analytical evaluation focus on perfect signal reconstruction and are adequate to objectively predict and assess the influences of imperfections of acquisition and rendering on the output video quality.

In the core of the proposed theory there is the concept of effective sampling density (ESD). ESD is shown to be an analytically tractable metric that represents the combined impact of the imperfections of LF acquisition and rendering and can be used to directly predict/estimate output video quality from system parameters. The ESD for the commonly used LF acquisition configurations and rendering methods are derived and analyzed for evaluation and comparison. This claim is verified by extensive numerical simulations. Furthermore, an empirical relationship between the rendering quality (in PSNR) of a system and its ESD is established to allow direct prediction of the overall video quality without the actual implementation of the system. A small scale subjective user study is also conducted which indicates a high correlation between ESD and perceived quality.

In addition to comparison and evaluation of LF acquisition and rendering components and objective quality assessment of LF-based FVV systems, ESD theory is also applied to

several other significant problems. The first problem is *LF acquisition optimization*. In particular for a simplified regular grid acquisition, this optimization leads to calculation of the number of cameras required to capture the scene. Existing methods calculate the Nyquist density by assuming a band-limited signal and perfect reconstruction of an arbitrary view using linear interpolation, which often results in an impractically high number of cameras. In contrast, by employing ESD to solve this problem, it is possible to study the problem for under-sampled LF under realistic conditions (non-Lambertian reflections and occlusions) and rendering with complex interpolations. Theoretical and numerical results show that the resulting number of cameras is significantly lower than what was reported in the previous studies with only a few percent reduction in the rendering quality. Moreover, it is shown that the previous methods are special cases of the one derived from ESD theory.

The second problem is *LF rendering optimization*. The ESD theory is utilized to provide an estimation of the rendering complexity in terms of optimum number of rays employed in interpolation algorithm so as to compensate for the adverse effect caused by errors in depth maps for a given rendering quality. The proposed method is particularly useful in designing a rendering algorithm with inaccurate knowledge of depth to achieve the required rendering quality.

The third problem is a *joint optimization of both LF acquisition and LF rendering* to achieve a desired output quality. In particular, the trade-off among acquisition camera density, ray selection, depth error and rendering quality is studied using ESD and methods are presented to optimize these parameters for a system with a desired output quality in terms of ESD or PSNR by applying a Lagrangean method to ESD. Employing the proposed method on a regular grid camera system shows that the number of cameras can be reduced by 8 times if 32 rays, instead of 8 rays, are employed during rendering to achieve a similar rendering quality for a typical 20% error in depth estimation.

While in original presentation of ESD, the scene complexity is assumed to be fixed, the fourth problem focuses on the scene complexity and how a *non-uniform/irregular acquisition* can lead to a higher output quality. LF acquisition is theoretically considered as a problem of plenoptic signal sampling. It is typically performed by using a regular acquisition such as a regular camera grid. While a regular acquisition itself results in non-uniform sampling density, this non-uniformity does not match the scene complexity and frequency variations. To give a solution to the fourth problem the ESD theory is superimposed with the scene complexity and an irregular acquisition method is proposed for optimum non-uniform LF sampling corresponding to the variations of the scene complexity. Specifically, scene complexity is measured through analyzing DCT coefficients of reference images of the scene, describing the frequency behavior of the plenoptic signal over the scene space. An optimization model is formulated to calculate the optimum configurations of the acquisition cameras including positions and orientations. The theoretical analysis and numerical simulations demonstrate that the rendered video quality can be significantly improved (around 20% in mean PSNR) by employing the proposed irregular acquisition compared with the regular camera grid.

To validate the proposed theory, a simulation system is proposed. The simulator takes a 3D model of a scene and generates both reference cameras images and ground truth images. The proposed simulation system is highly flexible and efficient to automatically generate different datasets and objectively compare and analyze any LF-based FVV systems for any given experiment design scheme.

While the fundamentals of ESD theory is studied and reported in this thesis, the theory requires significant further research. The author is working on extending the ESD theory and applying it to more problems and will report the results in future publications.

Table of Contents

List of Figures

List of Tables

List of Notations

Notations introduced in Chapter 3:

Θ *a set of all known rays captured by LF acquisition*

Ω *a subset of rays in* Θ *intersecting the area A*

$\boldsymbol{\omega}$ *a subset of rays in* Ω *selected by mechanism M*

$\boldsymbol{M}$ *ray selection mechanism of an LF rendering method*

$|\boldsymbol{\omega}|$ *number of rays in* ω

$|\Omega|$ *number of rays in* Ω

$\boldsymbol{r}$ *unknown ray*

$\boldsymbol{p}$ *the intersection point of unknown ray* r *with the scene*

$\boldsymbol{p'}$ *the estimated intersection point of unknown ray* r *with the scene*

$\boldsymbol{d}$ *the depth of point p*

$\boldsymbol{\Delta d}$ *the error of depth estimation, the difference between d (depth of p) and estimated depth of* p'

$\boldsymbol{A}$ *the area of interpolation in vicinity of* p *(an imaginary convex hull around p)*

$\boldsymbol{A'}$ *the synthetic departure area produced in camera grid proportional to area of interpolation A*

$\boldsymbol{\mu}$ *a function to calculate the effect of pixel interpolation*

$\boldsymbol{R}$ *LF rendering method*

$\boldsymbol{F}$ *interpolation function employed in LF rendering process*

$\boldsymbol{G}$ *geometric information employed by LF rendering method*

SD *Sampling Density*

ESD *Effective Sampling Density*

$\boldsymbol{\beta_i}$ *boundary rays intersect the vertexes of the convex hull A*

$\boldsymbol{u_i}$ *a camera in the camera grid*

$\boldsymbol{s_j}$ *a pixel in the reference image*

$\boldsymbol{k}$ *the distance between the cameras in the camera grid*

$\boldsymbol{l}$ *the pixel length*

$\boldsymbol{n}$ *the number of cameras bounded by boundary rays* $\beta_i s$

$\boldsymbol{m}$ *the number of pixels bounded by boundary rays* $\beta_i s$

$\boldsymbol{uv}$ *camera plane*

$\boldsymbol{st}$ *image plane*

$\boldsymbol{NN}$ *LF rendering methods with Nearest Neighbourhood estimation*

$\boldsymbol{ST}$ *LF rendering methods using a 2D interpolation in image plane* st

$\boldsymbol{UV}$ *LF rendering methods using a 2D interpolation in camera plane* uv

$\boldsymbol{UVST}$ *LF rendering methods using a full 4D interpolation, e.g., quadlinear in both camera and image planes*

$\boldsymbol{UV-D}$ *LF rendering methods using focusing depth as geometric information for ray selection and employing UV interpolation*

$\boldsymbol{UVST-D}$ *LF rendering methods using focusing depth as geometric information for ray selection and employing UVST interpolation*

$\boldsymbol{UV-DM}$ *LF rendering methods using full depth maps as geometric information for ray selection and employing UV interpolation*

$\boldsymbol{UVST-DM}$ *LF rendering methods using full depth maps as geometric information for ray selection and employing UVST interpolation*

$\boldsymbol{UVDM}(\boldsymbol{d}, \boldsymbol{\Delta d}, \boldsymbol{k}, \boldsymbol{l}, |\boldsymbol{\omega}|)$ *a generic representation for UV-DM rendering method*

$\boldsymbol{MSE}$ *Mean Square Error used for PSNR calculation*

$\boldsymbol{PSNR}$ *Peak Signal to Noise Ratio*

$\boldsymbol{Observed\ PSNR}$ *the PSNR observed from experiments*

$\boldsymbol{ESD_{method}}$ *effective sampling density for a given rendering method*

$\boldsymbol{Expected\ MSE}$ *the MSE expected from observed PSNR*

$\boldsymbol{f}$ *the empirical function to map observed PSNR to calculated ESD*

$\boldsymbol{fESD_{method}}$ *the empirical function* f *for* ESD_{method}

$\boldsymbol{Q}$ *a parameter used to define* f

$\boldsymbol{P}$ *a parameter used to define* f

Notations introduced in Chapter 4 and 5:

$\boldsymbol{b}(u, v, s, t)$ *a continuous light field*

$\boldsymbol{q}$ *the sampling pattern*

$\boldsymbol{w}$ *a low-pass filtering and interpolation mechanism*

$\boldsymbol{r}$ * *the rendering output*

$\boldsymbol{B}(\boldsymbol{U}, \boldsymbol{V}, \boldsymbol{S}, \boldsymbol{T})$ *corresponding spectra of b in frequency domain (The Fourier transform of b)*

$\boldsymbol{Q}$ * *corresponding spectra of q in frequency domain*

$\boldsymbol{W}$ *corresponding spectra of w in frequency domain*

$\boldsymbol{R}$ * *corresponding spectra of r in frequency domain*

$\boldsymbol{f}$ * *the focal length*

$\boldsymbol{\delta}(.)$ *Dirac delta function*

$\boldsymbol{c_1}$, $\boldsymbol{c_2}$, $\boldsymbol{c_3}$ ***and*** $\boldsymbol{c_4}$ *integer constants used to define q*

$\boldsymbol{k_u}$ *the distance between cameras in u direction*

$\boldsymbol{k_v}$ *the distance between cameras in v direction*

$\boldsymbol{l_s}$ *the horizontal pixel length.*

$\boldsymbol{l_t}$ *the vertical pixel length.*

$\boldsymbol{B'}(\boldsymbol{S}, \boldsymbol{T})$ *the 2D Fourier transform of the captured image*

$\boldsymbol{H}$ *the scene complexity in frequency domain*

$\boldsymbol{N_d}$ *depth layers*

$\boldsymbol{d_{min}}$ *minimum depth of the scene*

$\boldsymbol{d_{max}}$ *maximum depth of the scene*

$\boldsymbol{k_u^{max}}$ *maximum allowed k in u direction to avoid aliasing*

$\boldsymbol{k_v^{max}}$ *maximum allowed k in v direction to avoid aliasing*

$\overline{\mathbf{ESD}}$ *the average of ESD throughout the scene space*

$\mathbf{ESD}_{\boldsymbol{Ideal}}$ *the ideal ESD with accurate depth estimation $\Delta d = 0$ and n rays interpolation*

$\boldsymbol{n}$ * *number of rays used for ideal rendering based on the scene reflection complexity*

[*] ***Please note that these notations are used more than once in the thesis and should be interoperated based on the context***

Notations introduced in Chapter 6:

$\boldsymbol{T}$ *minimum required rendering quality as desired* $\overline{ESD}$

$\boldsymbol{\nabla}\left(\mathbf{ESD}_{\boldsymbol{UVDM}(\boldsymbol{d},\boldsymbol{\Delta d},\boldsymbol{k},\boldsymbol{l},|\boldsymbol{\omega}|)}\right)$ *the gradient of* $\mathrm{ESD}_{UVDM(d,\Delta d,k,l,|\omega|)}$

$\boldsymbol{C}(\boldsymbol{k},|\boldsymbol{\omega}|)$ *the cost function to express the combined rendering and acquisition cost*

$\boldsymbol{C_k}$ *the cost associated with each camera in camera grid*

$\boldsymbol{C_\omega}$ *the rendering computational cost associated with* $|\omega|$

$\boldsymbol{\Lambda}\ (\boldsymbol{k},|\boldsymbol{\omega}|,\boldsymbol{\lambda})$ *Lagrangean representation of* $C(k,|\omega|)$

$\boldsymbol{\lambda}$ *Lagrange multiplier*

$\boldsymbol{X}$ *an auxiliary variable equal to* $\sqrt{|\omega|}$

Notations introduced in Chapter 7:

$\boldsymbol{h}$ *a parameter representing the scene complexity*

$\boldsymbol{U_h}(\boldsymbol{ESD})$ *a set of utility functions of LF signal reconstruction accuracy vs. ESD*

$\boldsymbol{p_i}$ *a small block in the 3D scene discretization*

$\boldsymbol{\varepsilon_i}$ *length of* p_i *in* X, Y *and* Z *directions*

$\boldsymbol{h_i}$ *complexity of* p_i *in term of frequency variations*

$\boldsymbol{p_i}(\boldsymbol{x_i},\boldsymbol{y_i},\boldsymbol{z_i},\boldsymbol{\varepsilon_i},\boldsymbol{h_i})$ *representation of* p_i *with Cartesian position of* (x_i,y_i,z_i)

$\boldsymbol{m}$ [*] *number of 3D blocks* p_i *in the scene*

$\boldsymbol{c_i}$ *a camera in acquisition component*

$\boldsymbol{n}$ [*] *number of cameras in the acquisition*

$\boldsymbol{G_n}$ *the acquisition component of n cameras* $\{c_1,c_2,\dots c_n\}$

$\boldsymbol{c_i}(\boldsymbol{x_i},\boldsymbol{y_i},\boldsymbol{z_i},\boldsymbol{\alpha_i},\boldsymbol{\beta_i},\boldsymbol{\theta_i},\boldsymbol{f_i})$ *representation of camera* c_i *with Cartesian position of* (x_i,y_i,z_i)

$\boldsymbol{\alpha_i}$ *the camera* c_i*'s orientation in* X *direction*

$\boldsymbol{\beta_i}$ [*] *the camera* c_i*'s orientation in* Y *direction*

$\boldsymbol{\theta_i}$ *the camera c_i's orientation in Z direction*

$\boldsymbol{f_i}$ *the camera c_i's internal parameters*

$\boldsymbol{U_{h_i}(ESD_R(G_j, p_i))}$ *the rendering accuracy for block p_i for acquisition G_j*

$\boldsymbol{W(h_i)}$ *the weight of h_i based on U_{h_i}*

$\boldsymbol{O(G_n)}$ *objective function for non-uniform plenoptic sampling*

$\boldsymbol{\delta_{U_{h_i}}}$ *marginal utility*

$\boldsymbol{\beta_1}$ * *a parameter in describing $\delta_{U_{h_i}}$ ($\beta_1 \geq 1$)*

$\boldsymbol{\beta_2}$ * *a parameter in describing $\delta_{U_{h_i}}$ ($\beta_2 \geq U_{h_i}$)*

$\boldsymbol{Q(c_i)}$ *a function which gives a set of scene blocks that can be seen/captured by camera c_i*

$\boldsymbol{E(G_j, p_i)}$ *a function which gives the number of cameras that are able to see/capture p_i for a given acquisition G_j*

$\boldsymbol{du}$ *camera grid discretization length, i.e., the minimum space required for a camera in u direction*

$\boldsymbol{dv}$ *camera grid discretization length, i.e., the minimum space required for a camera in v direction*

$\boldsymbol{O_u}$ *possible number of discrete orientations in u direction*

$\boldsymbol{O_v}$ *possible number of discrete orientations in v direction*

$\boldsymbol{N}$ *all possibilities for positioning the cameras in a given camera grid uv*

$\boldsymbol{s_r}$ *one potential camera in the camera grid*

$\boldsymbol{S}$ *a set of all potential cameras in uv plane.*

$\boldsymbol{\tilde{p}_i(\tilde{x}_i, \tilde{y}_i, \tilde{h}_i)}$ *approximated 2D scene blocks*

$(\tilde{x}_i, \tilde{y}_i)$ *position of the projected scene over image plane st*

* ***Please note that these notations are used more than once in the thesis and should be interoperated based on the context***

Notations introduced in Appendices:

$\boldsymbol{FOV_{camera}}$ *Camera's Field of View*

Chapter 1: Introduction

Free Viewpoint Video (FVV) [1-5] aims to provide users with the ability to select arbitrary views of a dynamic scene in real-time. It allows the user to interactively control the viewpoint and generate new views from any 3D (three-dimensional) position and orientation. FVV systems widely adopt simplified plenoptic signal [6] representations, in particular, using all the rays reflected from every point of the scene in all directions, referred to as a light field (LF) [7, 8]. This is referred as an LF-based FVV system in this thesis. The conventional LF model assumed no geometric information about the scene and, hence, required a large number of cameras to capture the scene. However, recent extensions integrate the conventional model with geometric information, in particular depth maps of the scene [9-15], to improve the rendering quality with less number of cameras. Compared to the traditional depth image based rendering (DIBR) methods, these extensions are relatively insensitive to errors in depth maps [16, 17].

An LF-based FVV system consists of three components: *LF acquisition* [18-23], *LF rendering* [9-14, 24, 25] and *LF compression/transmission* [1, 2, 26-29]. The efficacies of these components directly affect the quality of the output video. This research focuses on the analytical evaluation and optimization of the LF acquisition and rendering components, hence, for the remainder of the thesis, quality degradation as a result of compression and transmission is ignored. Figure 1.1 shows a general system diagram for an LF-based FVV system. The main research questions are also illustrated in Figure 1.1. Simplified transmission component is represented as a direct communication link between acquisition and rendering components.

Acquisition component typically involves two processes, *ray capturing* and *depth estimation*. The light field is often sampled by multiple cameras through the *ray capturing* process, which results in a certain number of rays acquired per unit area of the convex hull of

the surface of the scene at a location, referred to as sampling density (SD). SD is typically *not* uniform across the scene even for a regular camera grid. In addition, the *depth estimation process* provides an estimation of depth (e.g. depth maps) to be used in rendering. This can be obtained by specialized hardware, such as depth cameras, or computed from the images obtained by the multiple cameras. In either case, the depth estimation will have some error.

The *rendering component* aims to reconstruct an unknown ray r from the acquired rays. In general, this component can be decomposed into two processes: (i) the *ray selection process* that chooses a subset of acquired rays, purported to be in the vicinity of r; and (ii) the *interpolation process* that estimates or computes r from these selected rays.

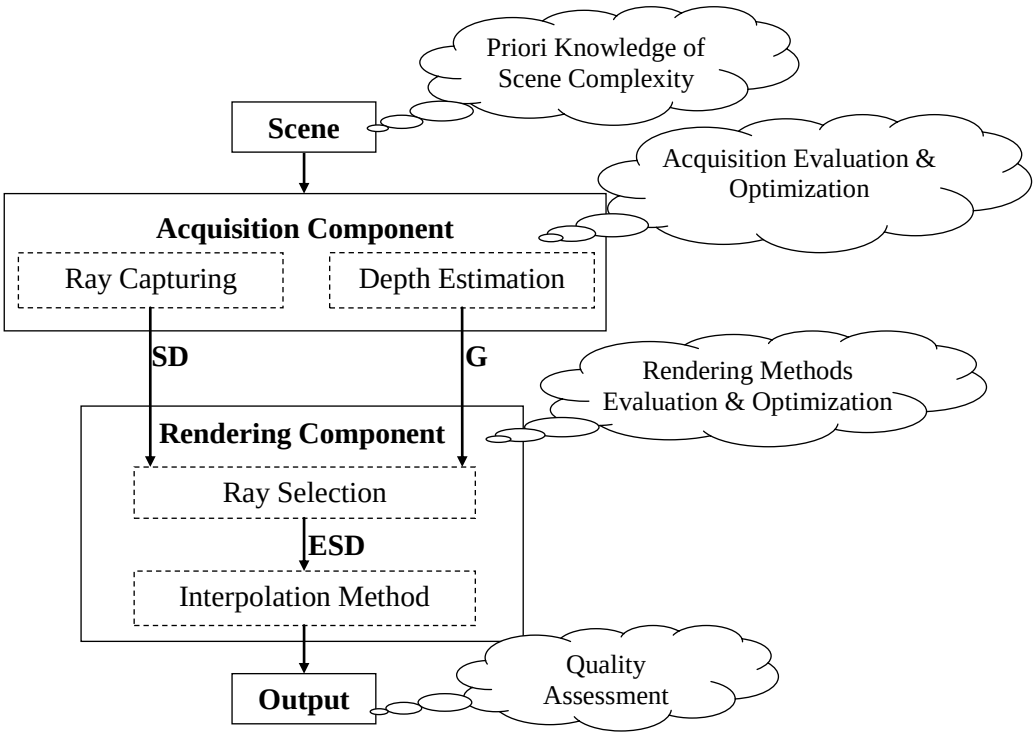


Figure 1.1. LF-based FVV system schematic diagram

The *ray selection process*, in particular, is often prone to error. For example, imperfect knowledge of depth may cause this process to miss some neighboring rays and choose others that are indeed sub-optimal (with respect to proximity to r) for interpolation. Also, constraints

on computational load (imposed due to, say, real-time rendering requirements) may necessitate this process to select only a subset of neighboring rays, less than what is available. In both cases, there is some loss of information and the output of this process represents an *effective sampling density* (ESD) [16, 30, 31] which is *lower* than the SD obtained by the acquisition component and distortion will inevitably introduced in the reconstructed video. ESD is defined as the number of rays per unit area of the scene that have been captured by *acquisition* component and chosen by *ray selection* process to be employed in the rendering.

1.1 LF-based FVV Research Questions

There are several research fields associated with light field (LF) and Free Viewpoint Video/TV (FVV/FTV) systems. The main research questions that this thesis is dealing with can be categorized based on the LF-based FVV system components as demonstrated in Figure 1.1.

1.1.1 Scene

The main research question dealing with the scene is that how the scene complexity can be acquired and represented. This includes *geometric information* of the scene, e.g., depth maps and *complexity information* of the scene in particular the geometric and texture frequency variations that can be employed for non-uniform acquisition.

1.1.2 Acquisition

There are several research questions dealing with acquisition component of an LF-based FVV system. In particular:

- Development of *depth estimation mechanism* for estimation of the depth of the scene.
- Analytical evaluation of the acquisition component for prediction of the influence of the acquisition on the output video quality. This allows analytical comparison of different acquisition architectures and evaluation of the effect of acquisition parameters, e.g., camera density on the output quality.
- Optimization of acquisition component for a given criteria on the output quality and acquisition cost. In particular acquisition optimization can be expressed in term of problems such as how many cameras are required in a regular camera grid to satisfy a given output quality? Or where to place the cameras in an irregular camera grid to produce the highest rendering quality for a given scene with known frequency variations.

1.1.3 Rendering

There are several research questions dealing with rendering component of an LF-based FVV system. In particular:

- Analytical evaluation of rendering method for prediction of the influence of the rendering method on the output video quality. This allows analytical comparison of different LF rendering methods and evaluation of the effect of rendering parameters on the output quality.
- Optimization of rendering component in particular *ray selection process* for a given criteria on the output quality and computational cost. For example how many rays should be selected by the *ray selection process*? And which rays? Or what type of *interpolation method* should be applied to these rays?

1.1.4 Output Video

The main research question dealing with the output video is development of an analytical model for quality assessment of the video, allowing direct prediction of output quality from acquisition and rendering parameters. This can be employed for theoretical evaluation and optimization of the system.

1.2 Current Approaches to the LF-based FVV Research Problems

Most of the current researches on LF analytical evaluation focused on one of these main fields: *signal processing theory*, for calculating the minimum sampling rate by LF frequency analysis and computing the signal Nyquist density for perfect signal reconstruction and optimum signal filtering, *optical analysis* of the LF system by assuming LF as a discrete synthetic aperture and calculating the optimum filtering, *geometric analysis of acquisition and rendering* to calculate the minimum number of cameras, and *computer vision and computer graphics* to design better rendering methods in term of efficacy and efficiency. On the other hand, more realistic *under-sampled* LF and *approximated depth information* with errors in depth map have not been investigated widely. As the result, most of the current analytical models are not adequate to objectively predict and assess the influences of imperfections of both acquisition and rendering on the output video quality. The details of these related works will be given later in chapter 2.

1.3 Thesis Contributions

The main aim of this research is to address the above research problems and to propose a theory for analytical comparison, evaluation, and optimization of the LF acquisition and

rendering components to overcome the limitations of other approaches in more realistic scenarios where (i) LF is *under-sampled*; and (ii) there is error in *depth information*. In the core of the proposed theory there is the concept of Effective Sampling Density (ESD) for LF-based FVV systems. ESD provides an analytically tractable way for evaluating the influence of the imperfections of *both* acquisition and rendering components on output quality. This has been demonstrated and verified in [31] and will be discussed later in chapter 3.

In particular, it is shown that for a fixed scene complexity and a given interpolation algorithm, ESD can objectively determine the quality of an LF rendering method for a given LF acquisition configuration and, hence, can be potentially used as an effective indicator for the quality of video generated from a corresponding LF-based FVV system.

According to the main research questions discussed before and refer to the proposed ESD, the main contributions of this thesis are discussed in this section.

1.3.1 Analytical Quality Assessment of Output Quality

- Introduction of the novel concept of effective sampling density (ESD) and an approach to calculate ESD for an LF-based FVV system.
- Verification that ESD is an effective indicator for quality, which can be employed to quantify the impact of acquisition and rendering on the final video quality separately as well as jointly. Calculation of ESD requires neither a reference/ground truth nor the actual output images/video. It can be derived from the key parameters of the acquisition and rendering components,
- Analytical evaluation and comparison of the widely used LF rendering methods with and without depth information for several LF acquisition configurations.

- Establishment of an empirical relationship between the estimated objective rendering PSNR as well as subjective perceived quality that can be used for predicting the output quality and/or optimizing system parameters.

1.3.2 Acquisition Evaluation and Optimization without Considering the Scene Complexity

- Study under-sampled LF under realistic conditions (non-Lambertian reflections and occlusions) and rendering with complex interpolations with ESD and discussing the relation of ESD with spectral analysis of light field.
- Demonstrating that both rendering quality and tolerance to errors in depth can be improved significantly by increasing ESD.
- Optimization of ESD with respect to the camera density for a given output quality and level of depth map estimation error.
- Calculating the minimum number of required cameras for a simplified acquisition component – a regular camera grid – to achieve a desired rendering quality for a given depth estimation error.
- Comparing the proposed method for calculating the minimum number of cameras with two well-known methods, Chai's LF spectral analysis [18] and Lin's LF geometric analysis [32] at different levels of errors in the depth estimation, and showing that the proposed method theoretically and numerically always results in a much lower number of required cameras with only a slight reduction in rendering quality (3-4% in PSNR). Moreover, it is shown that both Chai's and Lin's methods are special cases of the proposed method.

1.3.3 Rendering Evaluation and Optimization

- Proposing an analytical model based on ESD to study the impact of depth estimation errors on *ray selection process* and rendering quality.
- Demonstrating that the degradation of rendering quality caused by the errors in depth estimation can, to some extent, be quantified by ESD and compensated for by selecting more rays during interpolation.
- Optimization of ESD with respect to the number of rays within the interpolation area employed during rendering for a given output quality and depth estimation error.
- Deriving a mathematical expression to calculate the optimal number of rays required to compensate for errors in depth map in order to meet the specified rendering quality and computational efficiency.

1.3.4 Joint Optimization of Acquisition and Rendering Subsystems

- Study the trade-off among acquisition camera density, ray selection, depth error and rendering quality using the concept of ESD and present methods to optimize these parameters for a system with a desired output quality in terms of ESD or Peak-to-Signal Noise Ratio (PSNR).
- Joint optimization of ESD with respect to both variables, i.e., camera density and number of rays by using Lagrangean method. Notice that optimization with respect to density of the camera grid is associated with the system cost and optimization with respect to the interpolation reflects the computational complexity.

1.3.5 Acquisition Evaluation and Optimization with Consideration to the Scene Complexity

- Non-uniform sampling of plenoptic signal based on the scene complexity variation by proposing an optimal irregular LF acquisition.
- Formulating an optimization model from ESD by considering the scene complexity to calculate the optimum configurations of cameras, e.g., positions and orientations for an LF acquisition component.
- Representing the scene complexity as *scene complexity maps* by analyzing the spatial frequencies of references images of the scene.

1.4 List of Publications from This Thesis

1.4.1 Published Papers

[1] H. Shidanshidi, F. Safaei, and W. Li, "A Quantitative Approach for Comparison and Evaluation of Light Field Rendering Techniques," in *IEEE International Conference on Multimedia & Expo (ICME)*, 2011, pp. 1-4

[2] H. Shidanshidi, F. Safaei, and W. Li, "Objective Evaluation of Light Field Rendering Methods using Effective Sampling Density," in *IEEE International Workshop on Multimedia Signal Processing (MMSP)*, 2011, pp. 1-6

[3] H. Shidanshidi, F. Safaei, and W. Li, "A Method for Calculating the Minimum Number of Cameras in a Light Field Based Free Viewpoint Video System," in *IEEE International Conference on Multimedia & Expo (ICME)*, 2013, pp. 1-6

[4] H. Shidanshidi, F. Safaei, A. Zamani-Farahani, and W. Li, "Non-uniform Sampling of Plenoptic Signal based on the Scene Complexity Variations for a Free Viewpoint Video System," in *IEEE International Conference on Image Processing (ICIP)*, 2013, pp. 3147 – 3151

1.4.2 Under Review Papers

[5] "H. Shidanshidi, F. Safaei, and W. Li, "Estimation of Signal Distortion using Effective Sampling Density for Light Field based Free Viewpoint Video," *IEEE Transactions on Multimedia*, received a qualified accept in 2014 and is now under second revision.

1.4.3 Ready to be Submitted

[6] H. Shidanshidi, F. Safaei, and W. Li, "Optimization of Acquisition and Rendering Subsystems Using Effective Sampling Density for Light Field based Free Viewpoint Video," *IEEE Transactions on Multimedia*, TBA

[7] H. Shidanshidi, F. Safaei, and W. Li, "On The Minimum Number of Cameras to Capture the Light Field for Free Viewpoint Video Systems," TBA

[8] H. Shidanshidi, F. Safaei, and W. Li, "Non-Uniform Light field Acquisition based on the Scene Complexity Variations for a Free Viewpoint Video System," *IEEE Transactions on Image Processing*, TBA

[9] H. Shidanshidi, F. Safaei, and W. Li, “Optimization of the Number of Rays in Interpolation for Light Field Based Free Viewpoint Systems,” TBA

1.5 Thesis Structure

Chapter 2 summarizes the literature review and related works. Chapter 3 presents the main theory of ESD for LF based FVV systems, acquisition and rendering components comparison and evaluation, and objective quality assessment of output quality by using ESD. A simplified regular grid acquisition optimization for calculation of the minimum number of cameras to capture the light field for FVV systems is described in chapter 4. A simplified rendering component optimization in term of optimization of the number of rays during interpolation is demonstrated in chapter 5. Chapter 6 presents the joint optimization of acquisition and rendering subsystems by applying a Lagrangean method to ESD. Chapter 7 superimposes the ESD theory with scene complexity and demonstrates a non-uniform/irregular light field acquisition based on the scene complexity variations. Chapter 8 concludes the thesis. Extensive mathematics and the quantitative analysis of LF based FVV systems is given separately as appendices.

Following subsections give a summary for each main chapters of the thesis.

1.5.1 The Theory of ESD for Evaluation and Comparison of Acquisition and Rendering Components and Analytical Assessment of Output Quality

Quality assessment of a light field (LF) based free viewpoint video (FVV) system is usually confined to subjective evaluation of output since, typically, limited or no ground truth data is available. This chapter introduces the concept of effective sampling density (ESD). ESD is shown to be an analytically tractable metric that represents the combined impact of the

imperfections of LF acquisition and rendering. By deriving and analyzing ESD for the commonly used LF acquisition and rendering methods, it is shown that ESD is an effective indicator of quality determined by system parameters and can be used to directly estimate output video distortion without access to the ground truth. This claim is verified by extensive numerical simulations. Furthermore, an empirical relationship between the rendering quality (in PSNR) of a system and its ESD is established to allow direct prediction of the overall video quality without the actual implementation of the system. A small scale subjective user study is also conducted which indicates a high correlation between ESD and perceived quality.

1.5.2 Acquisition Optimization and Calculation of the Minimum Density of Cameras for a Regular Grid

Calculation of the number of cameras required to capture the scene is an essential problem in a practical light field based free viewpoint video (FVV) system. Existing methods calculate the Nyquist rate by assuming a band-limited signal and perfect reconstruction of an arbitrary view using linear interpolation, which often results in an impractically high number of cameras. This chapter proposes a new method based on the concept of effective sampling density (ESD). Specifically, the method assumes the availability of some depth information and explores the trade-off among the depth information accuracy, the required number of cameras, and the desired rendering quality. Theoretical and numerical results show that the resulting number of cameras is significantly lower than what was reported in the previous studies with only a few percent reduction in the rendering quality. Moreover, it is shown that the previous methods are special cases of the one presented in this chapter.

1.5.3 Rendering Optimization and Calculation of the Number of Rays in Interpolation

Light field (LF) rendering is widely used in free viewpoint video systems (FVV). Different methods have been proposed to employ depth maps to improve the rendering quality. However, estimation of depth is often error-prone. In this Chapter, a new method based on the concept of effective sampling density (ESD) is proposed for evaluating the depth-based LF rendering algorithms at different levels of errors in the depth estimation. In addition, for a given rendering quality, an estimation of the rendering complexity is provided in terms of optimum number of rays employed in interpolation algorithm so as to compensate for the adverse effect caused by errors in depth maps. The proposed method is particularly useful in designing a rendering algorithm with inaccurate knowledge of depth to achieve the required rendering quality. Both the theoretical study and numerical simulations have shown that the proposed method is reliable and accurate.

1.5.4 Joint Optimization of Acquisition and Rendering Subsystems by Applying Lagrangean Method to ESD

Quality of output video is an important usability objective in Free Viewpoint Video (FVV) systems. As shown in previous chapters, the density of a camera grid for acquisition on one hand and the complexity of the rendering algorithm on the other hand directly influence the FVV output video quality. As shown before ESD is an analytically tractable metric that can be used to predict and evaluate the FVV video quality for a given acquisition and rendering. To increase ESD and hence improve the video quality, two parameters can be altered: (i) the density of cameras in the acquisition grid; and/or (ii) the density of rays within the interpolation area employed during rendering. While in previous chapters individual optimizations of these parameters are demonstrated, in this chapter, a method is presented to optimize these parameters jointly for a target output video quality using ESD. Study on a regular grid camera system has

shown that the number of cameras can be reduced by 8 times if 32 rays, instead of 8 rays, are employed during rendering to achieve the similar rendering quality for a typical 20% error in depth estimation.

1.5.5 Non-Uniform/Irregular Acquisition based on the Scene Complexity Variations

Acquisition of a free viewpoint video (FVV) system is theoretically considered as a problem of plenoptic signal sampling. It is typically performed by using a regular camera grid. While a regular acquisition itself results in non-uniform sampling density, this non-uniformity does not match the scene complexity and frequency variations. This Chapter shows how to superimpose the ESD theory with the scene complexity and proposes an irregular acquisition method for optimum non-uniform LF sampling corresponding to the variations of the scene complexity. Specifically, scene complexity is measured through analyzing DCT coefficients of reference images of the scene, describing the frequency behavior of the plenoptic signal over the scene space. An optimization model is formulated to calculate the optimum configurations of the acquisition cameras including positions and orientations. The theoretical analysis and numerical simulations demonstrate that the rendered video quality can be significantly improved (around 20% in mean PSNR) by employing the proposed irregular acquisition compared with the regular camera grid.

Chapter 2: Literature Review

In this chapter, firstly, a brief review of the general background is given. Then a full review of the state of the art with respect to key research questions of this thesis is presented.

2.1 General Background

2.1.1 Free Viewpoint Video (FVV) Systems

Free Viewpoint TV (FTV) or Free Viewpoint Video (FVV) [1-5] aims to provide users with the ability to select arbitrary views of a dynamic scene in real-time. It allows the user to interactively control the viewpoint and generate new views from any 3D position and orientation. The main idea behind FVV is to change the focus of attention and story-telling from director's fixed chosen view to be controlled by the viewers, meaning that each viewer may be observing a unique viewpoint and can navigate the scene based on his preference.

A practical scenario is a user sitting in front of a 3D TV which also has FVV capability. The user not only can watch a 3D movie or a 3D football match, but a head tracking device can detect his head and eyes movements and change the viewpoint of the movie/match to his desired viewpoint accordingly. For the football match scenario the viewer can navigate the stadium and watch the match from any position and direction even from their favourite player viewpoint or goalkeeper viewpoint, similar to a 3D computer game experience. FVV will overcome the current shortcomings of 2D TV and Cinema, which is becoming less attractive for modern generation used to computer games and the freedom to control their viewpoint of the scene.

An FVV system consists of three main components: *acquisition component* [18-22], *rendering component* [9-14, 24, 25], and *compression/transmission component* [1, 2, 26-29]. Figure 2.1 illustrates a sample FVV system. There are a limited number of cameras with predefined positions and orientations around a stadium to capture the scene from several viewpoints. This is called *acquisition component* of an FVV system. There are different architectures and parameterizations for an FVV acquisition component which will be discussed in detail later. For this basic example, let's assume the cameras are located uniformly around the stadium placed in a number of parallel belts.

Rendering component of an FVV system aims to reconstruct the scene for any arbitrary/virtual viewpoints from these limited samples captured by *acquisition component*. Several arbitrary viewpoints are illustrated with an eye symbol in Figure 2.1. There are different rendering methods proposed for FVV systems that again will be discussed later in the related work section.

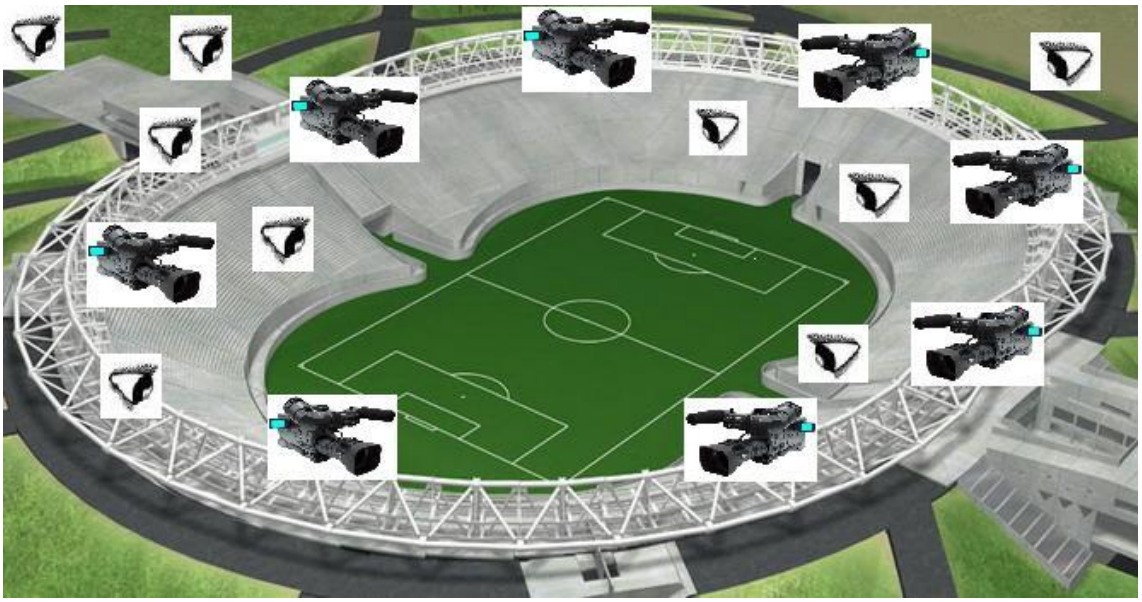


Figure 2.1. Overview of a sample FVV system

Since there are many users, simultaneously navigate the scene from their own desired perspective, FVV *rendering method* is aimed to be a client-side process. Hence all or a subset of captured video streams should be compressed and transmitted to the client side as a multi channel video. This would be carried out by *compression/transmission component* of an FVV system.

One of the main research questions on FVV is how to reconstruct or synthesize the scene in a virtual viewpoint which is the main responsibility of *rendering component*. Figure 2.2 shows the image reconstruction/synthesize process. A number of reference images captured by *acquisition component* cameras are employed to reconstruct the scene in an arbitrary virtual/ observation camera position and orientation.

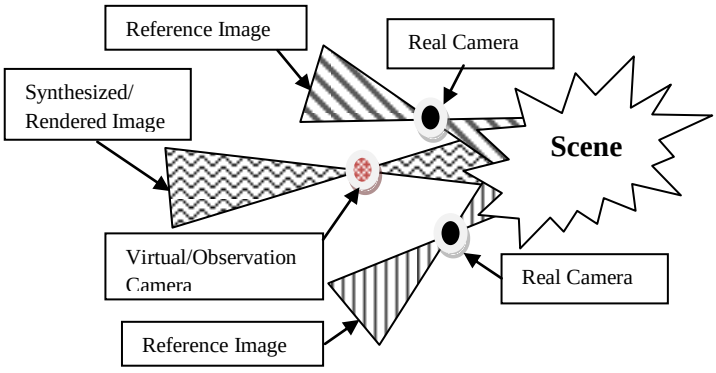


Figure 2.2. Image reconstruction/synthesize process

Different methods have been proposed for this image/video reconstruction process under three main categories [33, 34]:

a) Model based rendering, which employs explicit geometry based on computer vision techniques and scene geometry is described with 3D graphic elements such as meshes. Despite the advances in computer vision 3D reconstruction algorithms, reliable and accurate construction of full 3D scene models remains unsolved. Most of the proposed

methods are not real-time which is an important requirement in FVV systems. In addition, typically the rendering output is not photo realistic.

b) Image based modelling (IBM), image based representation (IBR), or image based rendering (IBR) which do not explicitly employ any 3D geometry, but attempt to reconstruct the scene directly from given reference images as demonstrated in Figure 2.2. The main advantage is that it can potentially produce high quality of synthesized views when dense sampling of the real world is obtained with sufficiently large number of cameras.

c) Hybrid systems, which expand IBR with implicit geometric information such as depth or disparity map. The original two-dimensional image and the depth map forms 2.5D representation of the scene. When full depth information is available, 3D warping techniques can be used to synthesize virtual views [35]. However, the quality of the warped images is subject to the sampling resolution and degree of occlusion.

2.1.2 Image Based Representation (IBR)

Due to simplicity of data acquisition, photo realistic rendering and real-time computation, IBR has been widely used in FVV systems. Note that category (b) and (c) in previous section are both IBR methods which are distinguished here based on using no or some implicit geometric information during image/video synthesis process. IBR includes a variety of models, representations, and methods employing reference images as the primary elements from which arbitrary virtual views are synthesized without the full 3D model reconstruction. In IBR [6-8, 12, 33, 34, 36-44], novel views (also called rendered images or synthesized images) are generated/reconstructed from a set of sampled images or videos of the scene, called the references images/videos. Essentially, the novel view rendering is based on data prediction, estimation or interpolation of a multidimensional space which is densely sampled as represented

by the reference images. Typical IBR rendering techniques include the panoramas [38] , light field and its variants [7, 8, 12, 40, 41], and concentric mosaics [39].

2.1.2.1 IBR Spectrum based on the Geometry Information

IBR methods employ different amounts of geometric information of the scene which can be demonstrated as a spectrum [45]. At one extreme, there are methods with no usage and representation of geometric information such as 2D panoramas[38], 3D concentric mosaics[39], 5D McMillan and Bishop's plenoptic modeling [37], 4D rays pace representation [42, 43], and 4D conventional light fields [7] or lumigraph [8]. At the other extreme there are methods employing explicit geometry such as 3D computer graphics models and other more sophisticated representations [46-48]. In the middle of the spectrum, there are methods employing different extents of implicit geometric information from minimum information such as focusing depth to depth layers or object-based representations such as [12, 40, 49] to full depth map information such as depth-based image rendering (DIBR) [29]. It is obvious that methods with less geometry require higher number of reference images to compensate for inaccurate or lack of knowledge of scene geometry.

2.1.3 Plenoptic Signal

From the signal processing point of view, IBR can be mathematically modelled as a process of sampling and reconstruction of the complex 7D plenoptic signal [6], $P^7 = (V_x, V_y, V_z, \theta, \varphi, \lambda, \tau)$, representing the radiant energy that is perceived at any 3D viewing point (V_x, V_y, V_z), from every possible angle (θ, φ) for every wavelength λ and at any time τ. FVV *acquisition component* is responsible to collect and record samples from plenoptic signal by mapping the reference images to the signal space. Note the term *ray* is used to demonstrate one

sample of the plenoptic signal. Each pixel of reference images is considered as a sample of a ray reflected from the scene at the time of taking the image and received by the camera sensor. FVV *rendering component* is responsible to map the rendering image/video to the signal space and reconstruct the unknown values of the signal from acquired samples. Hence, usual signal processing methods for signal reconstruction can be customized for the rendering process. With the same approach, FVV *compression/transmission* is responsible for compression and transmission of the plenoptic signal.

2.1.3.1 Plenoptic Signal Simplifications for Different IBR Methods

Due to complexity associated with the high dimensions of the plenoptic signal, complete sampling of the function is impractical if not impossible. Hence, assumptions have been introduced to reduce the signal dimensions. Following are common assumptions to reduce the number of dimensions for plenoptic signal:

1. The wavelength can be restricted to three digital RGB channels. Each channel represents a range of wavelength captured by the camera sensor.

2. The radiance along a light ray in empty space remains constant. Hence, it is not required to record the radiance of a ray on different points on its path, but just one value for the whole line for each ray leaving the convex hull of a bounded scene. The plenoptic signal can be represented by its values along an arbitrary surface surrounding the scene or a set of lines which can reduce the plenoptic function by one dimension.

3. By assuming the scene to be static, the time dimension could be eliminated. This assumption is still valid for video as each frame in the video can be considered as an image.

4. The viewer is restricted to move on a surface rather than freedom of moving in the 3D space which makes the viewer space as desired values of plenoptic function become 2D.

5. By introducing more restriction similar to assumption 4, the viewer movement could be restricted to a certain path. The user can move forward or backward but not out of the path.

6. Assuming fixed position for the viewer.

Table I summarizes different IBR representations, the assumptions applied and the space dimension for each simplified plenoptic signal [34, 36, 45]. Comprehensive reviews of the problems of sampling, rendering, and compression of IBR methods and representations are available in [33], [34] and [36].

Table 2.1

IBR representations and plenoptic signal space simplification [34]

Dimension	Representation	Assumptions
7D	Plenoptic Function	Nothing
6D	Surface Plenoptic Function [23, 50]	2
5D	Plenoptic Modelling [37]	1,3
5D	Light Field Video [51]	1,2
4D	Light Field/Lumigraph [7, 8, 42, 43]	1,2,3
3D	Concentric Mosaics [39]	1,2,3,4
3D	Panoramic Video [52-54]	1,6 or 1,3,5
3D	Branch Movies [55, 56]	1,3,5
3D	Normal Video	1,6
2D	Image Mosaicing [38]	1,3,6
2D	Normal Image	1,3,6

Though in an ideal world, the 7D plenoptic signal is the best IBR representation but due to several limitations and most importantly acquisition restrictions, the best practical simplified plenoptic signal representation for FVV systems is light field (LF) or ray space representation. By using a typical FVV *acquisition component* which captures the scene outside of the scene convex hull and by digitizing the wavelength to RGB channels, each pixel in the reference images corresponds to one ray (a sample of plenoptic signal). LF is a simplified 5D plenoptic signal or with ignoring the time a 4D signal (by assuming a static scene or by time freezing in each frame of the video) that can be geometrically defined as a set of oriented lines (rays) in Cartesian 3D space. There are several acquisition architecture and parameterizations for 4D LF that will be discussed later.

2.1.4 Light Field

Light field can be expressed as a simplified four dimensional plenoptic signal [6], first introduced by Levoy and Hanrahan [7] and Gortler et al [8] (as Lumigraph) in the mid-1990s. LF acquisition aims to sample the plenoptic signal by using a limited number of cameras configured in 3D space. Several parameterization schemes have been proposed to represent the camera configurations and the rays captured by the cameras. For instance, Levoy and Hanrahan [7] employed a regular grid of cameras and represented the rays by using their intersection points with two parallel planes/slabs defined by variables (s, t, u, v) respectively, where (s, t) represents the image plane and (u, v) represents the camera plane. The 4D space is then represented as a set of oriented lines, i.e., *rays* in 3D space. The parallel planes parameterization has been enhanced by more complicated representations or parameterization schemes such as Two-Sphere Parameterizations (2SP) , Sphere-Plane Parameterizations (SPP) and a Direction and Point Parameterizations (DPP) [57, 58].

2.1.5 Light Field Acquisition

As discussed before, in IBR the acquisition usually consists of a number of cameras located in predefined positions and orientations with a specific architecture. Generally the cameras in the acquisition are normal digital cameras/video cameras, however, recently with the advances in computational photography and light field cameras [51, 59, 60], the cameras in the *acquisition component*, can be replaced by light field cameras. Several pre-processing methods should be applied to the raw images of the acquisition cameras before mapping them to the LF signal samples/rays (or as some literature call *ray space*). These pre-processing includes common computer graphic and vision methods such as multi-camera calibration and colour correction as well as signal processing pre-filtering.

Regular camera grids are widely used for LF acquisition due to their simplicity, low setup cost and computationally effective mapping between the output images and LF samples/rays. The well-known 2PP LF representation (two parallel planes/slabs) can also be easily employed in a regular camera grid acquisition. The regular camera grid or multi camera array has been employed in many studies such as Stanford multi-camera array [61] which consists of 128 cameras and is intended for large-environment applications by using low-cost CMOS sensors and dedicated hardware for real-time compression, the 3D rendering system of Naemura et al. [62], and the (8×8) light field camera of Yang et al. [63]. There are also some studies on dynamic rearrangement of the grid such as (8×6) self reconfigurable camera array of Zhang and Chen [64] which for a given virtual view point, moves the cameras on a set of rails to perform active rearranged capturing to improve the rendering quality. Figure 2.3 shows the Stanford regular camera grid [65] as an example.

Other architecture for acquisition are also used, such as multi camera ring which is mainly used for bullet-effect generation [40] , locating the cameras in a semi-cylinder or semi-sphere topology [66], moving the cameras by robots [8] or hand-held cameras[67].

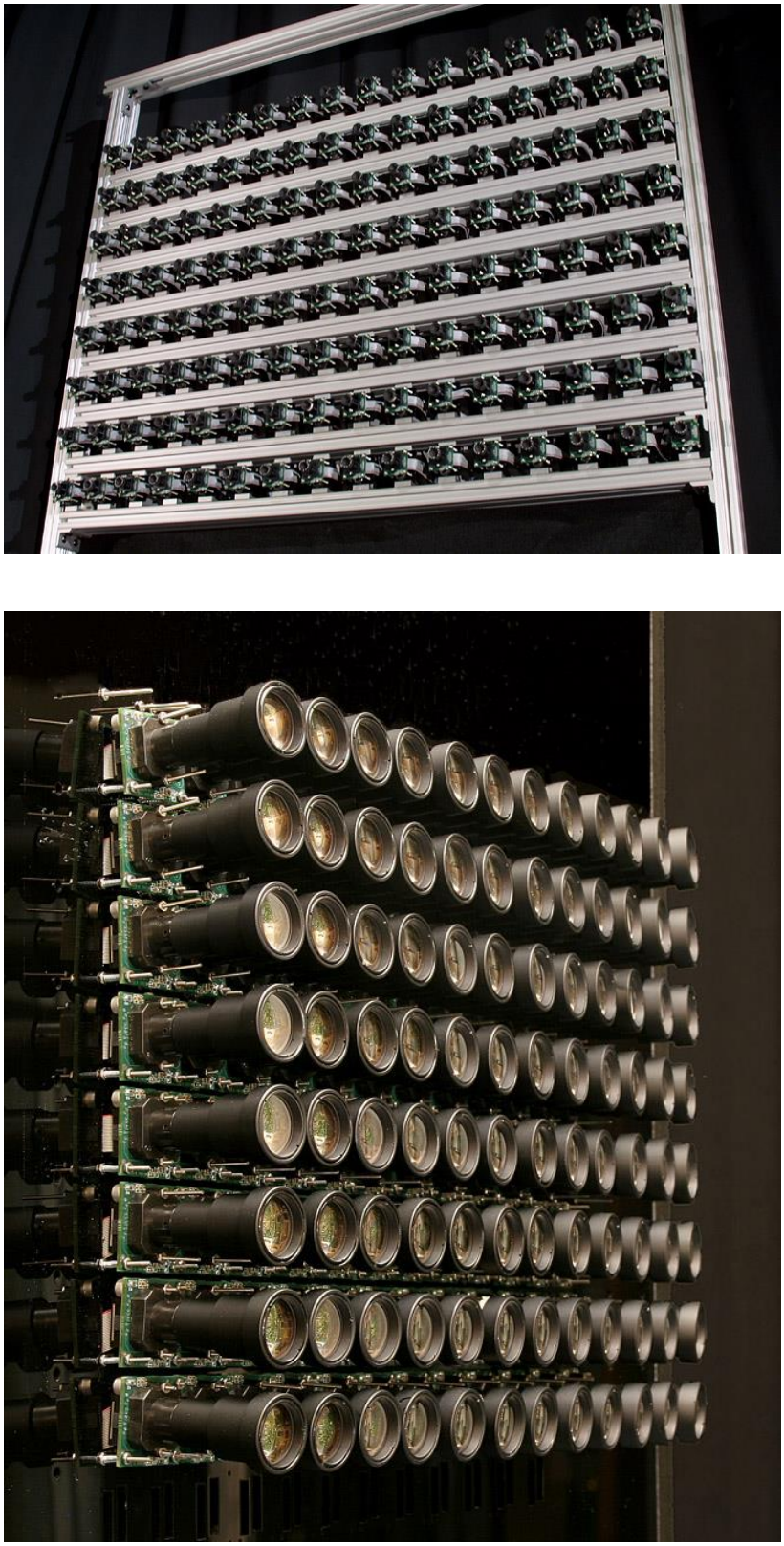


Figure 2.3. Stanford regular camera grid

2.1.6 Light Field Rendering

To synthesize the image in a virtual camera position, a Light field rendering method is required. The synthesized image would be directly rendered from acquired light field and depth information of the scene. Figure 2.4 demonstrates a very simple mechanism of the rendering with no depth information.

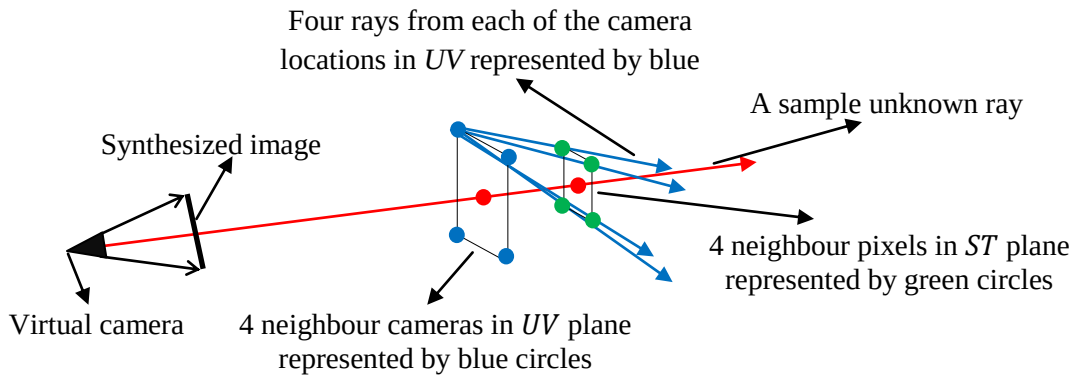


Figure 2.4. A simple light field rendering mechanism

The rendering process starts by posing a virtual camera in 3D space. The position, orientation and other parameters of this virtual camera is calculated from the user preferred view point. Each pixel in the synthesized image corresponds to one ray in the space. One example of an unknown ray is shown in Figure 2.4. In a simplified 2-planes representations, this unknown ray will intersect both camera plane UV at (u, v), and image plane ST at (s, t). In this simple representation, there are 16 rays as immediate neighbours of this unknown ray sourcing from 4 neighbouring real cameras in UV and for each camera passing through 4 known neighbouring pixels in ST. Four of these rays are illustrated in Figure 2.4. At least four different interpolations can be employed to estimate the unknown ray from these captured rays. Neighbourhood estimation (NN) method estimates the unknown ray with the closet ray among this 16 rays. UV interpolation selects 4 rays – from each neighbouring cameras to the closest known pixel in ST – , and apply a 2D interpolation such as a bilinear interpolation to estimate the unknown ray. ST interpolation selects the closest camera in UV and applies a bilinear pixel interpolation to the 4

neighbouring pixels in *ST*. UVST method applies a 4D interpolation such as quadlinear interpolation to all 16 immediate neighbouring rays to estimate the unknown ray.

The mechanism of ray selection and interpolation can be more complex and more intelligent compared with this simplified rendering method, in particular by superimposing the rendering method with geometric information of the scene such as depth maps. The full mathematical formulation of a generic light field rendering method will be discussed later in chapter 3 of the thesis.

By superimposing the rendering method with depth information specially in *under-sampled regime* different LF rendering methods have been developed to generate images for arbitrary viewpoints from the captured rays by implicitly or explicitly using geometric information about the scene [15]. These include layered light field [9], surface light field [10] , scam light field [11], pop-up light field [12], all-in-focused light field [13], dynamic reparameterized light field [14], and unstructured light field [68, 69].

The effect of *under-sampling* and amount of depth information and computational efficiency restrictions on the rendering quality will be discussed later in the thesis.

2.1.7 LF Compression and Transmission

The light field compression and transmission is out of the scope of this thesis. However extensive research has been carried out on multi-view video and multi-view video coding (MVC) such as [2, 28]. The standardization of MVC is also carried out as an amendment to H.264/MPEG-4 AVC video compression standard.

This concludes the introduction of the general background. The remaining of this chapter will demonstrate the state of the art and related work on each research questions discussed before in the Introduction chapter.

2.2 Quality Assessment of LF-based FVV Systems

As discussed before FVV [1, 2] aims to provide users with the ability to select arbitrary views of a dynamic scene in real-time. FVV systems widely use simplified plenoptic signal [6] representations, in particular, light field (LF) [7, 8]. Quality assessment of LF-based FVV systems is usually concerned with the quality of the rendered scene in comparison with the ground truth. Clearly, this is an important consideration in evaluating the effectiveness of rendering algorithms and acquisition configurations and is a key objective for optimizing the system. While subjective and objective empirical evaluation of output, with limited or no ground truth data, has been widely used, to the best of our knowledge, no analytical evaluation framework has been reported so far.

Light field is often represented as a simplified four-dimensional (4D) plenoptic signal [6]. Acquisition is to sample the signal and the rendering process is to reconstruct the signal from the acquired samples. Therefore, assuming a band-limited signal and linear interpolation, a perfect signal reconstruction would require sampling the rays from the scene at the Nyquist density. Several studies on LF acquisition analysis, such as [18-21, 23], have mainly focused on this minimum sampling density of such a perfect acquisition system. In practice, however, it is often infeasible to deploy sufficient number of cameras to capture the rays in all positions and directions at the Nyquist density (the term *ray* is used in LF to denote a sample of plenoptic signal). Thus, a realistic LF acquisition almost always results in an under-sampled signal, which may result in imperfect reconstruction with visual artefacts and signal aliasing. To overcome this problem, several LF rendering methods such as [9-14] have been proposed to work specifically *in the under-sampled regime* with additional auxiliary information about the scene. Most of these methods employ certain scene geometric information, such as depth map, to compensate to some extent for the insufficiency of samples.

Accordingly, an analytical evaluation of an LF system *in the under-sampled regime* that captures the impact of both acquisition and rendering is crucial for assessing the overall quality

of an LF-based FVV system. Nevertheless, the existing assessment methods are mainly based on subjective evaluation and comparison with the ground truth [70-73] and are limited to case-based studies. Typically, after discussing the new method, the method is applied to several test cases for validation. The rendering output then demonstrated with some sample images. To show the method improvement over older methods, some primitive real or artificial objects are chosen. Then the rendering outputs for some viewpoints for new and old methods are given for a limited subjective visualized comparison. The evaluation methodology is therefore primarily subjective, which is highly related to observer and test cases and therefore may not be reliable. In addition, a main hurdle in conducting this comparison is the unavailability of ground truth data to compare and validate the rendered/synthesized images.

2.3 Evaluation of the LF Acquisition Component

Existing approaches for evaluating LF acquisition mainly focus on the minimum required sampling density by assuming an ideal and perfect signal reconstruction. Two major approaches have been studied so far. The first one is based on plenoptic signal spectral analysis [18, 23] and, more specifically, the light field spectral and frequency analysis [19, 20]. In this approach the spectral analysis is applied to a surface plenoptic function (SPF) representing the light rays starting from the object surface and the minimum sampling density is estimated based on the sampling theory by computing the Fourier transform of the light field signal. However, the spectrum of a light field is usually not band-limited due to non-Lambertian reflections, depth variations and occlusions. Therefore, the first-order approximation is often applied to the signal by assuming that the range of depth is limited.

The second approach is based on the view interpolation geometric analysis rather than frequency analysis. This approach is based on blurriness and ghost (shadow)-effect error measurements and elimination in rendered images. In [21] the artifact of “double image” (a

geometric counterpart of spectral aliasing) is proposed to measure the ghost effect for a given acquisition configuration. This artifact is geometrically measured by calculating the intensity contribution of rays employed in interpolation. Finally, the minimum sampling density is calculated to avoid this error for all points in the scene. This approach can be used to derive the minimum sampling curve against scene depth information, showing how the adverse effect of errors in depth information can be compensated for by increasing the number of images or effectively the sampling density. This method is more flexible, especially for irregular capturing and rendering configurations, and leads to a more accurate and smaller sampling density compared with the first approach.

In addition to these two approaches, optical analysis by considering light field as a virtual optical imaging system is also employed in acquisition analysis [74, 75]. The original light field [7] shows that the distance between two adjacent cameras can be considered as the aperture for ray filtering. This concept is generalized in [14] by introducing a "discrete synthetic aperture", encompassing of several cameras. It is also shown in [14] that the size of this synthetic aperture can change the field of view very similar to an analog aperture. This optical analysis is mostly used to calculate the optimum light field filtering [76].

Due to the assumption of perfect signal reconstruction, all of these approaches result in very high sampling densities, which are hardly achievable in practice. For instance [18] shows that for a typical scenario a camera grid with more than 10,000 cameras is required. They also assume general Whittaker–Shannon interpolation method for signal reconstruction. However, having some geometric information about the scene, such as estimated depth map, could enable more sophisticated interpolation for signal reconstruction and rendering. Consequently, an indicator to measure signal distortion without any reference or ground truth, that works in the *under-sampled regime,* is desirable.

For a simplified camera grid acquisition, the problem of LF acquisition evaluation would reduce to a more familiar problem: what is the minimum number of cameras required to

capture the scene for a desired output video quality? Above three approaches have been addressed this question from a number of perspectives [18-21, 74, 75]. Please note that these studies refer to this problem as "LF minimum sampling rate" rather than "minimum number of cameras". In this thesis, the two terms will be used interchangeably hereafter since cameras are usually the basic sampling devices. These studies can be categorized into three different approaches.

However, these approaches are essentially based on several common unrealistic assumptions and also often suggest an impractically high number of cameras. Specifically,

- They assume a simple Lambertian scene with no occlusions to make the light field signal band-limited for frequency analysis. In real applications, these assumptions are seldom valid.

- They often assume a linear interpolation over 4 rays in the camera plane or 16 rays in both camera and image planes in the rendering process. However, in a practical scenario, there are usually more rays available for interpolation. In addition, by employing nonlinear interpolation techniques or incorporation of auxiliary information, such as depth, the quality of rendering could be enhanced further. Employing more rays in rendering can improve the rendering quality even without increasing the number of cameras.

- They aim for perfect signal reconstruction and calculate the Nyquist sampling rate. This typically results in a very high sampling rate which is not feasible in most of real FVV systems. In contrast it is a common practice to employ controlled under-sampling with additional processing such as anti-aliasing filtering and lossy signal reconstruction, to attain a desired output quality with significantly reduced number of samples.

Therefore, an analytical model to study under-sampled LF under realistic conditions (non-Lambertian reflections and occlusions) and rendering with complex interpolations is required to calculate the minimum number of cameras.

In this thesis, a method is derived from the proposed ESD [16, 30] to calculate the minimum number of required cameras to achieve a desired rendering quality by taking into consideration the possible error in depth estimation.

2.4 Evaluation of the LF Rendering Methods

Along with the acquisition configurations and parameterization schemes, different LF rendering methods have been developed to generate images for arbitrary viewpoints from the captured rays by implicitly or explicitly using geometric information about the scene [15]. Previous works on FVV evaluation and quality assessment are mainly based on the methods proposed for Image based Rendering (IBR) and they are not specifically for LF rendering. Often pixel-wise error metrics with respect to ground-truth images are employed for quality assessment [40]. Ground-truth data is provided by employing a 3D scanner for a real scene or virtual environments such as [77]. In [78], two scenarios are analysed: human performance in a studio environment and sports production in a large-scale environment. A method was introduced to quantify error at the point of view synthesis. This method is used as a full-reference metric to measure the fidelity of the rendered images with respect to the ground-truth as well as a no-reference metric to measure the error in rendering. In the no-reference metric, without explicitly having the ground truth, a virtual viewpoint is placed at the mid-point between the two cameras in camera grid. From this viewpoint, two images are rendered, each using one set of the original cameras. These images are then compared against each other with the same metrics as before.

Quality evaluation has also been carried out with two different categories of metrics, modelling the human visual system (HVS) and employing more direct pixel fidelity indicators. HVS-based measures of the fidelity of an image include a variety of techniques such as measuring mutual information in the wavelet domain [70], contrast perception modelling [71]

and modelling the contrast gain control of the HVS [72]. However, HVS techniques and objective evaluation of a visual system are not able to fully model the human perception as discussed in [79-81]. Pixel-wise fidelity metrics such as MSE and PSNR are simple fidelity indicators but with a low correlation with visual quality [82]. In [73] a full review of pixel-wise fidelity metrics is discussed. Also [83] shows a statistical analysis of pixel metrics and HVS-based metrics.

While the need for analytical quality evaluation of FVV systems are highlighted in several studies such as [84, 85], the current research on LF rendering evaluation and quality assessment is focused mostly on case-based study of applying these metrics. Little development has been reported on an analytical model that can evaluate LF rendering methods. In contrast, the proposed ESD in this thesis provides an analytical evaluation of the effect of LF rendering methods as well as LF acquisition on the final video quality.

2.4.1 Scene Geometric Information Representation and Application

While most of the conventional LF representations did not use much geometric information, recent studies proposed models for superimposing depth information [41, 52, 61-64, 86] and more dynamic LF representations to achieve higher rendering quality with less number of cameras. The idea initiated from the early work on the panoramic videos [52] and extended to construct LF video for different applications with employing geometric information about the scene. Below some of these studies are reviewed.

A simplified light field for dynamic environments (SDLF) [41] which also referred to as “plenoptic videos” employs a regular camera grid to capture the videos and uses the depth information for both LF dynamic representations and the rendering process. Despite the simplifications in their model, their experiments show that it can provide an acceptable video quality with no discontinuity of viewpoints as well as tolerating the lighting changes.

The main issue of SDLF representation is lack of tolerance to depth discontinuity and occlusions which results in aliasing and artifacts in the output video. SDLF assumes no occlusion and simplified Lambertian scene reflection and uses plenoptic sampling [18] to determine the number of cameras. Unfortunately, these assumptions are rarely the case in real applications.

To overcome SDLF limitations, LF representations by employing the full depth maps and object segmentation information [12, 40, 49, 86] have been proposed to reduce rendering artifacts at depth discontinuities. To integrate the depth information in LF rendering, methods such as [86] has been used. In [86], the Stanford camera grid is used for LF acquisition and warping techniques used for rendering. However it was shown that it is difficult to produce accurate depth maps with less than one pixel error at the boundaries which again results in errors in warping outputs.

In another study [40], an eight-camera [6] video capturing system is employed by using high resolution FireWire PtGrey cameras to capture 15 frames per second (fps) video with resolution of 1024*x*768. For rendering they used layered depth images [49] generated with stereo matching methods and Bayesian matting [87].

Another interesting work is object-based LF representation [49, 88], where the depth and shape information is only required for each object in the scene in each time frame. The main advantage of this representation is that it allows user interaction with each object in the scene.

Another extension to LF modeling with depth information is the concept of pre-computed radiance transfer (PRT) [89] and pre-computed shadow fields [90] which allows effective interactive visualization with low complexity for real-time relighting and soft-shadow computation.

2.4.2 Optimization of the Number of Rays in Interpolation for Rendering

Many LF rendering methods have been developed so far. Some of them assume that the light field has been sampled sufficiently and employ a simple view interpolation process. Recent studies [9-14] have shown that implicit or explicit use of geometric information, such as a depth map of the scene, can significantly improve the rendering quality since in most practical cases, the light field is highly under-sampled. However, none of these methods address the impact of depth map estimation error on the rendering quality and how the rendering method can compensate for the error in depth estimation

A typical approach to compensating for the errors in depth maps is to increase the number of cameras of acquisition component [18-21] to increase the SD. It has also been shown that the adverse effect caused by the depth errors can be to some extent compensated for by increasing the number of cameras used in acquisition, which may not be affordable in practice.

However, we are not aware of any results on the role of rendering component in the more realistic *under-sampled regime* on output quality and how the adverse effect caused by errors in depth maps can be compensated for by employing optimal number of rays in the *ray selection process* for a fixed acquisition camera grid. The current LF rendering methods often assume a linear interpolation over 4 rays in the camera plane or 16 rays in both camera and image planes in the rendering process, despite the fact that more rays may be available.

2.5 Non-Uniform/Irregular LF Acquisition based on the Scene Complexity Variations

Typically, regular acquisition such as a regular 2D camera grid, regular two-sphere (2SP) and regular sphere-plane parameterizations (SPP) [57] are employed in an FVV system. LF rendering methods that make use of this regularity are employed for view reconstruction.

While a regular LF acquisition itself results in non-uniform sampling density, this non-uniformity usually does not match the scene complexity and frequency variations. Instead, often homogeneous frequency variations throughout the scene space is assumed despite the fact that, in reality, plenoptic signal has significant spatial frequency variations. Hence, an irregular/non-uniform signal sampling and reconstruction is likely to be more appropriate.

The theory of irregular/non-uniform signal sampling has been widely investigated and it has been shown that irregular sampling can reduce the number of required samples for perfect reconstruction of the signal [91-93]. However to the best of our knowledge, this property has not been explored for LF acquisition and rendering. Several studies have been reported to address the plenoptic sampling by computing the Nyquist rate [18-21, 74, 75], all assuming regular/uniform acquisition. To address this problem, one of the main contributions of this thesis is to propose an optimal irregular LF acquisition.

Chapter 3: The Theory of ESD for Evaluation and Comparison of Acquisition and Rendering Components and Analytical Assessment of Output Quality

3.1 Summary

Quality assessment of a light field (LF) based free viewpoint video (FVV) system is usually confined to subjective evaluation of output since, typically, limited or no ground truth data is available. This chapter introduces the concept of effective sampling density (ESD). ESD is shown to be an analytically tractable metric that represents the combined impact of the imperfections of LF acquisition and rendering. By deriving and analyzing ESD for the commonly used LF acquisition and rendering methods, it is shown that ESD is an effective indicator determined by system parameters and can be used to directly estimate output video quality without access to the ground truth. This claim is verified by extensive numerical simulations. Furthermore, an empirical relationship between the rendering quality (in PSNR) of a system and the calculated ESD is established to allow direct prediction of the overall video quality without the actual implementation of the system. A small scale subjective user study is also conducted which indicates a correlation of .91 between ESD and perceived quality.

The rest of the chapter is organized as follows. Section 2 describes the problem statement and motivation. Section 3 describes the concept of ESD. Section 4 presents the application of ESD to analyze the commonly used LF rendering methods. Numerical simulation and validations are presented in section 5. Section 6 describes the empirical relationship between the ESD and overall video quality in PSNR. Section 7 reports the subjective test and its correlation with ESD. Section 8 concludes the chapter.

3.2 Problem Statement and Motivation

By assuming a simplified transmission component represented as a direct communication link between the acquisition and rendering components, a general FVV system diagram that utilizes the information of scene geometry can be illustrated as Figure 3.1.

The light field is often sampled by multiple cameras through the *ray capturing* process of the acquisition component, which results in a certain sampling density (SD). SD at a given location can be defined as the number of rays acquired per unit area of the convex hull of the surface of the scene in that location. SD is typically *not* constant across the scene even for a regular camera grid. The acquisition can have a variety of configurations, such as regular/irregular 2D or 3D camera grids or even a set of mobile cameras at random positions and orientations.

In addition, the *depth estimation* process provides an estimation of depth (e.g. depth map) to improve rendering. This could be obtained by specialized hardware, such as depth cameras, or computed from the images obtained by multiple cameras. In either case, the depth estimation will have some error.

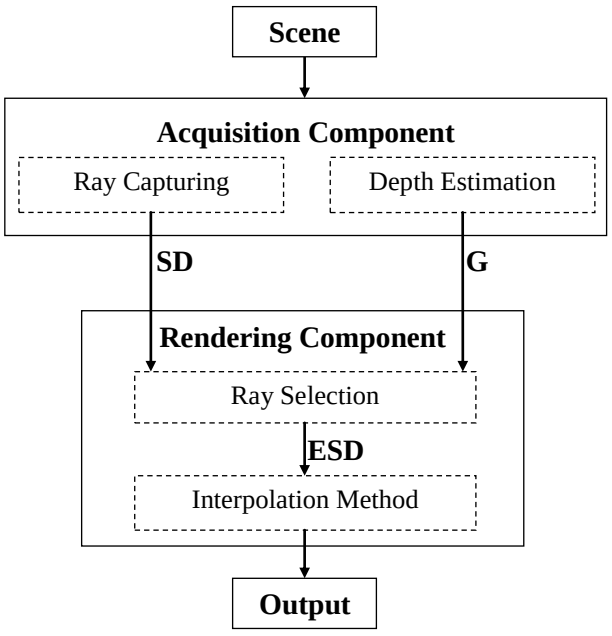


Figure 3.1. The schematic diagram of a typical LF-based FVV system that utilizes scene geometric information

The *rendering component* is responsible for estimating/reconstructing an unknown ray r from the acquired rays and depth information. In general, this component is composed of two processes: (i) the *ray selection process* that chooses a subset of acquired rays, purported to be in the vicinity of r, for the purpose of interpolation; and (ii) the *interpolation process* that provides an estimate of r from these selected rays.

The *ray selection process*, in particular, is often prone to error. For example, imperfect knowledge of depth may cause this process to miss some neighboring rays and choose others that are indeed sub-optimal (with respect to proximity to r) for interpolation. Also, constraints on computational load (imposed due to, say, real-time rendering requirements) may necessitate this process to select only a subset of neighboring rays, less than what is available. In both cases, there is some loss of information and the output of this process represents an *effective sampling density* (ESD) which is *lower* than the SD obtained by the acquisition component and distortion is inevitably introduced in the reconstructed video. ESD is defined as the number of rays per unit area of the scene that have been captured by *acquisition* component and chosen by *ray selection process* to be employed in the rendering. Clearly, ESD $\leq$ SD with equality holding only for a perfect rendering component with complete knowledge of the scene and sufficient computational resources. Not surprisingly, ESD is the true indicator of output quality, *not* SD, and its key advantage is that it provides an analytically tractable way for evaluating the influence of the imperfections of *both* acquisition and rendering components. In this thesis, we analytically derive ESD for a number of well-known rendering algorithms for a regular grid of cameras and verify that ESD is an effective indicator of output quality. Extension to irregular camera configuration is also discussed in chapter 7 of the thesis.

3.3 Effective Sampling Density (ESD)

To illustrate the concept of ESD, consider the case shown in Figure 3.2, where the actual surface is at depth d and the unknown ray r intercepts the object at point p. There are four rays r_1, r_2, r_3, and r_4 captured by the cameras that lie within the interpolation neighbourhood of p, shown as a solid rectangle, and could be used to estimate r. However, since the estimation of depth is in error by Δd, the algorithm selects four other rays, r_1', r_2', r_3', and r_4' as the closest candidates for interpolation. As a result, the sampling density has been effectively reduced from $4/A$ to $4/A'$, where A and A' are the areas of solid and dashed rectangles in the Figure respectively. In addition, the rendering algorithm may not be able to use all available rays for interpolation due to computational constraint. In typical Depth Image Based Rendering (DIBR), for example, only a single ray (supposedly, the closest) may be utilized even though more rays have been acquired.

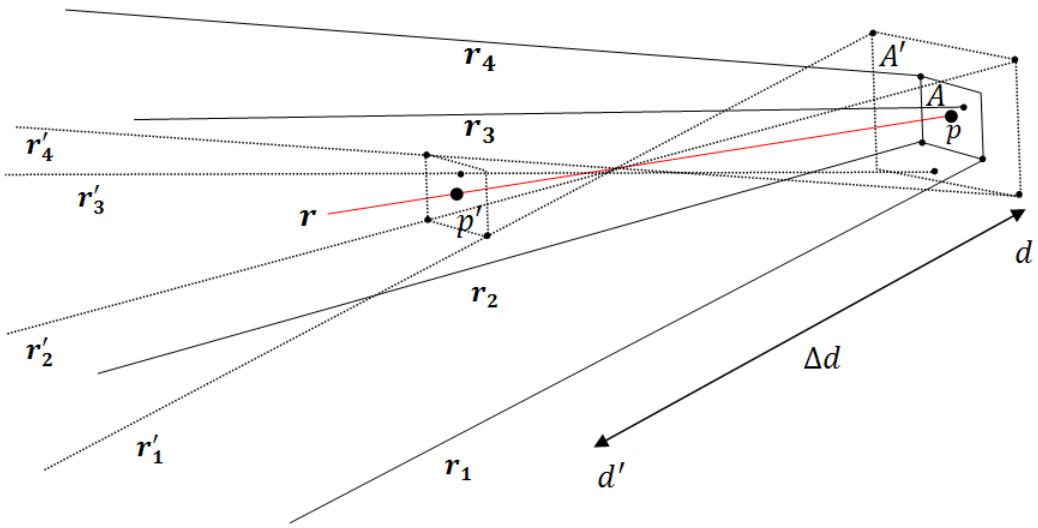


Figure 3.2. Selection of rays in an LF rendering and the concept of ESD

Let Θ be the set of all known rays captured by the cameras, that is, the samples of the scene obtained during the LF acquisition phase. A rendering method has a *ray selection mechanism* M to choose a subset ω of rays from Θ, purported to be surrounding the unknown ray r. Subsequently, an *interpolation function* F is applied to ω to estimate the value of the unknown ray r. Assume that r intersects with the scene at point p at depth d. A is an imaginary convex hull area around p which intersects with all the rays in ω at depth d. The size of A would depend on the choice of ω, hence, the rendering method. Since each squared pixel in an image sensor integrates light rays coming within a squared-based pyramid extending towards the scene. The cut area (square) of this pyramid at distance d is roughly $ld \times ld$, where l is the size of the pixel determined by camera resolution. Therefore, the minimum length of the sides of A is ld, which is referred to as the system resolution in this thesis.

There are usually more rays from Θ passing through A, but are not selected by the rendering process. However, using them could potentially enhance the interpolation and the rendering quality. Let all the captured rays passing through A be denoted by Ω. *Selection mechanism* M chooses a subset of rays ω from Ω to estimate the unknown ray r. Clearly:

$$\omega \subseteq \Omega \subseteq \theta \tag{3.1}$$

Both M and F may or may not use some kind of scene geometric information G such as focusing depth (average depth of the scene computed from automatic focusing algorithms or camera distance sensors) or depth map. Mathematically, the LF rendering can be formulated as (3.2) and (3.3) below. Different LF rendering methods differ in their respective M and F functions and their auxiliary information G.

$$\omega = M(\Theta, G) \tag{3.2}$$

$$r = F(\omega, G) \tag{3.3}$$

Sampling Density (SD) is defined as the number of acquired rays per unit area of the scene space (number of rays in Ω divided by the area A) and Effective Sampling Density (ESD)

as the number of rays per unit area of the scene that has been acquired and is employed during the interpolation process to estimate the unknown ray (number of rays in ω divided by the area A), that is,

$$\mathrm{SD} = \frac{|\Omega|}{A} \tag{3.4}$$

$$\mathrm{ESD} = \frac{|\omega|}{A} = \frac{|M(\Theta,G)|}{A} \tag{3.5}$$

where $|\Omega|$ and $|\omega|$ are the number of elements in Ω and ω respectively. A is the area of interpolation convex hull, and can be calculated by deriving the line equations for the boundary rays β_i's and finding the vertexes of convex hull A at depth d. Figure 3.3 shows this process for a simplified 2D light field generated by applying a 2D projection to a 3D light field with 2 planes parameterization, that is, camera plane uv and image plane st over (u, s). However, the approach can be generalized to 3D light field and any parameterization models [57].

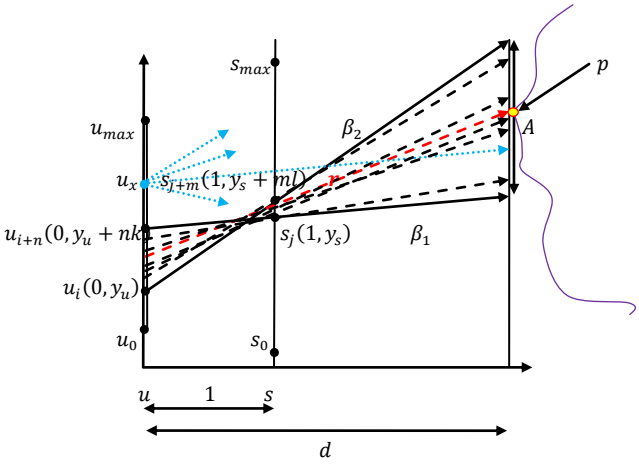


Figure 3.3. ESD calculation for a simplified 2D light field system

Figure 3.3 illustrates that r, the unknown ray, intersects the imaginary scene on point p and A is the area of interpolation surrounding p. The set Ω is determined by all the rays in the system that flow through A and potentially could be employed in a rendering method to estimate

r. However, practically, only ω, a subset of them would be used to interpolate the unknown ray r. Let us assume that rays in ω are surrounded by the boundary rays β_1 and β_2. The rays in ω are selected by the selection method M of a rendering method and potentially auxiliary information G and are bounded by $n+1$ cameras in u (u_i to u_{i+n}) and $m+1$ pixels in s (s_j to s_{j+m}). As it can be seen, A is at least a function of k, l , n, m and d, where k is the distance between the cameras, l is the pixel length, n and m are the number of cameras and pixels bounded by boundary rays respectively, and d is the depth of p. The rays intersect with A from these $n+1$ cameras are the rays employed by rendering method, i.e., ω set. However, as it is shown in Figure 3.3, there are more than $n+1$ cameras in the grid, (in addition to cameras bounded between u_i to u_{i+n}) that are able to see area A. u_x is shown as an example of these cameras. The rays from these cameras to A, make the difference between Ω and ω sets.

In 3D light field, boundary rays create a vicinity convex hull A (the area of sampling/interpolation) around the intersection point p. The area of A in 3D light field is the square of the length shown in the 2D model above if the configuration of the system is symmetrical or the product of vertical and horizontal sampling lengths for an asymmetrical system.

SD defined in (3.4) provides the upper bound of ESD. SD is a parameter to quantify the acquisition component. ESD is to quantify the combined effect of acquisition and rendering. Since ω ⊆ Ω in any point of the scene space, ESD is less or at best equal to SD. For a given LF acquisition configuration, it is possible to calculate SD on any point over the scene space. Importantly, SD is generally not uniform across the field of view of the LF system, even when a regular camera grid is used in capturing. Figure 3.4.a shows the SD contour maps at different depths, $d = 30m$, $60m$, and $90m$, for a regular camera grid of $30x30$ with $k = 2m$, camera field of view of 30°, image resolution of $100x100$ pixels, i.e., $l = 0.53cm$ in image plane st, and ideal area $A = (ld)^2$, i.e., LF system resolution. Figure 3.4.b shows a 2D slice, i.e., a 2D

light field of the same system with one row of 30 cameras and one row of 100 pixels for all d within $[2m, 100m]$.

Notice from Figure 3.4 that SD of a point in the scene decreases when it moves further from the camera grid and away from the centre of the scene, hence, rendering quality would possibly decrease as well. To have a desired rendering quality from any viewpoint for any proportion of the scene, the acquisition component should be designed in a manner to produce the required SD for every point of the scene. The rendering method can then be designed in such a way to provide optimum ESD at each point of the scene from this SD.

For the purpose of this thesis, it is assumed that the scene is located in the centre and not the boundary areas and the camera grid density and size is enough to provide sufficient SD at each point of the scene, that is, always higher than the required ESD for that point.

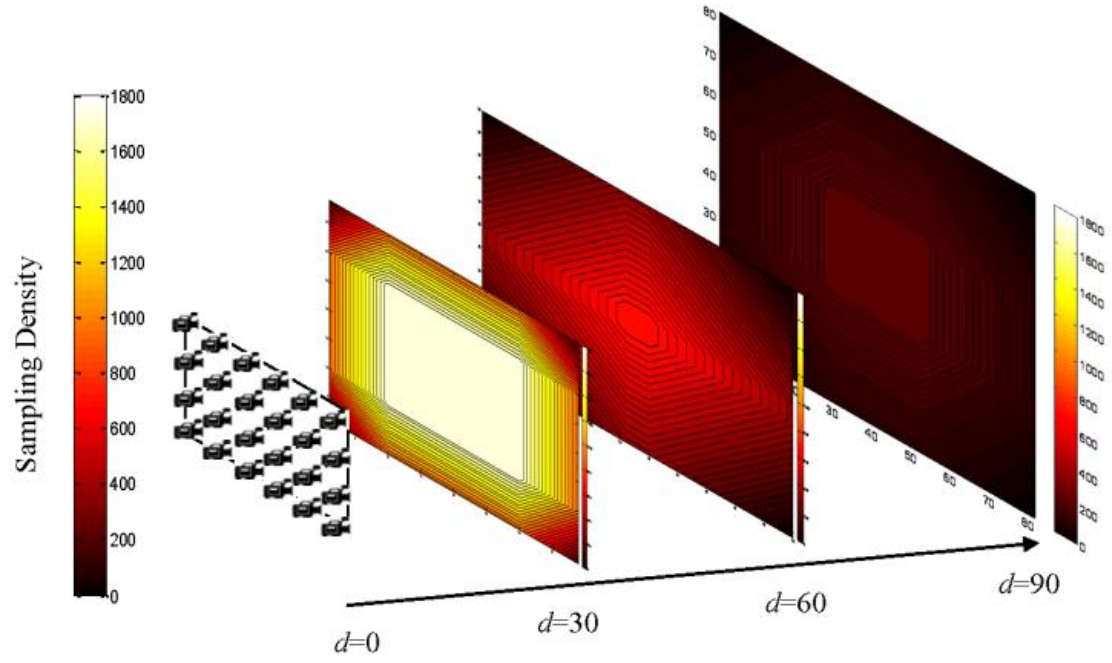
Based on the discussion above, it can be speculated that the output quality of an arbitrary view is determined by three key factors as outlined below:

1. The ESD in each area A, the vicinity of the unknown rays that compose the view, which could be mathematically derived for a given acquisition configuration and rendering method.

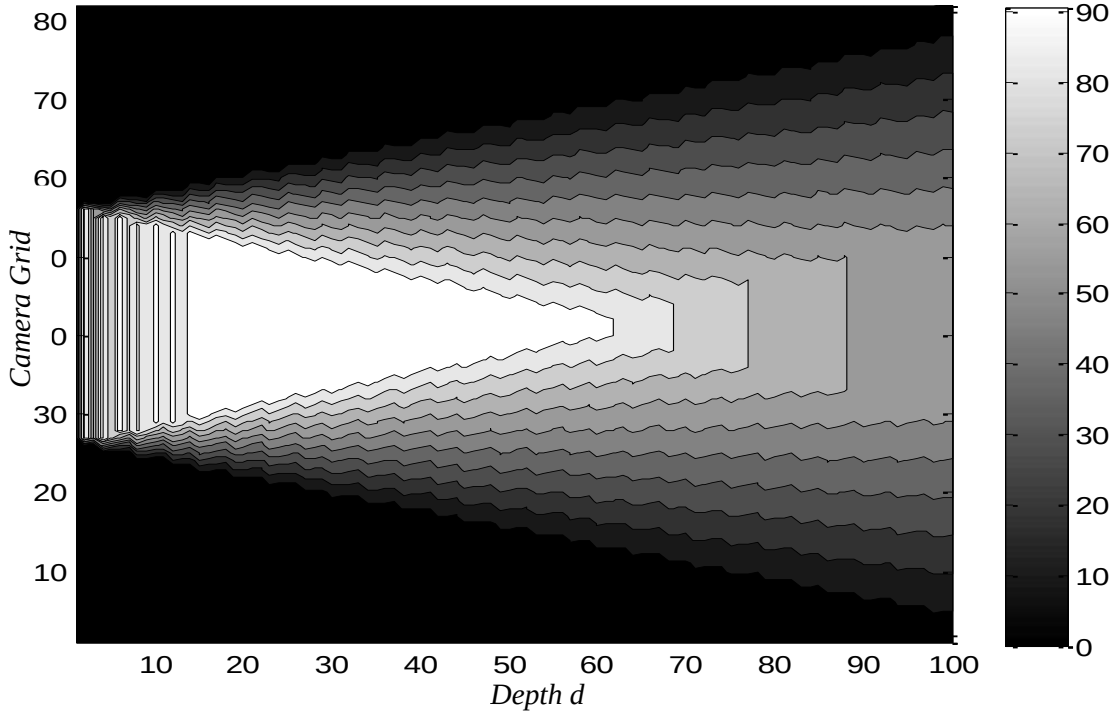
2. The scene complexity in each area A, which could be measured in terms of its spatial frequency components.

3. The accuracy and effectiveness of the interpolation function F employed for the estimation of the unknown rays.

In particular, for a fixed scene complexity and a given interpolation algorithm, ESD can objectively determine the quality of the LF rendering using equations (3.2) and (3.3) for a given LF acquisition configuration and, hence, can be potentially used as an objective indicator for the quality of video generated from a corresponding LF-based FVV system.



(a)



(b)

Figure 3.4. a) SD contour maps at different depths in 3D; b) SD contour map in 2D

It should be noted that ESD works in *under-sampled regime* and signal reconstruction accuracy cannot be improved indefinitely by increasing ESD [94]. For high values of ESD, the number of samples may reach the Nyquist density and further increase would not lead to any quality improvement. The maximum quality of the reconstructed signal is determined by many factors including, for instance, error in depth estimation.

3.4 ESD Analysis of LF Rendering Methods

According to its definition, ESD for an LF rendering method can be calculated from the ray selection mechanism M of that rendering method to choose ω and the LF acquisition parameters such as the camera configuration. In this section, we analyse ESD associated with a number of commonly used LF rendering methods. Without loss of generality, 2-plane parameterization scheme is adopted for the analysis. This can be extended to other parameterization schemes if needed. The selected LF rendering methods are divided into two main classes according to whether the methods utilize any depth information.

3.4.1 Rendering Methods without the Depth Information

All the LF rendering methods without the depth information (hereafter referred to as *blind* methods) can be categorized into four main groups based on their ray selection mechanism M, that is, methods using the Nearest Neighbourhood estimation (NN), methods using a 2D interpolation in camera plane (UV), methods using a 2D interpolation in image plane (ST) and methods using a full 4D interpolation in both camera and image planes (UVST). For the interpolation function F, a bilinear interpolation is often used for the 2D interpolation and a quadrilinear interpolation for the 4D interpolation. However, when $|\omega| > 4$ for UV and ST and when $|\omega| > 16$ for UVST, the convex hull A may not be a grid anymore and other types of 2D

and 4D interpolation could be employed as function F. Considering the regular geometry of the cameras shown in Figure 3.3, analytical form of ESD for these rendering algorithms can be derived. Table 3.1 summarizes the ESD derivation for the NN, ST, UV, and UVST methods where $|\omega| = 4$ for UV and ST and $|\omega| = 16$ for UVST. For each one of these rendering methods, the details of selection mechanism M and interpolation function F are given in the second and third columns. The fourth column summarizes the sampling /interpolation length A. Notice that A is a segment in the chosen 2D LF system whereas it is an area in 3D. The fifth column lists the corresponding ESD.

Table 3.1

ESD for the LF rendering methods without using depth information

Rendering method	*Selection Mechanism M*	*Interpolation Function F*	*Sampling length A in 2D LF*	ESD *for symmetric 3D light field*
NN	Select the nearest ray in 4D space, $\|\omega\| = 1$	No interpolation, neighbourhood estimation	$A_{NN} = (\frac{l+k}{2})d - \frac{k}{2}$	$ESD_{NN} = \frac{1}{{A_{NN}}^2}$
ST	Select 4 or more rays from the neighbourhood pixels in st plane to the nearest camera in uv plane, $\|\omega\| \geq 4$	Any type of 2D interpolation, e.g., bilinear interpolation for 2D grid selection of rays	$A_{ST} = (l + \frac{k}{2})d - \frac{k}{2}$	$ESD_{ST} = \frac{4}{{A_{ST}}^2}$
UV	Select 4 or more rays from the neighbourhood cameras in uv plane to the nearest pixel in the st plane, $\|\omega\| \geq 4$	Any type of 2D interpolation, e.g., bilinear interpolation for 2D grid selection of rays	$A_{UV} = (k + \frac{l}{2})d - k$	$ESD_{UV} = \frac{4}{{A_{UV}}^2}$
UVST	Select 16 or more rays from four neighbourhood cameras in uv to four neighbourhood pixels in st, $\|\omega\| \geq 16$	Any type of 4D interpolation, e.g., quadrilinear interpolation for grid selection of rays	$A_{UVST} = (l + k)d - k$	$ESD_{UVST} = \frac{16}{{A_{UVST}}^2}$

With the analytical ESD forms shown in Table 3.1, it is possible to objectively compare these rendering methods in terms of the output quality for the same acquisition. The higher the ESD is, the higher output quality is expected. Since when $|\omega|$ is fixed, ESD is a function of the sampling/interpolation area A. The ratio γ of A between two rendering methods is used as a factor for comparison.

Table 3.2 summarizes the comparison. The first column shows a pair of rendering methods to be compared, the second column is the ratio γ, the third column gives the relationship between the corresponding ESDs, the fourth column is the minimum value of γ for each pair. Specifically, three particular scenarios are analysed and their corresponding γ are shown in the fifth column of Table 3.2.

Scenario one: $d \to \infty$ and $k \gg l$, which represents a typical low density camera grid and a scene that is very far from the grid. In this case, the analysis shows that, $4\mathrm{ESD}_{\mathrm{NN}} < 4\mathrm{ESD}_{\mathrm{UV}} < \mathrm{ESD}_{\mathrm{ST}} < ESD_{\mathrm{UVST}}$. In other words, UVST has the highest ESD and is expected to produce the best video quality. NN has the lowest ESD and therefore would generate the lowest quality output.

Scenario two: $d \to \infty$ and $k \cong l$, a hypothetical very high density camera grid for a scene that is very far from the grid, the analysis shows that, $1.7\mathrm{ESD}_{\mathrm{NN}} < \mathrm{ESD}_{\mathrm{UV}} < \mathrm{ESD}_{\mathrm{ST}}$, $4\mathrm{ESD}_{\mathrm{NN}} < ESD_{\mathrm{UVST}}$, and $2.2\mathrm{ESD}_{\mathrm{UV}} < 2.2\mathrm{ESD}_{\mathrm{ST}} < ESD_{\mathrm{UVST}}$. This shows the same order as first scenario, but both NN and UV methods work much better in comparison with ST, though still UVST has the best performance.

Scenario three: $d \cong 1$, a hypothetical scene very close to the image plane, the analysis shows that, $4\mathrm{ESD}_{\mathrm{NN}} < 4\mathrm{ESD}_{\mathrm{ST}} < \mathrm{ESD}_{\mathrm{UV}} < ESD_{\mathrm{UVST}}$. This shows that UV outperforms ST in such a scenario with ESD more than four times higher than ST. Hence, for a scene close to the grid, UV is a better choice for rendering method compared with ST, which is intuitively appealing.

Similar analysis can be applied to other scenarios, which can offer a choice of rendering algorithms for a given acquisition system. In addition, for a given camera system, the analytic expression of ESD can be derived from the geometry of the camera system and used, in turn, to estimate, tune or optimize system parameters.

Table 3.2

Comparison of ESD of the LF rendering methods without using depth information

Methods	*Sampling length comparison*	ESD *comparison*	γ *(the ratio of* ESD*'s)*	γ *Analysis*
NN vs. ST	$A_{NN}.\gamma > A_{ST}$	$ESD_{NN}.\frac{4}{\gamma^2} < ESD_{ST}$	$\gamma > 1 + \frac{ld}{(l+k)d-k}$	$d \to \infty \; and \; k \gg l \Rightarrow \gamma = 1$ $d \to \infty \; and \; k \cong l \Rightarrow \gamma = 1.5$ $d \cong 1 \Rightarrow \gamma = 2$
NN vs. UV	$A_{NN}.\gamma > A_{UV}$	$ESD_{NN}.\frac{4}{\gamma^2} < ESD_{UV}$	$\gamma > 1 + \frac{kd-k}{(l+k)d-k}$	$d \to \infty \; and \; k \gg l \Rightarrow \gamma = 2$ $d \to \infty \; and \; k \cong l \Rightarrow \gamma = 1.5$ $d \cong 1 \Rightarrow \gamma = 1$
NN vs. UVST	$A_{NN}.\gamma > A_{UVST}$	$ESD_{NN}.\frac{16}{\gamma^2} < ESD_{UVST}$	$\gamma > 2$	$\gamma > 2$
ST vs. UVST	$A_{ST}.\gamma > A_{UVST}$	$ESD_{ST}.\frac{4}{\gamma^2} < ESD_{UVST}$	$\gamma > 1 + \frac{d-1}{(\frac{2l}{k}+1)d-1}$	$d \to \infty \; and \; k \gg l \Rightarrow \gamma = 2$ $d \to \infty \; and \; k \cong l \Rightarrow \gamma = 1.33$ $d \cong 1 \Rightarrow \gamma = 1$
UV vs. UVST	$A_{UV}.\gamma > A_{UVST}$	$ESD_{UV}.\frac{4}{\gamma^2} < ESD_{UVST}$	$\gamma > 1 + \frac{ld}{(l+2k)d-2k}$	$d \to \infty \; and \; k \gg l \Rightarrow \gamma = 1$ $d \to \infty \; and \; k \cong l \Rightarrow \gamma = 1.33$ $d \cong 1 \Rightarrow \gamma = 2$
ST vs. UV	$A_{UV} > \gamma.A_{ST}$	$ESD_{UV}.\gamma^2 < ESD_{ST}$	$\gamma < 1 + \frac{(k-l)d-k}{(2l+k)d-k}$	$d \to \infty \; and \; k \gg l \Rightarrow \gamma = 2$ $d \to \infty \; and \; k \cong l \Rightarrow \gamma = 1$ $d \cong 1 \Rightarrow \gamma = 0.5$

3.4.2 Rendering Methods with the Depth Information

Utilization of depth information G in rendering can compensate to some extent for insufficient number of samples acquired in an *under-sampling* situation [17]. It can make the ray selection mechanism M more effective compared with blind rendering methods. The amount of depth information G could vary from a crude estimate, such as the focusing depth, to the full depth map or even full 3D geometric model of the scene. A mechanism M in this case may choose a number of rays intersecting the scene in the vicinity of point p at depth d. A rendering method whose interpolation function F is a 2D interpolation over uv plane and utilizes the focusing depth is referred to as UV-D (**UV**+**D**epth) and the one with a full depth map is referred to as UV-DM (**UV**+**D**epth **M**ap). By extending the selection mechanism M and interpolation function F to a full 4D interpolation over both uv and st planes, the rendering methods are referred to as UVST-D (**UVST**+**D**epth) and UVST-DM (**UVST**+**D**epth **M**ap) respectively, the former using focusing depth only. Many LF rendering methods with depth information can be mathematically expressed in the form of one of these 4 groups. The simplest forms of these methods are with $|\omega| = 4$ and bilinear interpolation as F for UV-D and UV-DM and $|\omega| = 16$ and quadrilinear interpolation as F for UVST-D and UVST-DM. The unknown ray r can be estimated as $r = F(\omega, G) = F(M(\Theta, G), G)$.

Figure 3.5 illustrates the rendering methods with depth information. If the exact depth d at point p, the intersection of unknown ray r with the scene, is known, applying a back projection could easily find a subset of known rays Ω intersecting the scene at the vicinity of p. Subsequently, an adequate subset ω of these rays can be selected by mechanism M to be employed in interpolation method F. Mechanism M selects rays required for a neighbourhood estimation or bilinear interpolation over image plane st, if rays intersecting the scene at the vicinity of p don't pass through known pixel values.

However, in practice, the estimated depth of p is unknown and the estimated depth has an error Δd. Again, the same procedure can be employed to form the Ω and ω sets by

employing the estimated depth of p. Though, this time the depth estimation error Δd, makes the rays intersect in an imaginary point p' in the space and going through the vicinity of area A on the scene instead of intersecting with the exact point p on the scene surface. Subsequently, this estimation error Δd would result in reduction of ESD and a lower rendering quality.

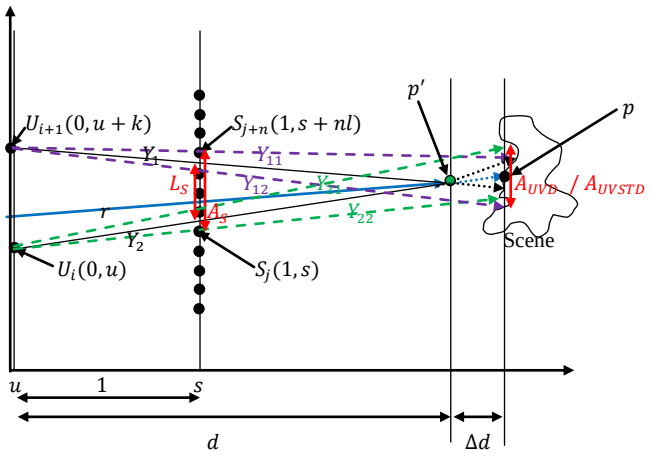


Figure 3.5. UV-D/UV-DM and UVST-D/UVST-DM rendering methods with depth error Δd

Note that forming Ω by applying a back projection to the imaginary point p is a simplification for effective estimation of Ω, in fact Ω has more rays passing through A but are not passing through vicinity of p. Hence, to compute the exact Ω, back projection should be applied to the vertexes of A and not p to find all the rays passing through A. The SD calculated from the simplified calculation of Ω is a lower bound estimate of the exact SD.

The size of area A depends on Δd and as Δd gets larger, it also increases. As the exact value of Δd error is not known, calculating the precise area of A is not possible. Usually only the upper bound of the error is known and therefore in this thesis, the worst-case scenario, i.e., largest A is computed in the LF analysis which corresponds to the lower bound of ESD.

Considering a simplified scenario in Figure 3.5, Y_1 and Y_2 are two immediate neighbour rays, intersecting with the desired ray r at depth d on object surface. If these two rays don't pass through the known s values in image plane, Y_1 from Y_{11} and Y_{12} and Y_2 from Y_{21} and Y_{22} can be

estimated. Finally, a bilinear interpolation in uv plane (or a linear interpolation over u in this 2D example) is applied to estimate r from Y_1 and Y_2.

In the simple 2D example demonstrated in Figure 3.5, ω includes only two samples for UV-D/UV-DM and four samples for UVST-D/UVST-DM. In fact, all the samples that intersect the object surface at point p in vicinity A at depth d can be employed in the rendering ($\omega = \Omega$) to achieve a better quality. Y_{12} and Y_{21} are boundary rays used for interpolation. In an ideal scenario, the depth estimation has no error, i.e., $\Delta d = 0$. It is obvious that in this case: $A_S = L_S + \frac{l}{2} + \frac{l}{2} = \frac{k(d-1)+ld}{d}$, $A_{UVD/UVDM} = ld$ and $A_{UVSTD/UVSTDM} = 2ld$. In a realistic scenario when $\Delta d > 0$, p is somewhere in the range of $d \pm \Delta d$, and the sampling area A would be increased as:

$$A = max[\,|Y_{11}(d+\Delta d) - Y_{22}(d+\Delta d)|, |Y_{12}(d+\Delta d) - Y_{21}(d+\Delta d)|\,] = l(d+\Delta d) + \frac{\Delta d.k}{d} \tag{3.6}$$

Using this approach, the difference between the rendering methods with focusing depth (UV-D/UVST-D) and the rendering methods with full depth map (UV-DM/UVST-DM) is in the scale of Δd. For focusing depth, a fixed depth is given for all points on the scene. This makes the depth estimation error, $\Delta d = \frac{object\ length}{2} + focusing\ depth\ estimation\ error$. When the full depth map of the scene is provided as G, the depth of each point p of the scene possibly with some estimation error Δd is known. This error is usually much less than the focusing depth error and makes the UV-DM/UVST-DM rendering more accurate than UV-D/UVST-D.

In UV-DM/UVST-DM, the scene depth information can be in the form of one or several separate depth maps on the camera plane or, in the best-case scenario, n depth maps for n cameras. For any unknown ray, either the closest depth camera information or an interpolation of several close depth points with the aim of reducing Δd , can be employed to estimate the depth of p.

3.4.2.1 General case of Rendering Methods with the Depth Information

Figure 3.6 demonstrates an LF rendering method with 2 plane parameterization using a depth map as the auxiliary information G. Ray r is the unknown ray that needs to be estimated for an arbitrary viewpoint reconstruction. r is assumed to intersect the scene on point p at depth d.

In Figure 3.6, seven rays from all rays intersecting imaginary p are selected by M, i.e., $|\omega| = 7$, assuming these rays pass through known pixel values or if neighbourhood estimation is used. In the case of bilinear interpolation, 28 rays are chosen by M to estimate these 7 rays. The chosen cameras in uv plane are bounded by a convex hull A'. It is easy to show that interpolation convex hull A is proportional to A'.

Finally a 2D interpolation F over convex hull A' on uv plane can be applied to estimate unknown ray r from the rays in ω. This rendering method with depth information is referred to as UV-DM when 2D interpolation is performed over neighbouring cameras in the uv plane and neighbourhood estimation, i.e., choosing the closest pixel in the st plane. The rendering method is called UVST-DM in the case of 2D interpolation over neighbouring cameras in the uv plane and bilinear interpolation over neighbouring pixels in the st plane.

In a simple form of UV-DM and UVST-DM, the rays in ω are selected in a way that A' becomes rectangular, i.e., 2D grid selection and therefore 2D interpolation over A' can be converted into a familiar bilinear interpolation.

Notice that all the existing LF rendering methods such as [9-14], in which depth map is utilized, are a special case of UV-DM and UVST-DM methods. The ESD for the UV-DM and UVST-DM demonstrated in Figure 3.6 can be derived as:

$$\mathrm{ESD}_{\mathrm{UVDM}} = \frac{|\omega|}{A} = \frac{|\omega|}{\frac{\Delta d}{d} A' + \mu(l(d+\Delta d), A')} \tag{3.7}$$

$$\mathrm{ESD}_{\mathrm{UVSTDM}} = \frac{|\omega|}{A} = \frac{|\omega|}{\frac{\Delta d}{d}A' + \mu(2l(d+\Delta d), A')} \tag{3.8}$$

where μ is a function to calculate the effect of pixel interpolation over st plane on the area A. A is mainly determined by A', but the pixel interpolation μ which is added to (3.7) and (3.8) also has small effect on A. The pixel interpolation over st even when $\Delta d = 0$ makes $A = (ld)^2$.

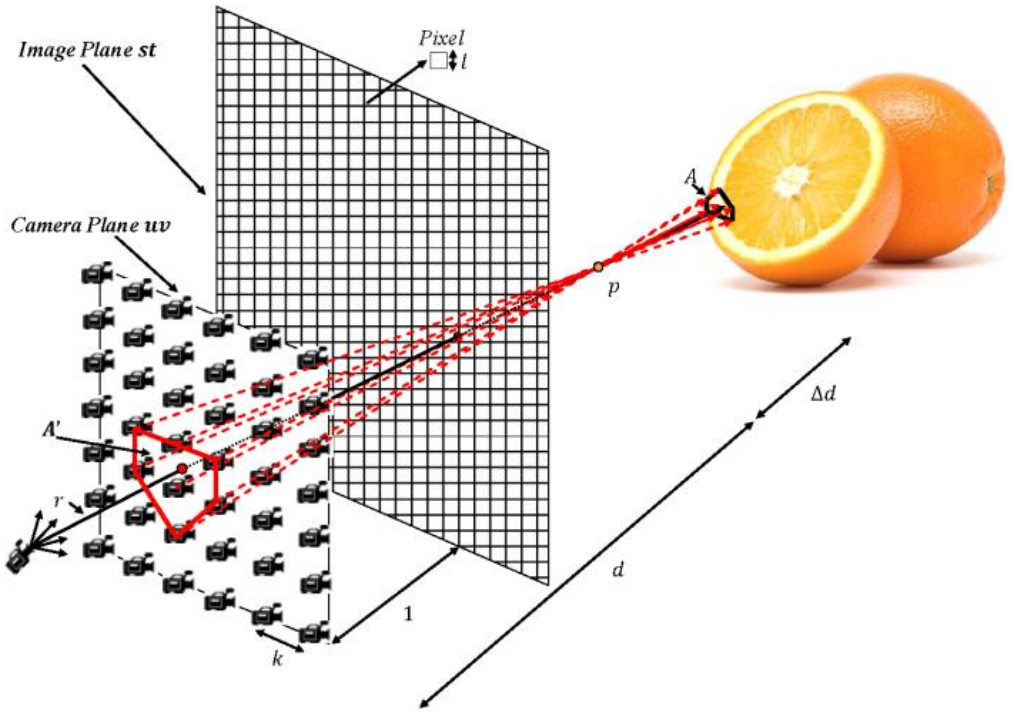


Figure 3.6. General light field rendering method using depth information (UV-DM /UVST-DM) with Δ*d* error in depth estimation

A simple form of UV-DM and UVST-DM can be formulated with employing a regular camera grid and 2D grid selection of rays, i.e., *A'* as a rectangular area with 4 and 16 samples in |ω| respectively. Subsequently (3.7) and (3.8) become:

$$\mathrm{ESD}_{\mathrm{UVDM}} = \frac{4}{\left(\frac{\Delta d.\, k}{d} + l(d + \Delta d)\right)^2} \tag{3.9}$$

$$\mathrm{ESD}_{\mathrm{UVSTDM}} = \frac{16}{\left(\frac{\Delta d.\, k}{d} + 2l(d + \Delta d)\right)^2} \tag{3.10}$$

where k is the distance between the two neighbouring cameras in the cameras grid and l is the length of the pixel in the image plane as illustrated in Figure 3.6. Note that the edge of A' rectangular is equal to k and that is how (3.9) and (3.10) are derived from (3.7) and (3.8).

It has to be pointed out that most existing rendering methods with depth information are equivalent to these simple versions of UV-DM and UVST-DM and choose only a very small subset of Ω, typically 4 or 16 rays, as ω. When the depth map is accurate, a small number of rays, say 4, would be sufficient, but for the case of less accurate depth maps, employing more rays in ω for interpolation could compensate for the adverse effect of errors in depth to some degree and improve the rendering quality since ESD is increased as can be seen from (3.7) and (3.8). This does not necessarily mean to increase the number of cameras, as there are already $|\Omega|$ rays passing through area A of the scene and potentially can be chosen as ω. These samples are already captured so if using more can result in rendering quality improvement, the added complexity of the rendering algorithm may be justifiable.

For the rest of this thesis the analysis is only carried out for UV-DM, which can easily be extended to UVST-DM. Consider the simple form of UV-DM described above (i.e., the rays in ω are selected in a way that A' becomes rectangular). Mathematically, a general representation of this simplified UV-DM rendering method is $r = \mathrm{UVDM}(d, \Delta d, k, l, |\omega|)$, where k is the distance between two neighbouring cameras and l is the length of the pixel, d and Δd are the estimated depth and its error and $|\omega|$ refers to the number of rays employed in interpolation.

3.4.2.2 ESD for $UVDM(d, \Delta d, k, l, |\omega|)$

By extending (3.9), the ESD could be calculated for $\mathrm{UVDM}(d, \Delta d, k, l, |\omega|)$ as follows:

$$\mathrm{ESD}_{\mathrm{UVDM}(d,\Delta d,k,l,|\omega|)} = \frac{|\omega|}{\left(l(d+\Delta d) + \frac{\Delta d.\, k}{d}\left(\sqrt{|\omega|} - 1\right)\right)^2} \tag{3.11}$$

Equation (3.11) assumes that the rays are chosen for interpolation symmetrically around the vertical and horizontal axes, such as $4x4$ samples. In this case, $\sqrt{|\omega|}$ would be an integer. For an asymmetrical choice of rays, (3.11) could be rewritten as follow:

$$\mathrm{ESD}_{\mathrm{UVDM}(d,\Delta d,k,l,|\omega|_{vertical},|\omega|_{horizontal})} = \frac{|\omega|_{vertical}.|\omega|_{horizontal}}{\left(l(d+\Delta d)+\frac{\Delta d.k}{d}(|\omega|_{vertical}-1)\right)\left(l(d+\Delta d)+\frac{\Delta d.k}{d}(|\omega|_{horizontal}-1)\right)} \tag{3.12}$$

$\mathrm{ESD}_{\mathrm{UVDM}(d,\Delta d,k,l,|\omega|)}$ predicts the rendering quality as shown in [30]. In the above expression, d is given by scene geometry and Δd is determined by the depth estimation method and cannot be altered by us. Changing the other three parameters could potentially improve the rendering quality. By assuming a given camera resolution, i.e., a fixed value of l, two other parameters can be tuned to compensate for the depth estimation error while maintaining the rendering quality. These parameters include k as a measure of density of cameras during acquisition and $|\omega|$ as an indicator of complexity of rendering method. ESD is proportional to $|\omega|$ and inversely proportional to k. It means higher camera density in camera grid (smaller k) and employing more rays for interpolation results in higher ESD. The influence of these two parameters on output quality and the ESD optimization based on them, individually and jointly will be discussed in detail in chapters 4, 5, and 6.

More discussion on ω, Ω, and Θ sets can be found in appendix I. Further discussion on SD calculation for a regular camera grid is presented as appendix II. In addition, appendix III demonstrates the details of ESD calculation.

3.4.2.3 ESD Analysis for Rendering Methods with Depth Information

ESD for the rendering methods using either focusing depth or depth maps can be analytically derived based on the geometry of the regular grid camera system as described in Figure 3.5 and Figure 3.6 and Equations (6) to (11). Table 3.3 summarizes derivation. The first column shows the rendering methods: UV-D and UVST-D methods that use focusing depth and UV-DM and UVST-DM that use depth maps, with $|\omega| = 4\ or\ 16$ and $|\omega| > 4\ or\ 16$. The second and third columns describe the selection mechanism M and interpolation function F respectively. The fourth and fifth column give the sampling/interpolation length A and ESD respectively.

Table 3.3

ESD for the LF rendering methods with depth information

Rendering method category	*Selection Mechanism M*	*Interpolation Function F*	*Sampling length in 2D*	ESD *for symmetric 3D light field*
UV-D $\|\omega\| = 4$	Select 4 rays sourcing from neighbourhood cameras in uv and intersecting with expected p	Neighbourhood estimation in st and 2D interpolation over uv	$A_{UVD} = l(d + \Delta d) + \frac{\Delta d.k}{d}$	$ESD_{UVD} = \frac{4}{{A_{UVD}}^2}$
UVST-D $\|\omega\| = 16$	Select 16 rays sourcing from neighbourhood cameras in uv, through known pixels in st and intersecting with expected p	4D interpolation over st and uv planes, e.g., quadlinear interpolation	$A_{UVSTD} = 2l(d + \Delta d) + \frac{\Delta d.k}{d}$	$ESD_{UVSTD} = \frac{4}{{A_{UVSTD}}^2}$

UV-DM $\|\omega\| = 4$	The same as UV-D but with more accurate depth estimation of p employing depth maps.	The same as UV-D	$A_{UVDM} = l(d + \Delta d) + \frac{\Delta d.k}{d}$	$ESD_{UVDM} = \frac{4}{A_{UVDM}^2}$
UVST-DM $\|\omega\| = 16$	The same as UVST-D but with more accurate depth estimation of p employing depth maps.	The same as UVST-D	$A_{UVSTDM} = 2l(d + \Delta d) + \frac{\Delta d.k}{d}$	$ESD_{UVSTDM} = \frac{16}{A_{UVSTDM}^2}$
UV-DM $\|\omega\| > 4$	Select $\|\omega\|$ rays sourcing from neighbourhood cameras in uv and intersecting with expected p	2D interpolation over chosen rays in ω and estimate each ray from closest known pixel in st	$A_{UVDM(d,\Delta d,k,l,\|\omega\|)} = l(d + \Delta d) + \frac{\Delta d.k}{d}(\sqrt{\|\omega\|} - 1)^*$	$ESD_{UVDM(d,\Delta d,k,l,\|\omega\|)} = \frac{\|\omega\|}{A_{UVDM(d,\Delta d,k,l,\|\omega\|)}^2}$
UVST-DM $\|\omega\| > 16$	Select $\|\omega\|$ rays sourcing from neighbourhood cameras in uv, through known pixels in st and intersecting with expected p	4D interpolation over chosen rays in ω in both uv and st planes	$A_{UVSTDM(d,\Delta d,k,l,\|\omega\|)} = 2l(d + \Delta d) + \frac{\Delta d.k}{d}(\sqrt{\|\omega\|} - 1)^*$	$ESD_{UVSTDM(d,\Delta d,k,l,\|\omega\|)} = \frac{\|\omega\|}{A_{UVSTDM(d,\Delta d,k,l,\|\omega\|)}^2}$

**This is calculated by assuming that chosen rays form a rectangular grid in uv plane for simplification*

Table 3.4 summarizes comparison of the ESD among UVST, UV-D, and UVST-D. It is clear from Table 3.3 that (UV-DM and UV-D) and (UVST-DM and UVST-D) have the same ESD, the difference between them being the scale of Δd, thus UV-DM and UVST-DM are omitted in Table 3.4. Similar to the analysis of the blind methods, ratio γ is used and two special scenarios, one with $d \to \infty, k \cong l \; and \; \Delta d \ll d$ and the other with $d \to \infty, k \gg l \; and \; \Delta d \ll d$ are analysed.

The second scenario corresponds to a typical FVV system where the scene is far from the camera grid, depth estimation error is small compared with the depth and there are a finite number of cameras.

The γ values allows us to compare the rendering methods with and without using depth information. Table 3.2 and Table 3.4 have shown that: $4\mathrm{ESD}_{\mathrm{NN}} < 4\mathrm{ESD}_{\mathrm{UV}} < \mathrm{ESD}_{\mathrm{ST}} < \mathrm{ESD}_{\mathrm{UVST}} \ll \mathrm{ESD}_{\mathrm{UVD/UVDM}} < \mathrm{ESD}_{\mathrm{UVSTD/UVSTDM}}$,i.e., for a given acquisition, the NN rendering method has the lowest ESD and hence results in the lowest video quality following by UV, ST, UVST, UV-D/UV-DM, and UVST-D/UVST-DM respectively. The experimental validation in next section will not only confirm this, but also show that ESD is highly correlated with PSNR.

Table 3.4

Comparison of the UVST, UV-D/UV-DM and UVST-D/UVST-DM methods

Methods	*Sampling length comparison*	ESD *comparison*	*γ Ratio*	*γ Analysis*
UVST vs. UV-D	$A_{UVST} > \gamma . A_{UVD}$	$ESD_{UVST}\frac{\gamma^2}{4} < ESD_{UVD}$	$\gamma < \frac{(k+l)d^2 - kd}{ld^2 + l\Delta\mathrm{dd} + \mathrm{k}\Delta\mathrm{d}}$	$d \to \infty , k \cong l \text{ and } \Delta d \ll d \Rightarrow \gamma = 2$ $d \to \infty , k \gg l \text{ and } \Delta d \ll d \Rightarrow \gamma = \infty$
UVST vs. UVST-D	$A_{UVST} > \gamma . A_{UVSTD}$	$ESD_{UVST}\gamma^2 < ESD_{UVSTD}$	$\gamma < \frac{(k+l)d^2 - kd}{2ld^2 + 2l\Delta\mathrm{dd} + \mathrm{k}\Delta\mathrm{d}}$	$d \to \infty , k \cong l \text{ and } \Delta d \ll d \Rightarrow \gamma = 1$ $d \to \infty , k \gg l \text{ and } \Delta d \ll d \Rightarrow \gamma = \infty$
UV-D vs. UVST-D	$A_{UVD} > \gamma . A_{UVSTD}$	$ESD_{UVD}4\gamma^2 < ESD_{UVSTD}$	$\gamma < 1 - \frac{ld^2 + l\Delta\mathrm{dd}}{2ld^2 + 2l\Delta\mathrm{dd} + \mathrm{k}\Delta\mathrm{d}}$	$d \to \infty \Rightarrow \gamma = \frac{1}{2}$

Equations shown in Table 3.3 and Table 3.4 can be used in LF based FVV system analysis and design. In addition to LF system evaluation and comparison, by knowing the upper bound of the depth estimation error, optimum system parameters such as camera density k, cameras resolution in terms of l , and rendering complexity in terms of number of rays employed in interpolation $|\omega|$ can be theoretically calculated. For example, in [16], we have used the above relationships to obtain the minimum camera density for capturing a scene. We will show in following three chapters of the thesis (chapter 4, 5, and 6) that how ESD can be

used to optimize the acquisition and rendering parameters of an LF system individually and jointly [95] for a target output video quality.

3.5 Theoretical and Simulation Results

To verify the effectiveness of ESD as an objective quality indicator of an LF-based FVV system, a computer simulation system employing a 3D engine has been developed to generate the ground truth data [96]. The details of the simulation model will be discussed later in appendix V . The system takes a 3D model of a scene and generates the sampling of the scene (reference images) for a given configuration of cameras. For any virtual views to be reconstructed by an LF rendering method, the system generates its ground truth image as a reference for quantitative comparison. Figure 3.7 illustrates a simulated regular-camera grid for acquisition. Virtual views were randomly generated as the ground truth and used to evaluate the performance of ESD as a distortion indicator.

In addition, since 3D models were used to represent the scene, a full precise depth map was available for rendering. Controlled amount of depth map error is simulated and added to the depth map in order to evaluate ESD when inaccurate depth is employed in the rendering. In the following, details on the depth error model and experimental settings are presented.

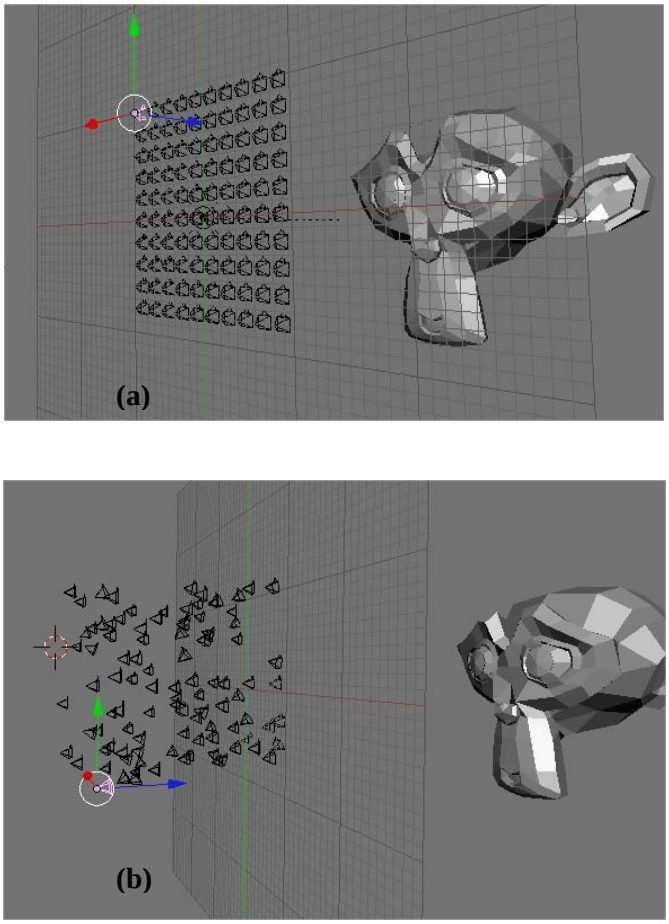


Figure 3.7. a) A simulated regular camera grid; b) Random virtual viewpoints.

3.5.1 Depth Error Model

There are two commonly used approaches to obtain depth information for FVV systems [97]: triangularization-based through either stereoscopic vision or structure light, and time-of-flight (ToF) based. When depth is estimated using the former approach, the error Δd is normally distributed whose standard deviation is proportional to the square of distance d^2, i.e. $\Delta d \approx \tau . d^2$, where τ depends on the system parameters [98]. For ToF, the error tends to be approximated coarsely as $\Delta d \approx \tau . d$ [99]. The linear model is adopted for the experimental validation in this chapter.

In the experiments, the exact depth map is known from the simulator. Based on the prescribed depth estimation error, for each pixel of the exact depth map, a random error with

normal distribution and standard deviation of $\Delta d = \tau . d$ is introduced to create a noisy depth map with average of $\tau\%$ error.

3.5.2 ESD of Scenes

The ESD equations summarized in Table I and Table III are all for a small vicinity of scene around a given point p. Clearly, ESD varies over the scene, depending on the depth. On the other hand, the overall distortion of output in addition to ESD is also scene dependent. Estimation of overall distortion for a given scene requires integration of ESD over the entire scene and at each point considering the scene texture complexity. In this section, an approximation is adopted by using the average depth of the scene. This allows analysing acquisition configurations or rendering methods based on ESD independently of the scene complexity. To compare acquisition configurations and rendering methods an $\overline{\mathrm{ESD}}$ for each configuration/method is calculated for comparison using an average depth of the scene $\bar{d}$ with an average $\overline{\Delta d}$ of absolute depth error.

3.5.3 Simulation Settings

For the experiments reported in this section, the LF engine is customized for the eight LF rendering methods: NN, UV, ST, UVST, UV-D, UVST-D, UV-DM and UVST-DM with $|\omega|$ = 1, 4, 4, 16, 4, 16, 4 and 16 respectively with default rectangular grid ray selection for M and bilinear and quadrilinear interpolations for F.

To assess the effect of scene complexity on output distortion, four 3D models, a "*room*", a "*chess board*", "*blender monkey*", and "*Stanford bunny*", as shown in Figure 3.8, were selected, where the complexity decreases in this order. In the simulation, the centre of the 3D model was placed at $d = 10m$ by default, if depth is not given in the experiment. A $16x16$

regular camera grid were placed for acquisition and the image resolution was originally set to $1024x768$ pixels, i.e., $l = 0.05cm$. However, for experiments reported in Figure 3.11, to evaluate the effect of the 3D model depth in output PSNR, $\bar{d}$ is changed between $[10m, 50m]$, in Figure 3.17 to evaluate the effect of the camera grid density in output PSNR, k is changed between $[0.1m, 0.9m]$, and in Figure 3.18 to evaluate the effect of the reference cameras resolution on output PSNR l is changed between $[0.02cm, 0.1cm]$, to analyse the effects of these factors on the output distortion.

Please note that the term pixel size in the following experiments refers to l, the projected pixel size on image plane st at depth $d = 1$. Hence, $l = 0.02cm$ on st plane corresponds to a real pixel size equal to $4.8x10^{-4}cm$ for a typical $1/2"$ camera sensor or capturing resolution of $2560x1920$. With the same assumptions, $l = 0.5cm$ corresponds to capturing resolution of $1024x768$ and $l = 0.1cm$ to resolution of $512x384$.

For each 3D model, 1000 random virtual cameras at different distances from the scene were generated and average PSNR between the rendering images and the ground truth was calculated for comparison.

Figure 3.8 shows the four 3D models used in the simulation.

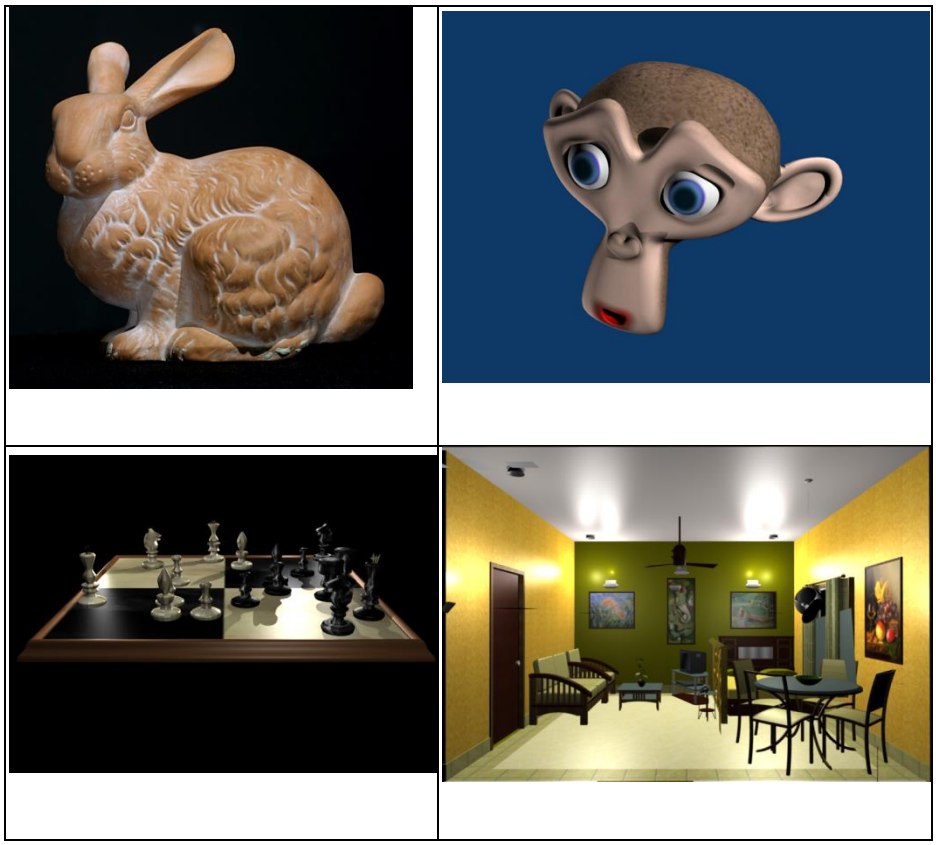


Figure 3.8. Four 3D scenes chosen for experimental validation

In the following, the theoretical expectations in terms of calculated $\overline{\mathrm{ESD}}$ and the actual measurement of output video distortion in PSNR are reported and compared for different rendering methods and different acquisition configurations.

3.5.4 Results on Rendering Methods

3.5.4.1 Theoretical Expectation

Figure 3.9 shows the ESD for the above-mentioned LF rendering methods in addition to the ideal rendering ($\Delta d = 0$) where $k = 0.4m$, $l = 0.05cm$, $d \in [10m, 50m]$, the object length is $5m$ and $\Delta d = 0.1d$ i.e., ten percent error in depth estimation. The ideal case is when there is no error in the depth map and refers to the maximum value for ESD at depth d. The

vertical axis is logarithmic. For UV-D and UVST-D the actual error is $\frac{object\ length}{2} + \Delta d$, which in this example is equal to $2.5m + 0.1d$.

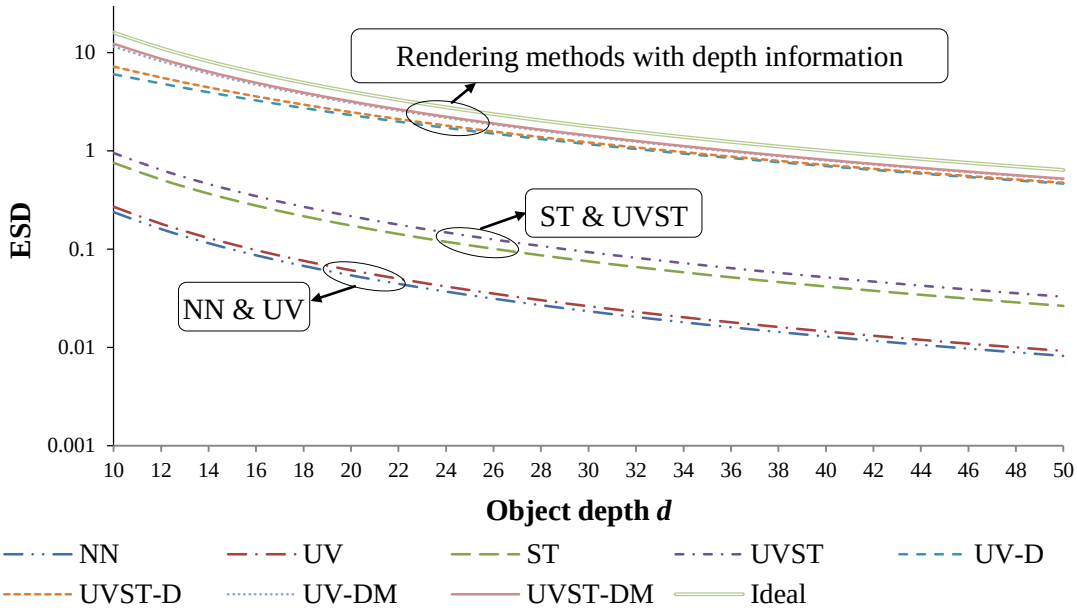


Figure 3.9. Theoretical $\overline{ESD}$ for different LF rendering methods based on object depth $\bar{d}$ for $k = 0.4m$ and $l = 0.05cm$ (i.e., camera resolution of $1024x768$)

It can be seen from Figure 3.9 that, for all depths, the expected relative relationship of ESD among the eight LF rendering methods is maintained. A quadrilinear interpolation over UVST makes UVST-D and UVST-DM perform slightly better than their corresponding UV-D and UV-DM, especially for small d. For large depths, UV-D/UVST-D performance approaches that of UV-DM/UVST-DM, because the object length is small compared to depth error in this case.

Figure 3.10 demonstrates a bar chart of theoretical ESD values for different rendering methods for $k = 0.4m$, $l = 0.05cm$, for a point p with $d = 10m$ and $\Delta d = 1m$.

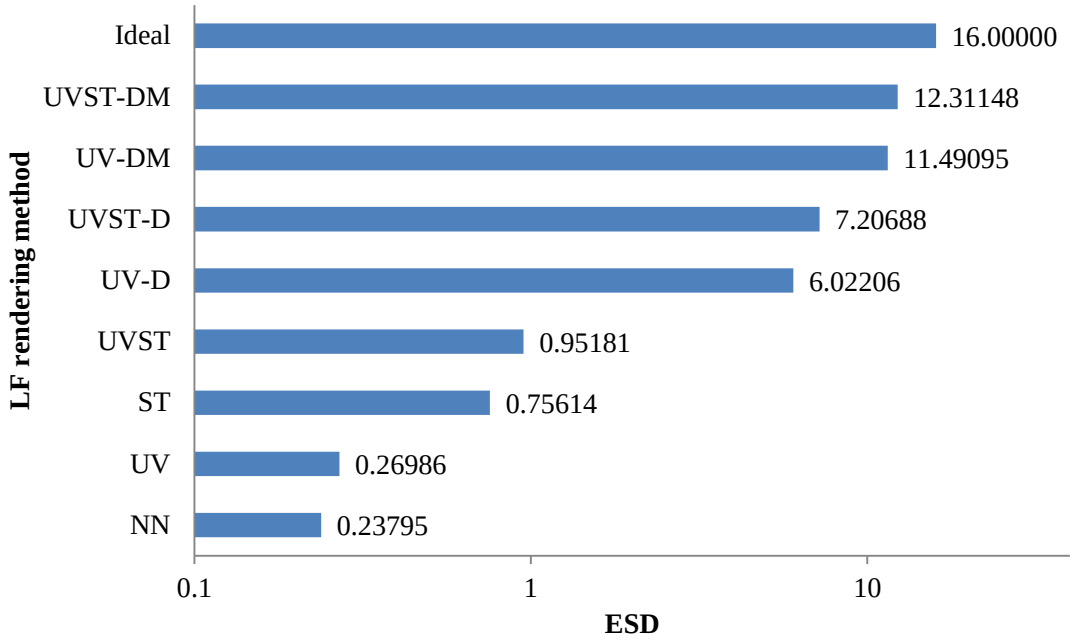


Figure 3.10. Theoretical $\overline{\mathrm{ESD}}$ for different rendering methods for $k = 0.4m, l = 0.05cm$, $\bar{d} = 10m$, and $\overline{\Delta d} = 1m$

3.5.4.2 Simulation Results

Figure 3.11 shows the simulated results, where the object depth d is changed from $10m$ to $50m$ with steps of $5m$ to analyze the effect of d on rendering output distortion in PSNR for different rendering methods. The acquisition parameters are: $k = 0.4m$ and $l = 0.05cm$ (i.e., camera resolution of $1024x768$). Notice that all the parameters for camera configuration and rendering algorithm were set the same as those used to obtain the theoretical results shown in Figure 3.9. 10% depth error was added in the experiments. Figure 3.11 shows the average results calculated from 288,000 experiments for 9 depths, 8 rendering methods, four 3D models and 1000 virtual viewpoints for each experiment.

As it can be seen, rendering methods with full depth information UVST-DM and then UV-DM performed the best with the least distortion (in PSNR) followed by rendering methods with focusing depth information UVST-D and then UV-D. Not surprisingly, the blind rendering methods with no depth information had the highest distortion with UVST performing the best

among blind methods followed by ST, UV and NN. The distance of the scene to the camera grid had a direct effect on output distortion, where further distance caused higher distortion for all methods, more significantly for methods with depth information and less pronounced for blind methods. More importantly, the results show the same trends with the theoretical ESD values shown in Figure 3.9.

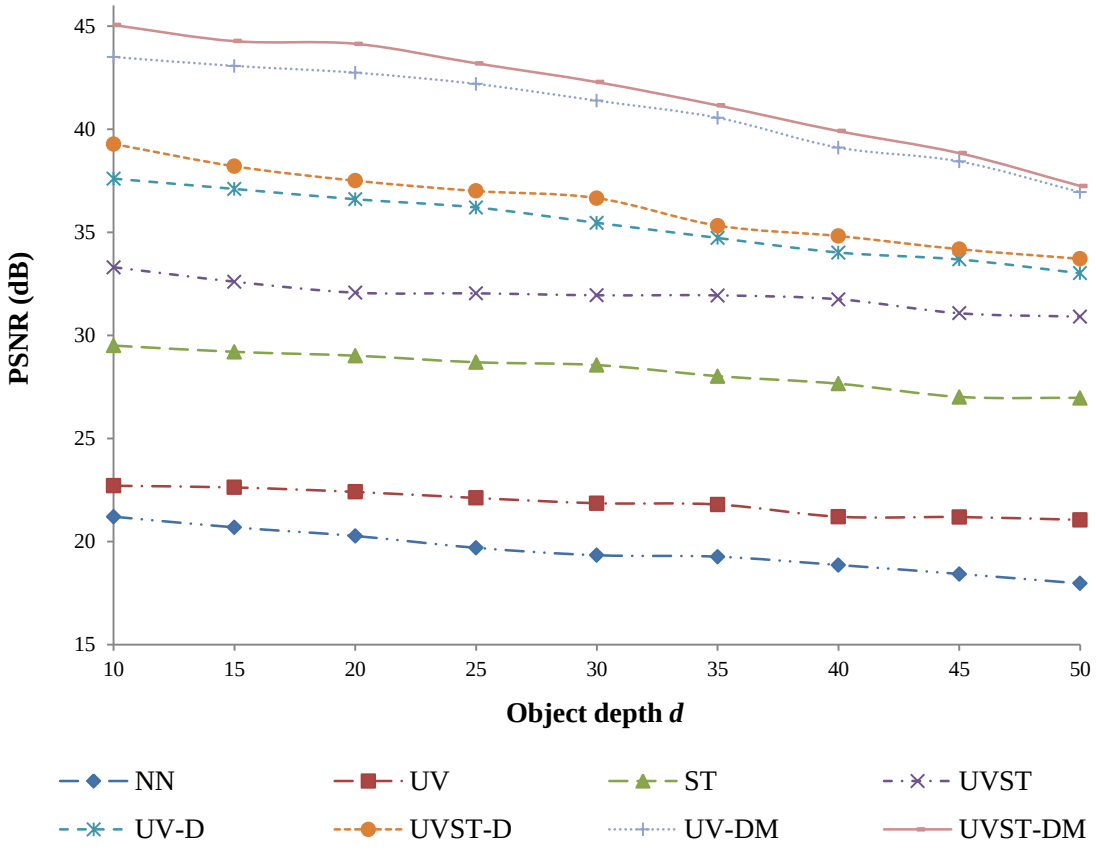


Figure 3.11. Experimental rendering quality in PSNR for different LF rendering methods vs. object depth $\bar{d}$

Figure 3.12 shows the average PSNR values over 32,000 simulations at $d = 10m$. NN interpolation performs the worst; UVST-DM is the best while UVST is the best blind rendering method. This order is consistent with the theoretically calculated ESD shown in Figure 3.10.

Figure 3.13 shows the mean PSNR from 144,000 experiments for different rendering methods, categorized based on the complexity of the scene. As can be seen, more complex scenes result in reduced rendering quality. This can be explained due to fixed ESD for different

scenes with different complexities in term of higher spatial frequency components. Nevertheless, ESD provides the right ranking on the performance amongst the various methods.

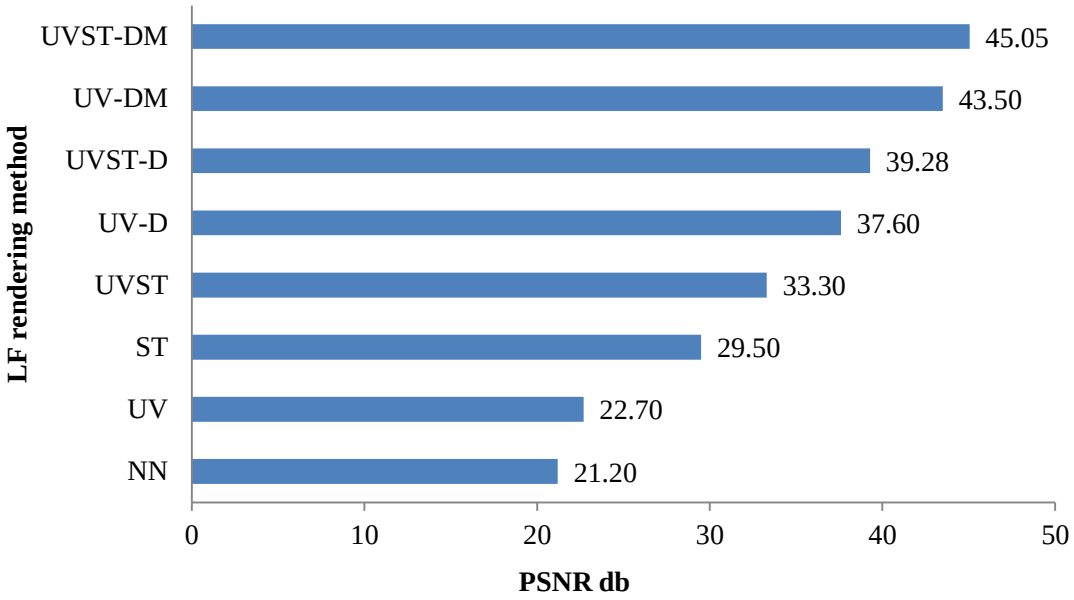


Figure 3.12. Experimental rendering quality in PSNR for different rendering methods

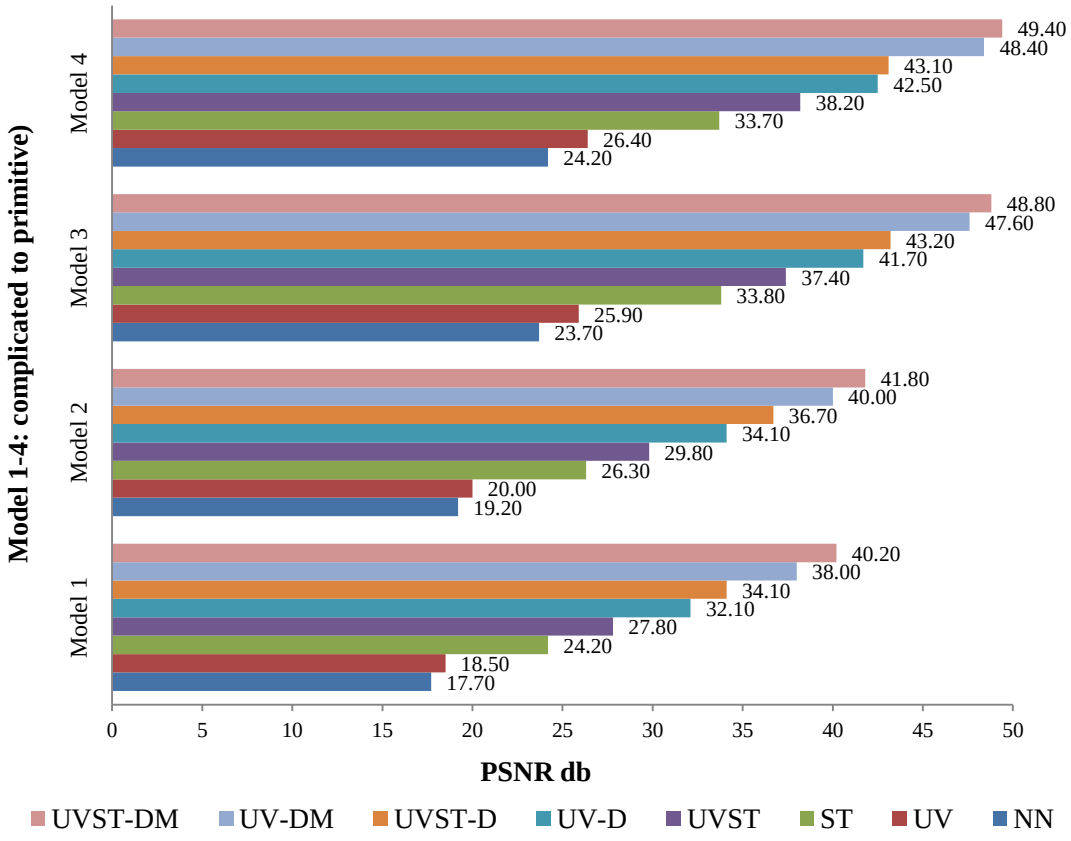


Figure 3.13. Rendering quality and scene complexity

Figure 3.14 shows the rendering distortion from 144,000 experiments based on the distance of the virtual camera to the scene. As it is shown, far navigation results in higher rendering quality compared with closer observations. Again, this can be explained as a consequence of reduction in the required high frequency components to be sampled. Note that this experiment is different from experiments demonstrated in Figure 3.11 and that is why the results are different. In this experiment, the light field system was fixed and the depth of virtual cameras was changed. In the previous experiment, the object depth is changed and the PSNR is calculated as the mean of 1000 random virtual cameras.

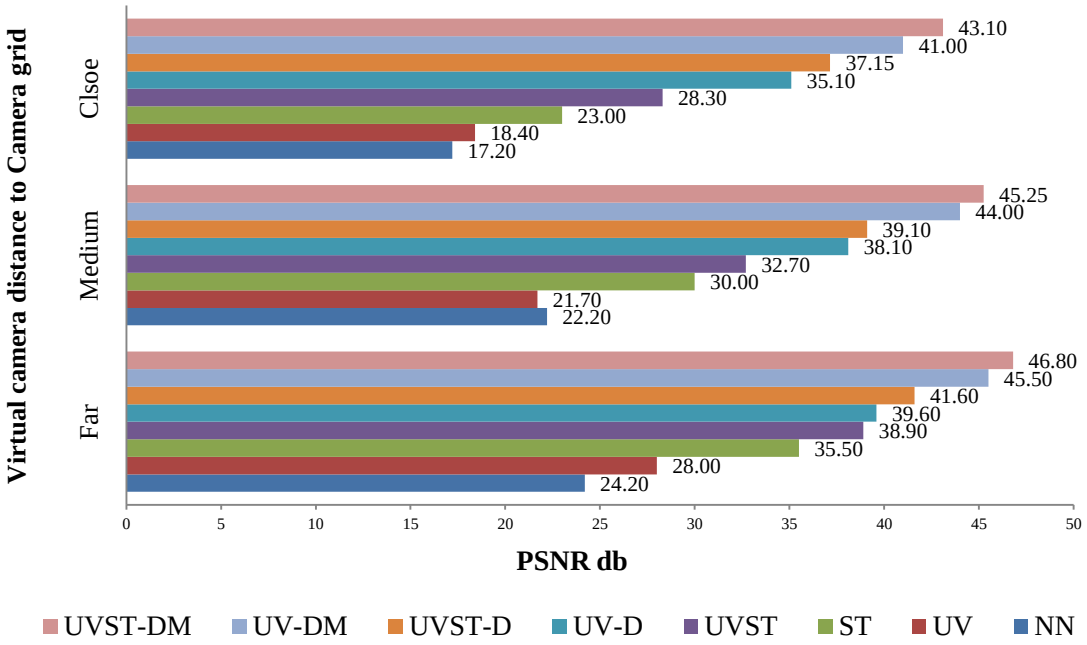


Figure 3.14. Rendering quality and observation distance

Figure 3.14 demonstrates the rendering quality based on the distance of the virtual camera to the scene. As it is shown, far navigation results in higher rendering quality compared with closer observations. Again, this can be explained as a consequence of reduction in the required high frequency components to be sampled. Note that this experiment is different from experiments demonstrated in Figure 3.11 and that is why the results are different. In this experiment, the light field system was fixed and the depth of virtual cameras was changed. In

the previous experiment, the object depth is changed and the PSNR is calculated as the mean of 1000 random virtual cameras.

3.5.5 Results on Acquisition Configurations

By changing l and k respectively, various LF acquisition configurations were simulated.

3.5.5.1 Theoretical Expectations

Figure 3.15 demonstrates the theoretical relationship between k, the distance between the cameras in the camera grid, and ESD. As expected, for all methods, dense camera grid (small k) results in high ESD and therefore high rendering quality. In this Figure, $d = 50m$, $l = 0.05cm$ (camera resolution of $1024x768$), and $k \in [0.1m, 0.9m]$ with the same assumption for depth error as the case shown in Figure 3.9.

As it can be seen, changing the value of k has limited effects on UV-D/UVST-D and UV-DM/UVST-DM, though at large k, UV-D and UV-DM performance gets worse compared to UVST-D and UVST-DM respectively. Also ESD of the ideal case (when there is no error in depth) is independent of k as demonstrated before. However, for blind methods, k has a significant effect on ESD values. NN, UV, ST and UVST all perform poorly especially for a large k. This confirms the view that by utilizing depth information, the cost of acquisition system can be significantly reduced.

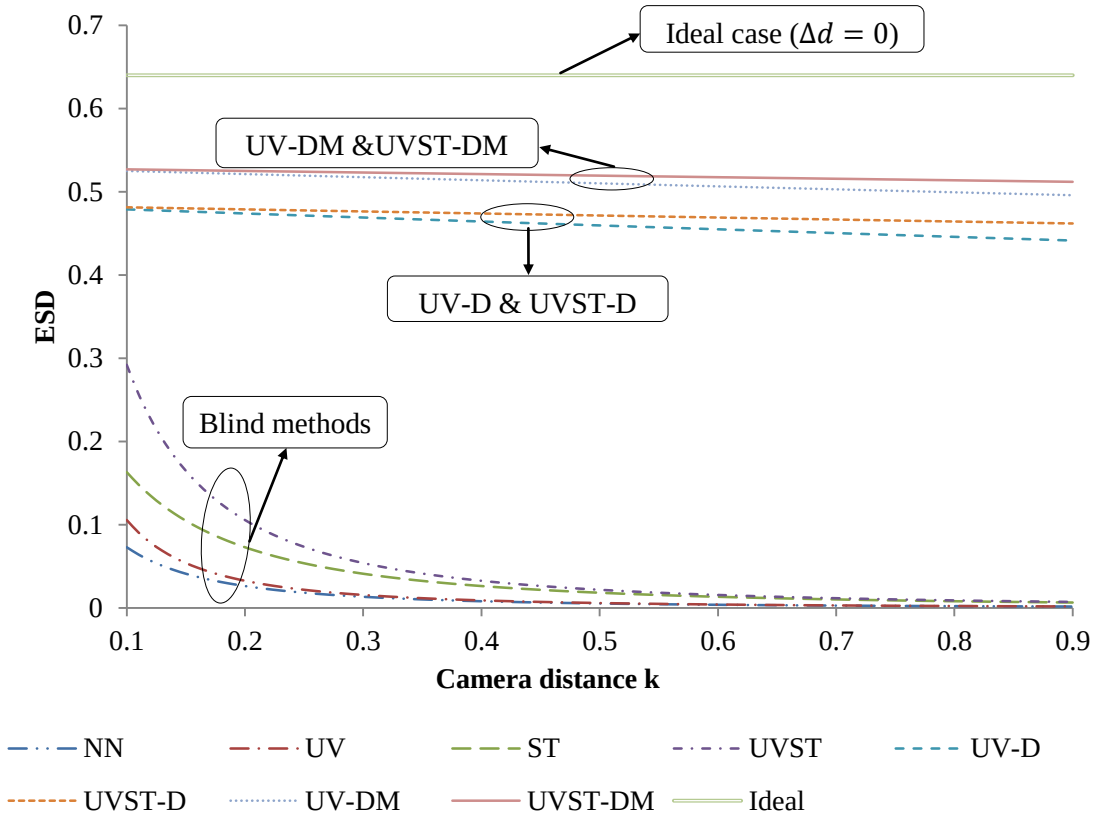


Figure 3.15. Theoretical $\overline{ESD}$ for different LF rendering methods based on camera distance k between $0.1m$ to $0.9m$ for $l = 0.05cm$

Figure 3.16 presents the theoretical relationship between l, the pixel size and ESD. It is clear that for all methods, high resolution (small l) results in high ESD and therefore high rendering quality. In this Figure, $d = 50m$, $k = 0.4m$ and $l \in [0.02cm, 0.1cm]$, i.e., camera resolution of $2560x1920$ to $512x384$ respectively, with the same assumption for depth error as the case shown in Figure 3.9.

As it can be seen, changing l has a direct effect on all methods. This effect is much more significant for UV-D, UVST-D, UV-DM, UVST-DM and the ideal case and less significant for blind methods. NN/UV and also ST/UVST performed similarly especially for a small l (high resolution).

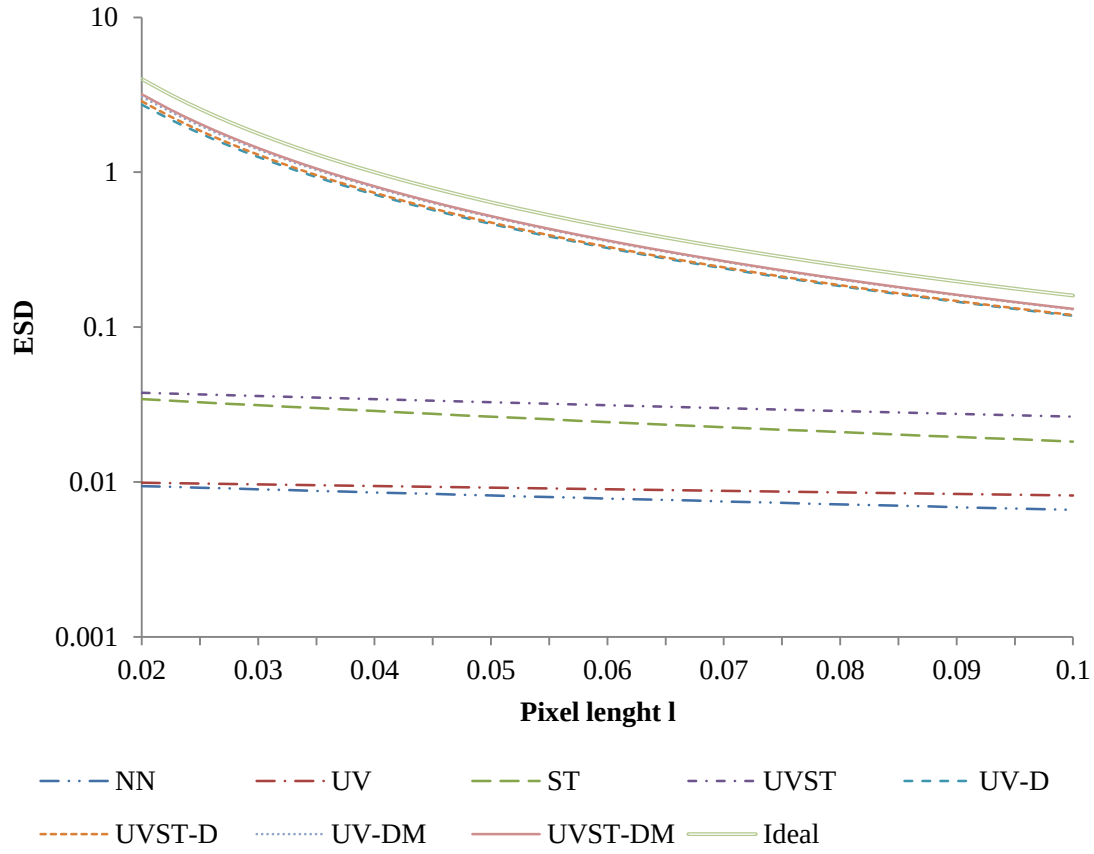


Figure 3.16. Theoretical $\overline{ESD}$ for different LF rendering methods based on pixel length l between $0.02cm$ (camera resolution of $2560x1920$) to $0.1cm$ (camera resolution of $512x384$)

3.5.5.2 Simulation Results

Experiments were carried out to see the effect of k in rendering distortion in term of PSNR so as to make a comparison to the theoretical ESD values. In first experiment, $d = 50m$, object length $= 5m$, $l = 0.05cm$ and $k \in [0.1m, 0.9m]$ and 10% depth error was added. Figure 3.17 shows the results calculated from random 288,000 trials. As it can be seen, large separation between the cameras decreases the rendering PSNR as expected. However, the impact of increasing k is less significant for UV-D, UVST-D, UV-DM and UVST-DM compared to the blind methods.

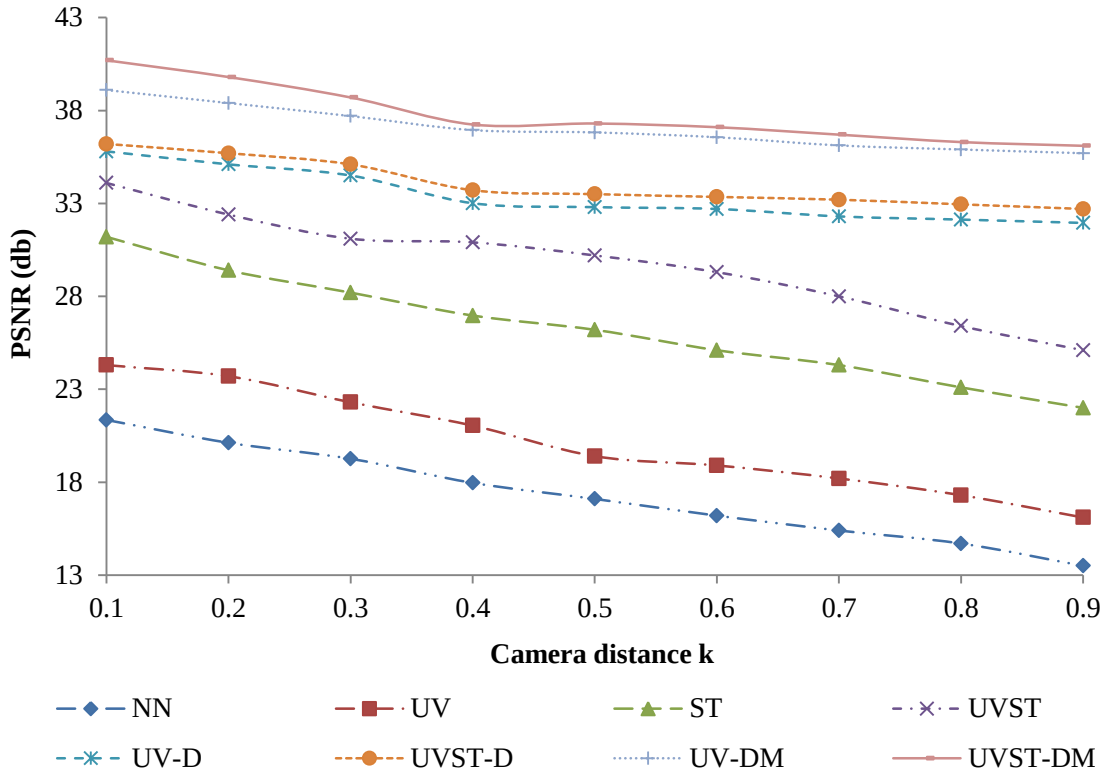


Figure 3.17. Experimental rendering quality in PSNR for different LF rendering methods vs. camera distance *k*

The second experiment shows the relationship between the resolution of cameras (in term of pixel length l) and the rendering distortion in term of PSNR. In this experiment $d = 50m$, object length $= 5m$, $k = 0.4m$ and $l \in [0.02cm, 0.1cm]$, i.e., camera resolution of $2560x1920$ to $512x384$ respectively, and 10% depth error. Figure 3.18 illustrates the results calculated from 288,000 trials. As it can be seen, high resolution (i.e. smaller value of l) increases the rendering PSNR as expected. However, l has less impact on the blind rendering methods and more on UV-D, UVST-D, UV-DM and UVST-DM.

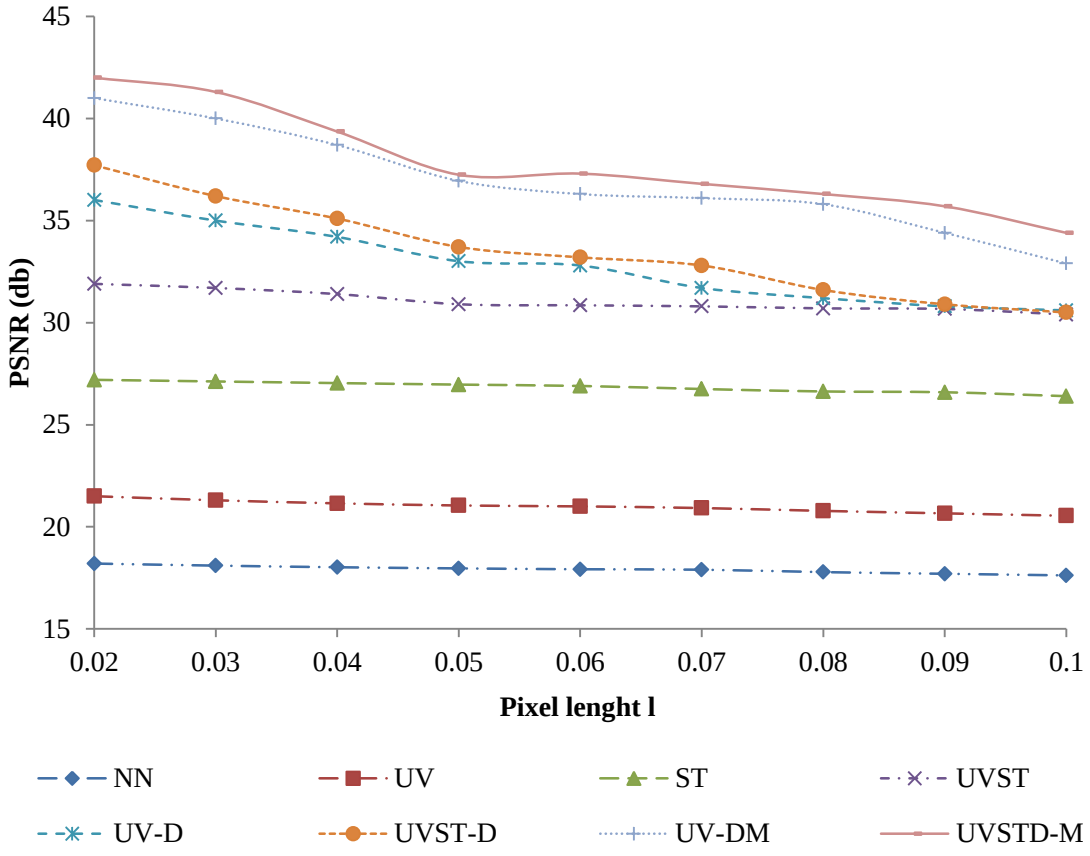
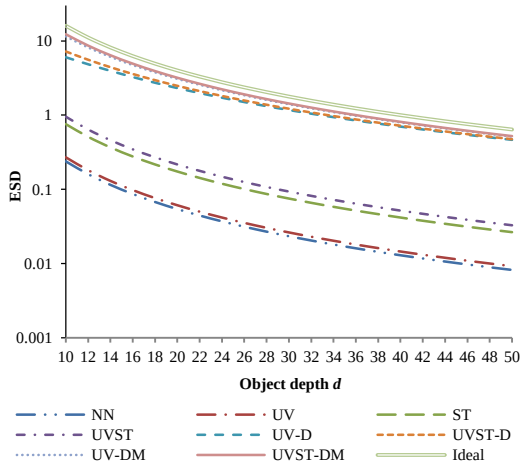


Figure 3.18. Experimental rendering quality in PSNR for different LF rendering methods vs. pixel length l

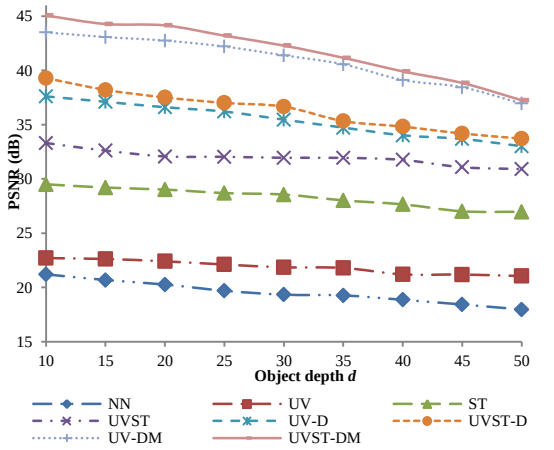
Therefore, the theoretical expectations based on ESD analysis are confirmed by the empirical results. This can be seen more obviously by comparing Figure 3.15 with 3.17 and Figure 3.16 with 3.18. Notice that the theoretical expectation is shown in ESD while the simulation results are shown in PSNR, and their relationship will be examined in the next section.

It should be noted that ESD is a function of d, the depth of a point in the scene space. Hence, it has different values at different points of the scene. Theoretical expectations demonstrated in Figures 3.9, 3.10, 3.15, and 3.16 show the mean $\overline{\mathrm{ESD}}$ for the entire scene by assuming the average depth of the scene/object $\bar{d}$ and average error in depth $\overline{\Delta d}$. In other words, the scene is sampled and rendered with $\overline{\mathrm{ESD}}$.

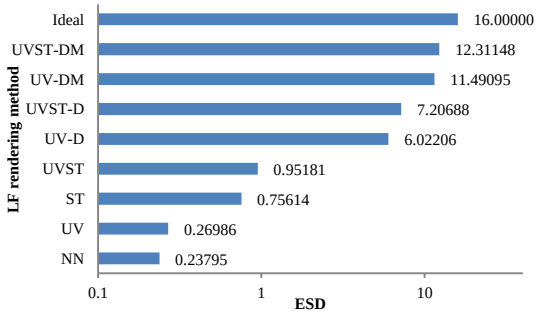
A summary of all the theoretical expectations and experimental results are demonstrated in Figure 3.19 for easy pair to pair comparison.



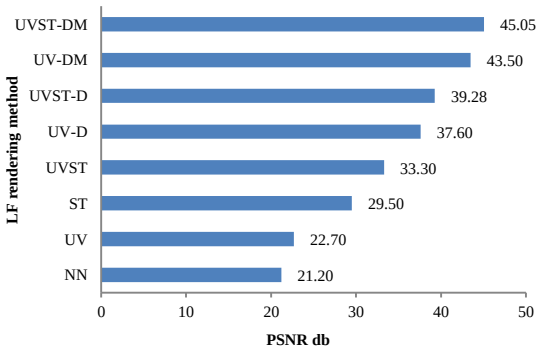
Theoretical $\overline{ESD}$ for different LF rendering methods based on object depth $\bar{d}$ for $k = 0.4m$ and $l = 0.05cm$ (i.e., camera resolution of $1024x768$)



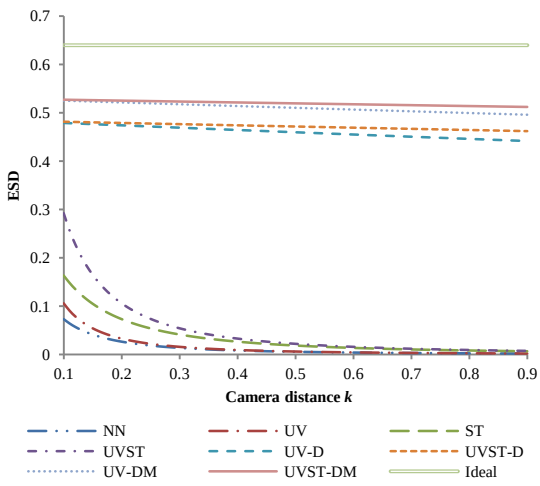
Experimental rendering quality in PSNR for different LF rendering methods vs. object depth $\bar{d}$



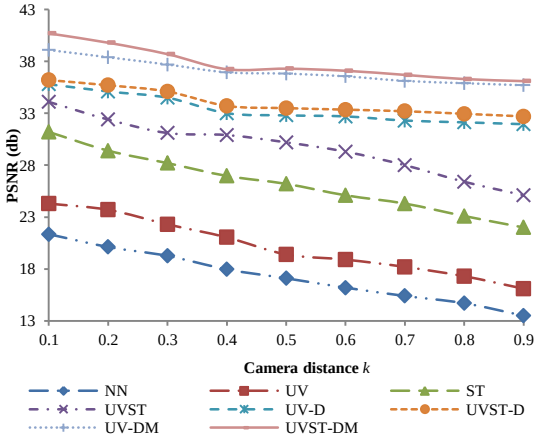
Theoretical $\overline{ESD}$ for different rendering methods for $k = 0.4m$, $l = 0.05cm$, $\bar{d} = 10m$, and $\Delta\bar{d} = 1m$



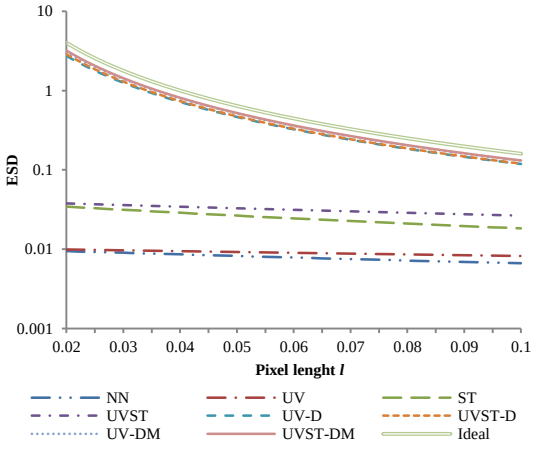
Experimental rendering quality in PSNR for different rendering methods



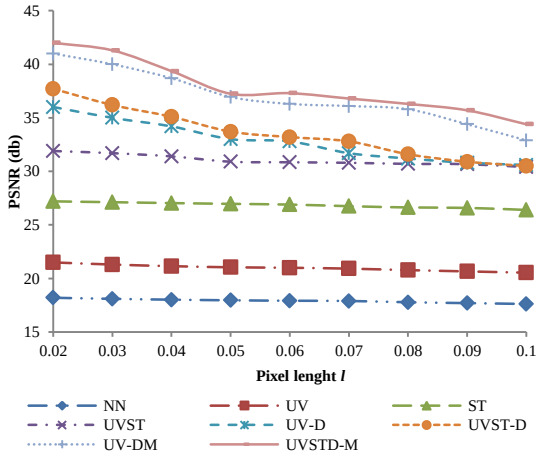
Theoretical $\overline{ESD}$ for different LF rendering methods based on camera distance k between $0.1m$ to $0.9m$ for $l = 0.05cm$



Experimental rendering quality in PSNR for different LF rendering methods vs. camera distance k



Theoretical $\overline{ESD}$ for different LF rendering methods based on pixel length l between $0.02cm$ (camera resolution of $2560x1920$) to $0.1cm$ (camera resolution of $512x384$)



Experimental rendering quality in PSNR for different LF rendering methods vs. pixel length l

Figure 3.19. A summary of LF theoretical expectation and experimental results for pair to pair comparison

3.5.6 Discussions

Figures 3.9 to 3.18 present the theoretical expectations in term of ESD and experimental results in term of PSNR for different scenarios. To verify whether ESD is a good distortion indicator, an analysis was conducted of ESD vs. its counterpart of PSNR, i.e. pairs of Figures (3.9, 3.11), (3.15, 3.17) and (3.16, 3.18). Figure 3.20 shows the average observed experimental PSNR vs. $\bar{d}$ from Figure 3.11 vs. theoretical calculated ESD vs. $\bar{d}$ from Figure 3.9, both obtained by changing the object depth $\bar{d}$. The trendline, covariance, and correlation of PSNR vs. ESD are also shown in Figure 3.20.

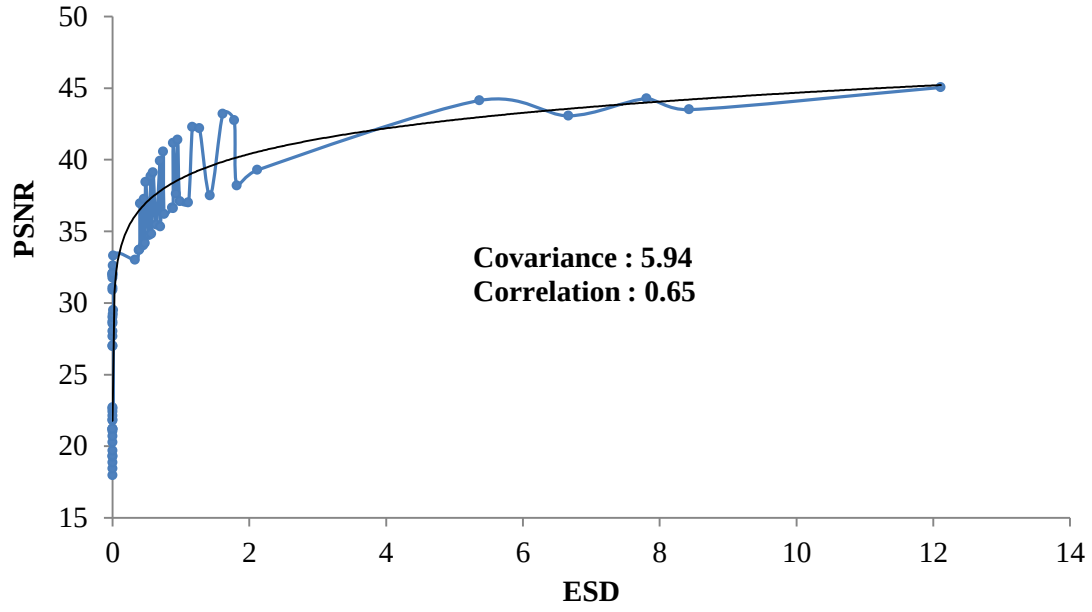


Figure 3.20. Theoretical calculated ESD from Figure 3.9 vs. experimental PSNR from Figure 3.11, both obtained by changing the object depth ($\bar{d}$ from $10m$ to $50m$)

Similarly, Figure 3.21 demonstrates the observed PSNR from Figure 3.17 vs. calculated ESD from Figure 3.15, both obtained by changing the camera density. Again, the trendline, covariance, and correlation of PSNR vs. ESD are shown.

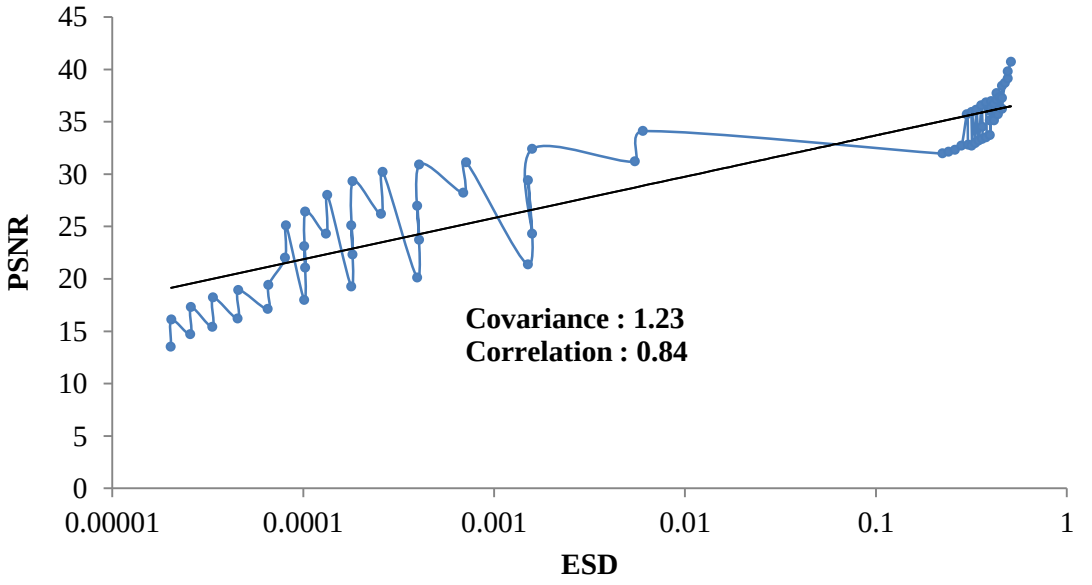


Figure 3.21. Theoretical calculated ESD from Figure 3.15 vs. experimental PSNR from Figure 3.17, both obtained by changing the camera density (k from $0.1m$ to $0.9m$)

Finally, Figure 3.22 demonstrates the observed PSNR from Figure 3.18 vs. calculated ESD from Figure 3.16, both obtained by changing the camera resolution.

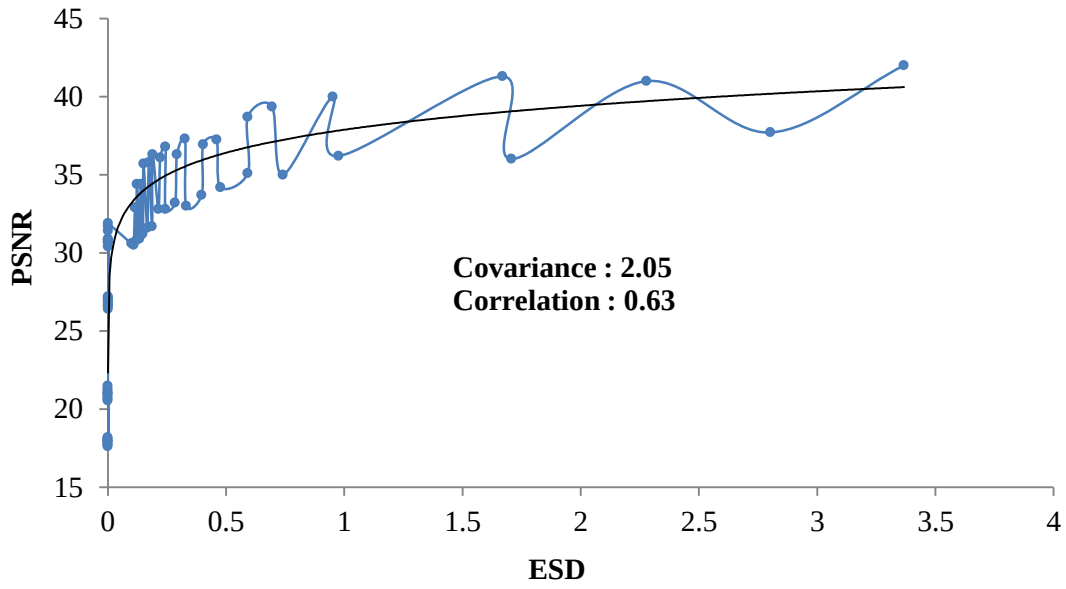


Figure 3.22. Theoretical calculated ESD from Figure 3.16 vs. experimental PSNR from Figure 3.18, both obtained by changing the resolution (l from $0.02cm$ to $0.1cm$)

Figure 3.20, Figure 3.21, and Figure 3.22 show a high correlation between theoretically calculated ESD and observed PSNR. In addition, as the trendlines demonstrate, there is an empirical relationship that can be used to estimate output distortion in PSNR directly from calculated ESD without experiments. This will be explored in the next section.

3.6 Prediction of Output Quality

As shown in the previous Section, there is an explicit conformity between theoretical ESD and the experimental results. This similarity suggests that it may be possible to establish an empirical relationship between the theoretical ESD and PSNR.

As discussed before, the rendering quality is a function of scene frequency components in interpolation area A, ESD in that area and effectiveness of interpolation function F. By

assuming a fixed scene and a fixed interpolation algorithm, it can be argued that overall rendering quality measured in PSNR is a function of $\overline{\text{ESD}}$ for a given LF rendering method. As the only variable determining PSNR is MSE (Mean Squared Error) so it can be said that MSE is a function of $\overline{\text{ESD}}$ for each given LF rendering method, denoted by ESD_{method}, for a given fixed scene, i.e., $\text{MSE} = f(\text{ESD}_{method})$. If function f is known, then the PSNR of a given LF rendering method and LF acquisition configuration can be directly estimated from its ESD.

To find f, a subset of existing data is chosen as the training set for curve fitting and the rest of the data as a validation set to test the accuracy of the empirical model f. To generate the curve fitting data, a map between observed PSNR and expected MSE is calculated as follows:

$$f(\text{ESD}_{method}) = \text{Expected MSE} = \frac{255^2}{10^{\left(\frac{\text{Observed PSNR}}{10}\right)}} \tag{3.13}$$

By feeding the observed PSNR from the training set to equation 3.13, the expected MSE is calculated. We used the data presented in Figures 3.9 and 3.11 (theoretical and experimental results based on changing the object depth) as the training set and data demonstrated in Figures (3.15, 3.17) and (3.16, 3.18) for validation. Figure 3.23 demonstrates the curve fitting for the training set for different LF rendering methods. This curve fitting is done on all the data and without clustering the data based on the LF rendering methods. Figure 3.24 shows the curve fitting for each LF rendering method separately (method-dependent). The optimum value for $f(ESD_{method})$ for best estimation is when it is equal to expected MSE. In both figures $\overline{\text{ESD}}$ is drawn with respect to expected MSE.

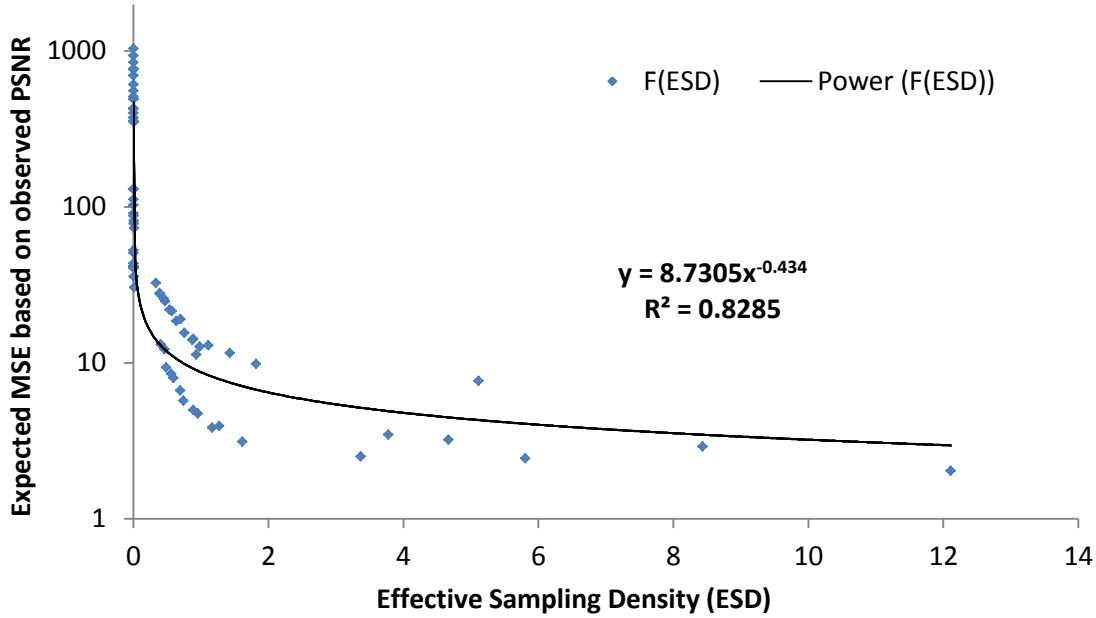


Figure 3.23. A general curve fitting for $f(\mathrm{ESD})$ estimation based on calculated $\overline{\mathrm{ESD}}$ vs. expected MSE

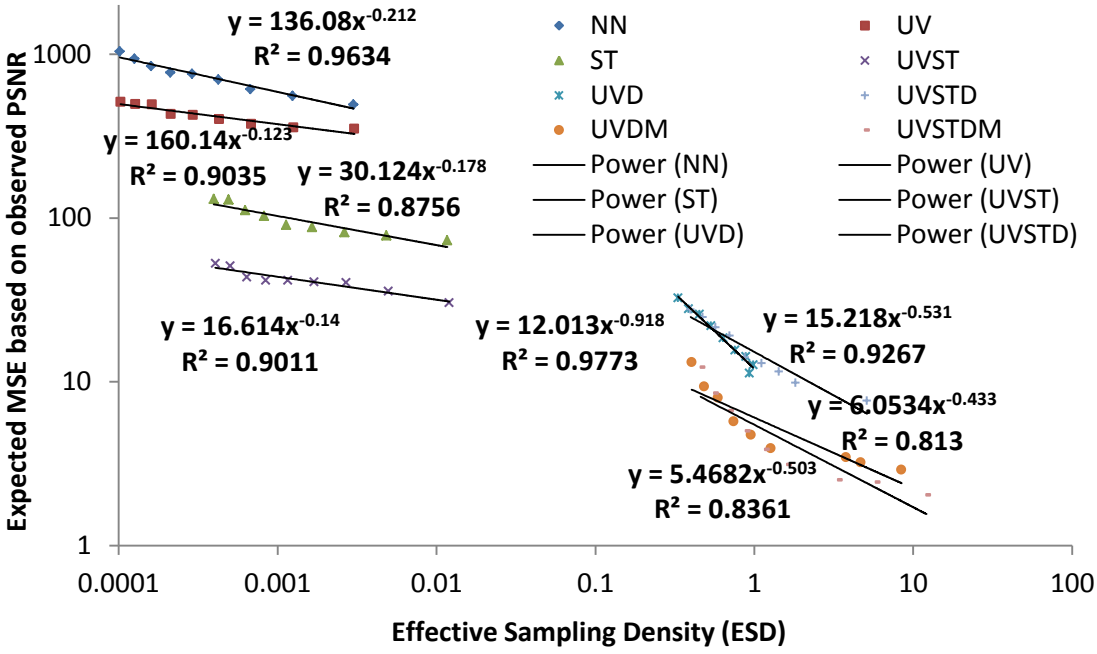


Figure 3.24. Method-dependent curve fittings for $f(\mathrm{ESD}_{method})$

Figure 3.25 shows a summary of curve fitting and validation errors of PSNR estimation for all LF rendering methods. As it can be seen from Figure 3.25, the estimation error for

validation tests for the case when the method is known is less than 3%. If the method-dependent equations are not available, the estimation error for general equation is less than 12%. This shows that empirical equations for $f(\mathrm{ESD}_{method})$ are accurate to predict the rendering quality in term of PSNR. These equations offer a way to directly estimate the overall rendering quality of an LF-based FVV system from calculated ESD without implementation and experiments.

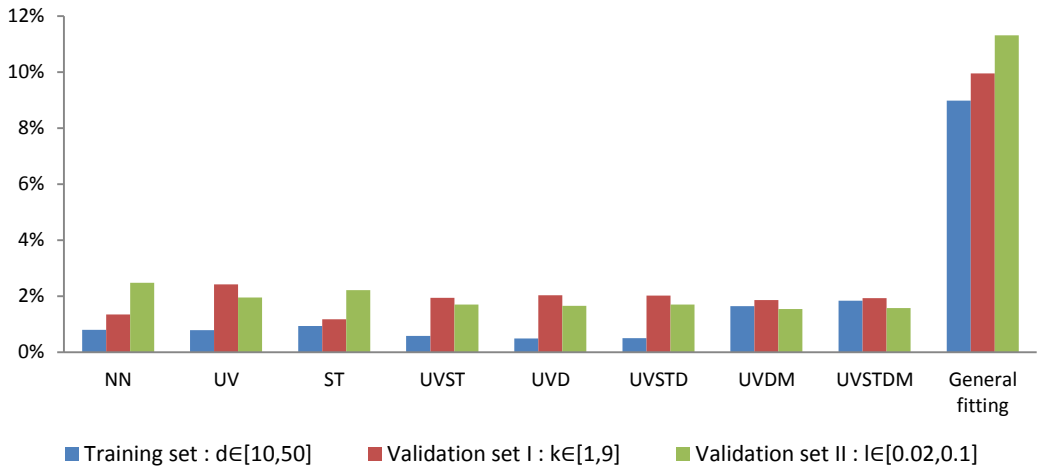


Figure 3.25. Summary of curve fitting training and validation errors of PSNR estimation

By applying the analytical ESD equations to the proposed empirical equations, a direct model to estimate the rendering quality in PSNR from LF system parameters can be formulated. This helps the system designers to optimize the LF acquisition and LF rendering components without exhaustive experimental implementation of each configuration. For instance, for a general UVDM$(d, \Delta d, k, l, |\omega|)$ method, by applying the ESD from Table 3.3, the rendering quality can be directly calculated as:

$$\mathrm{PSNR}_{\mathrm{UVDM}(d,\Delta d,k,l,|\omega|)} \cong 20\, log_{10} \frac{255}{\sqrt{3.4545\left(\frac{|\omega|}{[l(d+\Delta d)+\frac{\Delta d.k}{d}(\sqrt{|\omega|}-1)]^2}\right)^{-0.256}}} \tag{3.14}$$

In general, empirical f can be formulated as (3.15),

$$f(ESD_{method}) = Q.ESD_{method}{}^{P} \tag{3.15}$$

Table 3.5 summarizes the empirical boundaries of Q and P for different LF rendering methods, estimated for different scenes and acquisitions.

Table 3.5

Empirical boundaries of P and Q

LF rendering method type	*LF rendering method*	Q	P
LF rendering methods with no depth information $10 < Q < 300$ $-0.3 < P < -0.1$	NN	$50 < Q_{NN} < 300$	$-0.3 < P_{NN} < -0.2$
	ST	$20 < Q_{ST} < 200$	$-0.2 < P_{ST} < -0.1$
	UV	$20 < Q_{UV} < 250$	$-0.25 < P_{UV} < -0.1$
	UVST	$10 < Q_{UVST} < 200$	$-0.2 < P_{UVST} < -0.1$
LF rendering methods with focusing depth information $10 < Q < 40$ $-1.0 < P < -0.15$	UV-D	$10 < Q_{UVD} < 40$	$-1.0 < P_{UVD} < -0.15$
	UVST-D	$10 < Q_{UVSTD} < 40$	$-1.0 < P_{UVSTD} < -0.15$
LF rendering methods with full depth information $1 < Q < 15$ $-0.9 < P < -0.2$	UV-DM	$1 < Q_{UVDM} < 15$	$-0.9 < P_{UVDM} < -0.2$
	UVST-DM	$1 < Q_{UVSTDM} < 15$	$-0.9 < P_{UVSTDM} < -0.2$
	General Method	$1 < Q < 10$	$-1.4 < P < -0.2$

As discussed before, the differences in $f(\mathrm{ESD}_{method})$ equations can be directly explained due to differences in the scene complexities and interpolation methods. Despite these differences, the general model offers a good indication on what the overall quality in terms of PSNR should be expected by a given $\overline{\mathrm{ESD}}$.

3.7 Subjective Validation

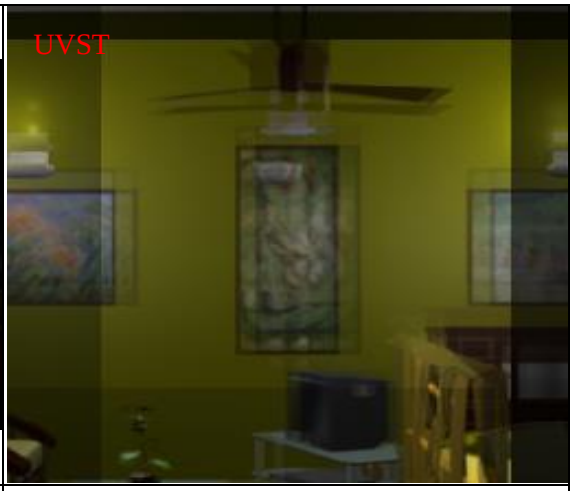
While previous section discussed the correlation between ESD and output video objective distortion in term of PSNR, this section demonstrates that ESD is also highly correlated with subjective assessment of the perceived video quality. A subjective quality

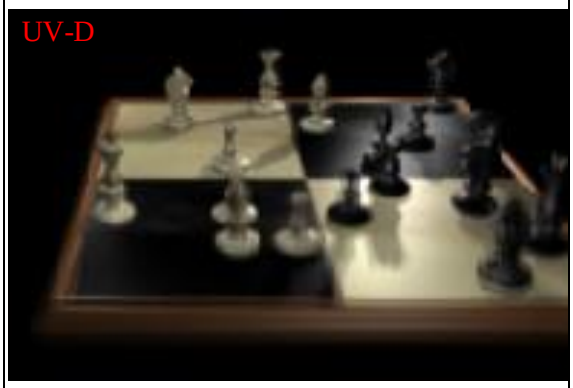
assessment based on ITU-T standardization and guidelines on "subjective video quality assessment methods for multimedia applications" [100] and using degradation category rating (DCR) method was carried out. The test procedure is based on recommendations proposed in VQEG reports [101, 102]. Three rendering method, UVST as a candidate of rendering methods with no depth information, UV-D with focusing depth and UV-DM with full depth information were selected for subjective test. The ground truth from the simulator and Stanford light field archive [103] was used as reference image. The original Stanford camera grid to capture real scenes is $17x17$, i.e., 289 reference images. To provide the ground truth for real scenes with real depth values, a subset of these reference images as a sparse $8x8$ camera grid was selected for acquisition component and a subset of other cameras were used as ground truth. 18 subjects participated in the test. For each of three candidate rendering methods, eight rendering outputs from different viewpoints for four different scenes, "*chess board*" and "*room*" from simulator and "*eucalyptus flowers*" and " *Lego knights*" from Stanford real data were generated. These 96 test sequences as a pair of reference and rendering output were presented to each subject with the recommended time pattern and experiment conditions as proposed in [100, 104]. The subjects were asked to rate the impairment of the second stimulus in relation to the reference into one of the five-level scales: 5:Imperceptible, 4:Perceptible but not annoying, 3:Slightly annoying, 2:Annoying, and 1:Very annoying.

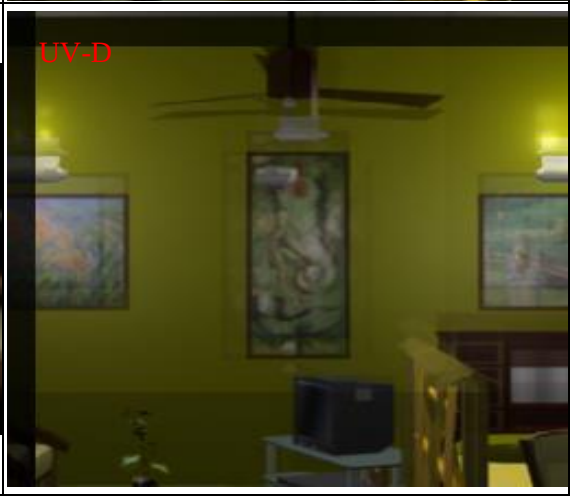
The ESD is also calculated for each pair of scene and rendering method using the equations presented in Table 3.1 and 3.3. There are totally 12 values for ESD (4 scenes and 3 rendering methods). Each value of ESD is corresponded to 8 different views.

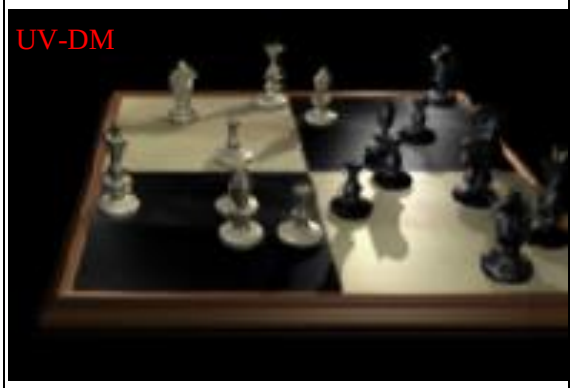
Figure 3.26 shows samples of the test sequences, presented to the subject panel. Note that Figure 3.26 shows 12 different pairs out of 96 test sequences which were presented to each subject.

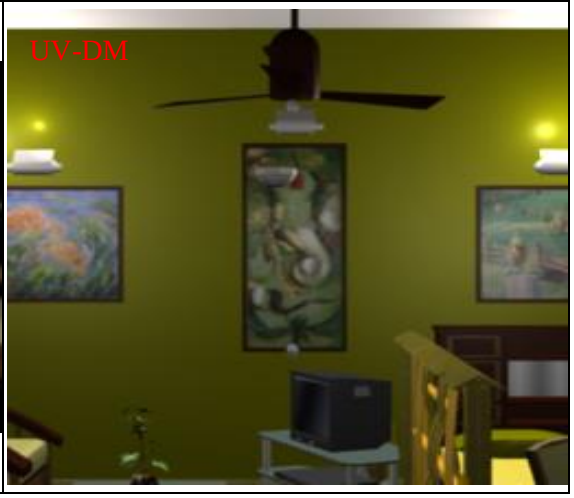

UVST

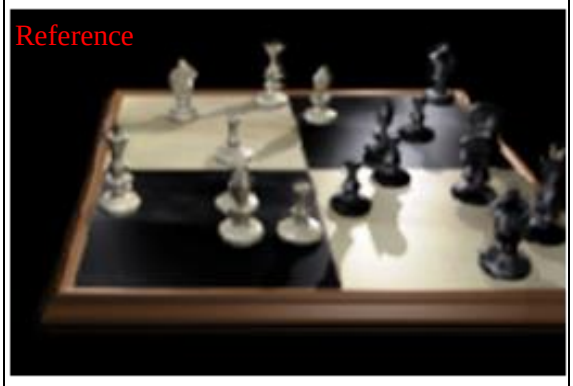

UVST

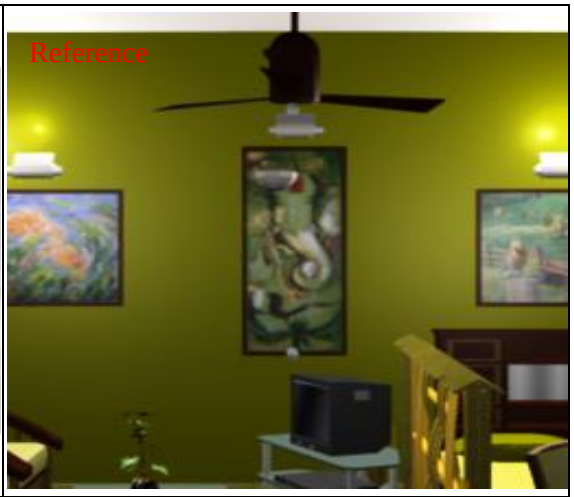

UV-D


UV-D


UV-DM


UV-DM


Reference


Reference


UVST


UVST


UV-D


UV-D

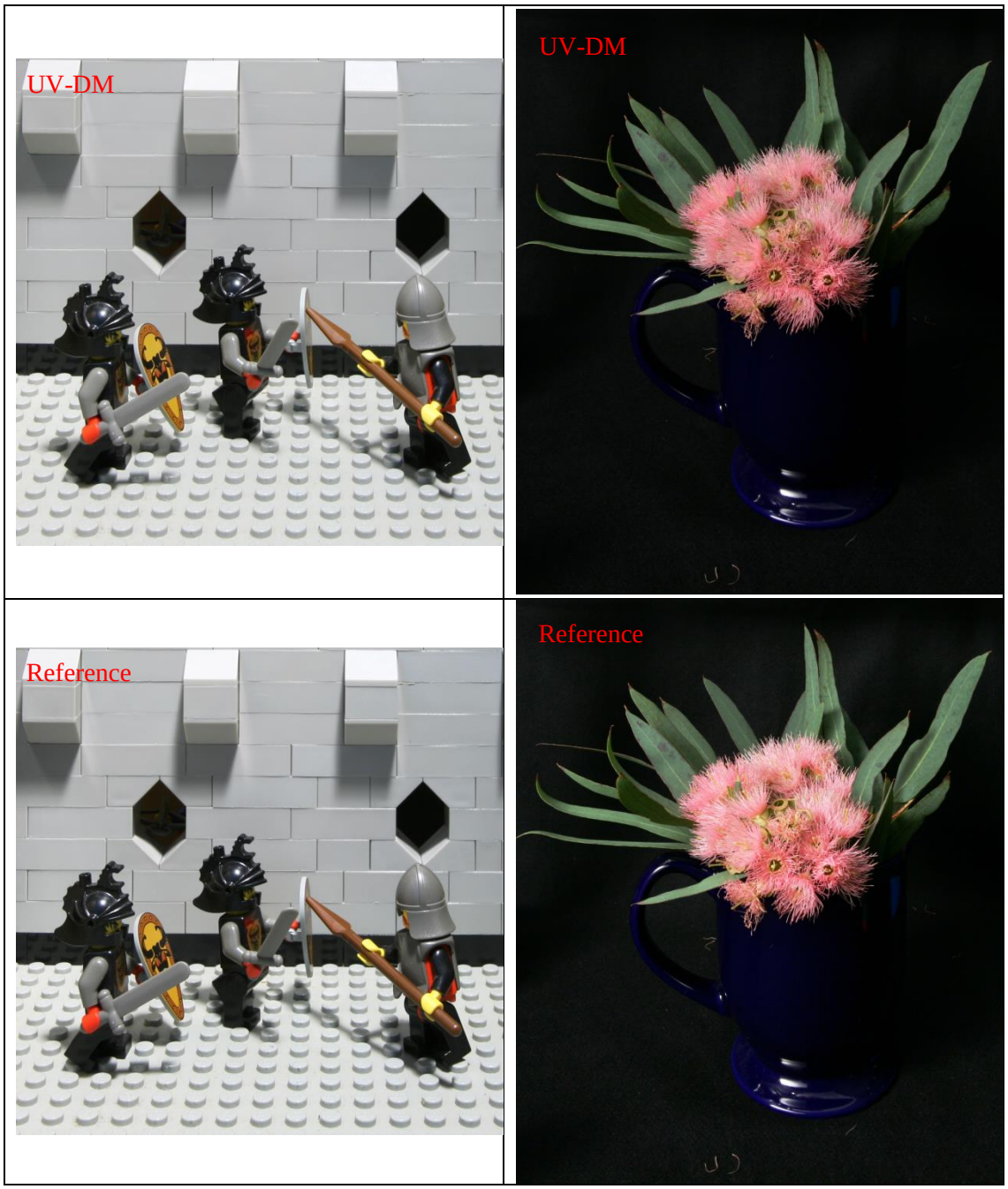


Figure 3.26. Samples of test sequences used in subjective assessment

Figure 3.27 illustrates the results of the subjective test for each rendering method. The average and variance of the impairment for each rendering method was calculated from 576 collected scores (32 test sequences among 18 subjects).

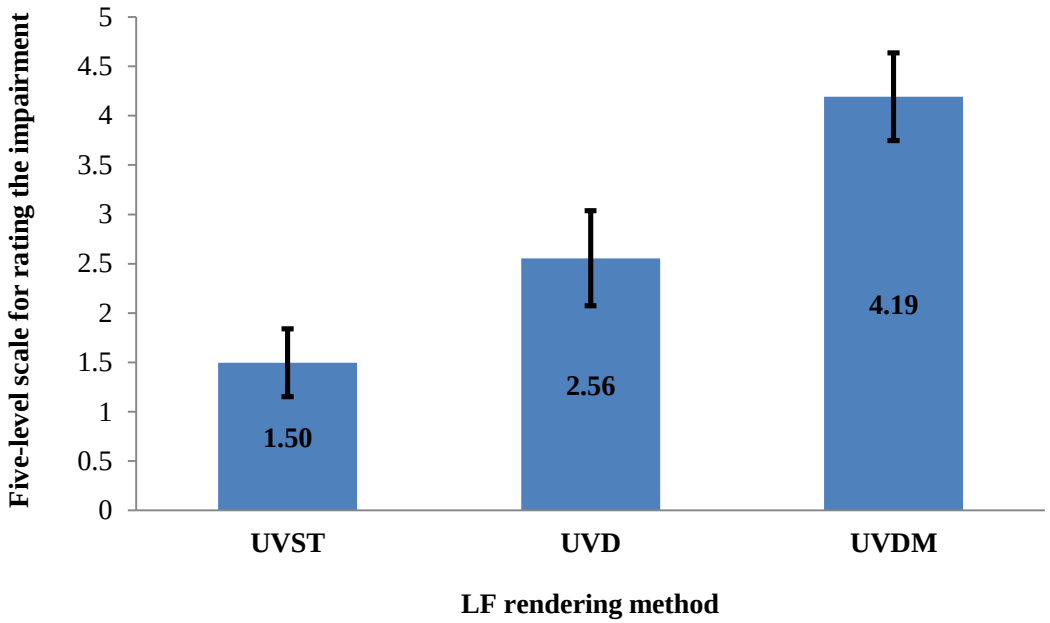


Figure 3.27. Subjective assessment of three LF rendering methods by using degradation category rating (DCR), showing the Mean and Variance of rating from 576 collected scores for each method (32 test sequences among 18 subjects) with a five-level scale for rating the impairment

To validate the relationship between ESD and subjective DCR rating, the procedure for specifying accuracy and cross-calibration of video quality metrics proposed in VQEG reports [101, 102] were employed. Figure 3.28 shows the scatter plot for the ESD-DCR couples for all 96 test sequences. Please note that for each 8 test sequences for different views, there is only one calculated ESD. To obtain the empirical relationship between DCR impairment rating and ESD, a polynomial curve fitting, as one of the candidates in VQEG reports, is applied over the data. The *Pearson correlation coefficient* is calculated as 0.91 which demonstrates a high relationship among ESD and DCR. The curve fitting has a *root mean square error* of 0.34 which demonstrate around 10% error of prediction DCR from calculated ESD which is technically satisfactory.

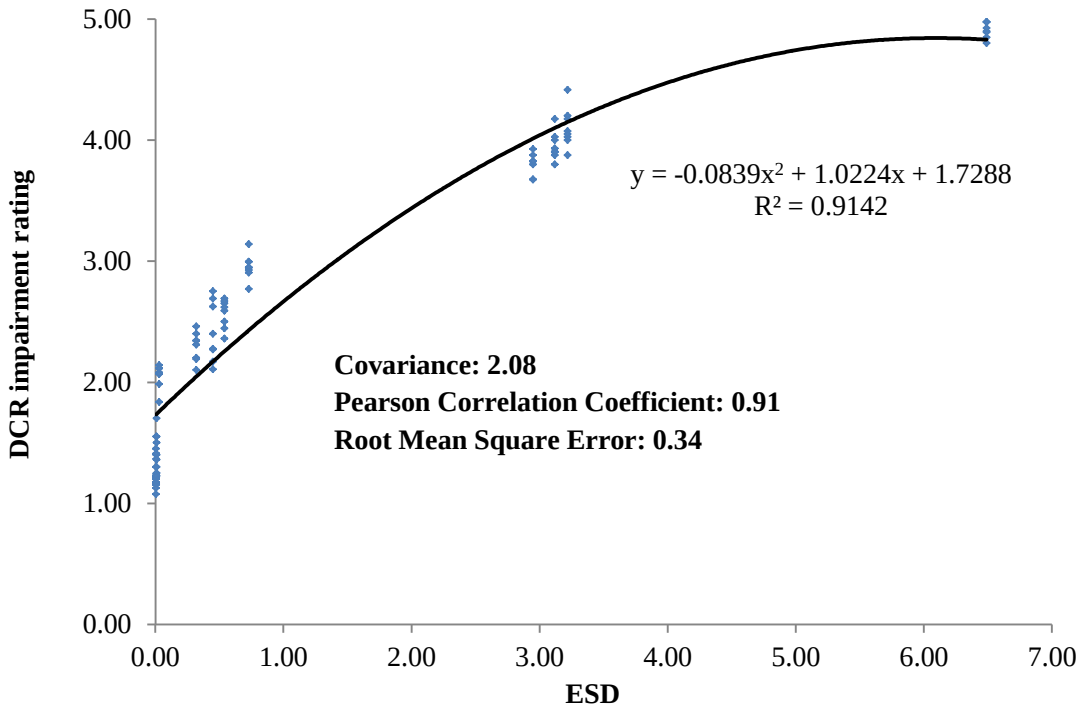


Figure 3.28. DCR impairment rating for subjective quality assessment vs. theoretical ESD and the empirical relationship between these two parameters

3.8 Discussion and Conclusion

This chapter introduced a novel theory based on the concept of ESD. Using ESD different LF rendering methods and LF acquisition configurations can be theoretically evaluated and compared. Eight well-known rendering methods with different acquisition configurations have been analyzed through ESD and simulation. The results have shown that ESD is an effective indicator of output distortion and quality that can be obtained directly form system parameters and that takes into consideration both acquisition and rendering. In addition, an empirical relationship between the theoretical ESD and achievable PSNR has been established. Furthermore, a subjective quality assessment has confirmed that ESD is highly correlated with the perceived output quality. Although this chapter focuses on the overall distortion of an LF-based FVV system, the concept is readily extended to measure the rendering quality at a specific location or part of the scene. A further study on the impact of depth estimation errors on ESD

and optimization of ESD with respect to the *camera density* and *ray selection complexity* for a given output quality will be discussed in the following chapters.

Chapter 4: Acquisition Optimization and Calculation of the Minimum Density of Cameras for a Regular Grid

4.1 Summary

Calculation of the number of cameras required to capture the scene is an essential problem in a practical light field based free viewpoint video (FVV) system. Existing methods calculate the Nyquist rate by assuming a band-limited signal and perfect reconstruction of an arbitrary view using linear interpolation, which often results in an impractically high number of cameras. This chapter proposes a new method based on the concept of effective sampling density (ESD). Specifically, the method assumes the availability of some depth information and explores the trade-off among the depth information accuracy, the required number of cameras, and the desired rendering quality. Theoretical and numerical results show that the resulting number of cameras would be significantly lower than what was reported in the previous studies with only a few percent reduction in the rendering quality. Moreover, it is shown that the previous methods are special cases of the one presented in this chapter.

This rest of the chapter is organized as follows. Section 2 reviews the LF spectral analysis and the LF lossy reconstruction with controlled amount of under sampling. Section 3 describes the proposed method by introducing LF analysis based on ESD and optimization of camera density by employing ESD. Experimental validation and simulation results are presented in Section 4. Section 5 describes the comparison of the proposed method with Chai's LF spectral analysis and Lin's LF geometric analysis. Section 6 concludes the chapter.

4.2 Overview of LF Spectral Analysis

This section presents the essence of LF spectral analysis introduced in [18-20] with a consistent terminology and notation and shows how this analysis has been employed to calculate the minimum number of cameras for LF acquisition.

Typical approach to the problem of calculating the minimum sampling rate of a signal is to derive the Fourier transform of the signal and compute the Nyquist rate for the signal. Without loss of generality, let's assume the LF acquisition can be parameterized using two parallel planes, that is, camera plane uv and image plane st. Light field rendering in spatial domain can be expressed as:

$$r(u,v,s,t) = w(u,v,s,t) * [\, b(u,v,s,t)q(u,v,s,t)] \tag{4.1}$$

where b is a continuous light field, q represents the sampling pattern, w describes a low-pass filtering and interpolation mechanism, r is the rendering output, i.e., the intensity of unknown ray r and $*$ represents convolution operation. Let B, Q, W and R be their corresponding spectra in frequency domain, thus:

$$R(U,V,S,T) = W(U,V,S,T)\big(\, B(U,V,S,T) * Q(U,V,S,T)\big) \tag{4.2}$$

Assume that the depth of each ray is known and denoted as $d(u,v,s,t)$. By applying image disparity to the light field representation b, the intensity of rays captured by camera position (u,v) can be calculated based on camera position $(0,0)$ as:

$$b(u,v,s,t) = b\left(0,0,s - \frac{fu}{d(u,v,s,t)}\,, t - \frac{fv}{d(u,v,s,t)}\right) \tag{4.3}$$

where f is the focal length of the cameras. Note that ray (s,t) is assumed to be seen by camera $(0,0)$. B, the Fourier transform of b, can be represented as:

$$B(U,V,S,T) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} b(u,v,s,t)e^{-2\pi i(Ss+Tt)}dsdt\, e^{-2\pi i(Uu+Vv)}dudv \tag{4.4}$$

Analytical computing of (4.4) is usually not available unless several simplifications and assumptions about the light field are made as discussed later.

By assuming rectangular sampling lattice, q can be represented as:

$$q(u,v,s,t) = \sum_{c_1,c_2,c_3,c_4 \in Z} \delta(u - c_1 k_u)\, \delta(v - c_2 k_v)\delta(s - c_3 l_s)\delta(t - c_4 l_t) \tag{4.5}$$

where $\delta(.)$ is a Dirac delta function, c_1, c_2, c_3 and c_4 are integers, k_u and k_v are the distance between cameras in u and v directions and l_s and l_t are the horizontal and vertical pixel lengths.

Discrete Fourier transform of $B * Q$ is:

$$B(U,V,S,T) * Q(U,V,S,T)$$
$$= \sum_{C_1,C_2,C_3,C_4 \in Z} B\left(U - \frac{2\pi C_1}{k_u}, V - \frac{2\pi C_2}{k_v}, S - \frac{2\pi C_3}{l_s}, T - \frac{2\pi C_4}{l_t}\right) \tag{4.6}$$

Equation (4.6) shows that discrete sampling of the light field signal creates replicas of B shifted to a 4D cube. These replicas are alias components of original signal value of $B(U,V,S,T)$ where $C_1 = C_2 = C_3 = C_4 = 0$ in 4D space. As normally B is not bandlimited, and if bandlimited the practical sampling rate is less than the Nyquist frequency of the signal, these replicas might overlap and create aliasing artefacts in rendering images/videos. To eliminate aliasing artefacts, it has been suggested to oversample the light field or apply a low frequency filter, e.g., Gaussian filter, Mean filter or Median filter. As in most practical applications oversampling is not feasible, most of LF systems adopt some kind of filtering. Several studies have been reported on optimum filtering of the LF signal such as [76].

For a simple Lambertian scene with a constant depth d_0, B is band-limited and can be calculated as [18]:

$$B(U,V,S,T) = 4\pi^2 B'(S,T)\delta\left(\frac{f}{d_0}S + U\right)\delta\left(\frac{f}{d_0}T + V\right) \tag{4.7}$$

where $B'(S,T)$ is the 2D Fourier transform of the captured image by the camera positioned at $(0,0)$. The frequency analysis of this simplified scene can be illustrated by the 2D projection of the spectral support of the 4D continuous signal B over (U,S) or (V,T) planes. For projection over (U,S), the spectral support of $B(U,S)$ is a line: $\frac{f}{d_0}S + U = 0$ and for projection over (V,T), the spectral support of $B(V,T)$ is a line: $\frac{f}{d_0}T + V = 0$. However, due to discrete sampling pattern q of the light field signal, replicas of $B(U,S)$ and $B(V,T)$ emerge at intervals $\left(\frac{2\pi C_1}{k_u}, \frac{2\pi C_2}{k_v}, \frac{2\pi C_3}{l_s}, \frac{2\pi C_4}{l_t}\right)$ in the U, V, S and T directions.

Figure 4.1 illustrates $B(U,S)$ and its replicas. As it is shown, the original signal in frequency domain is only dependent on k_u , the distance between cameras in U axis and has a width of $\frac{2\pi}{k_u}$. In S axis, it depends on l_s, the pixel length in S as well as H, the scene highest frequency component. The signal width in S axis is the minimum of $\frac{2\pi}{l_s}$ and $2\pi H$. The ideal signal sampling requires the light field signal to be filtered to remove all the replicas, but preserve the whole frequency band of the original signal. The red rectangular in Figure 4.1 demonstrates this ideal filtering.

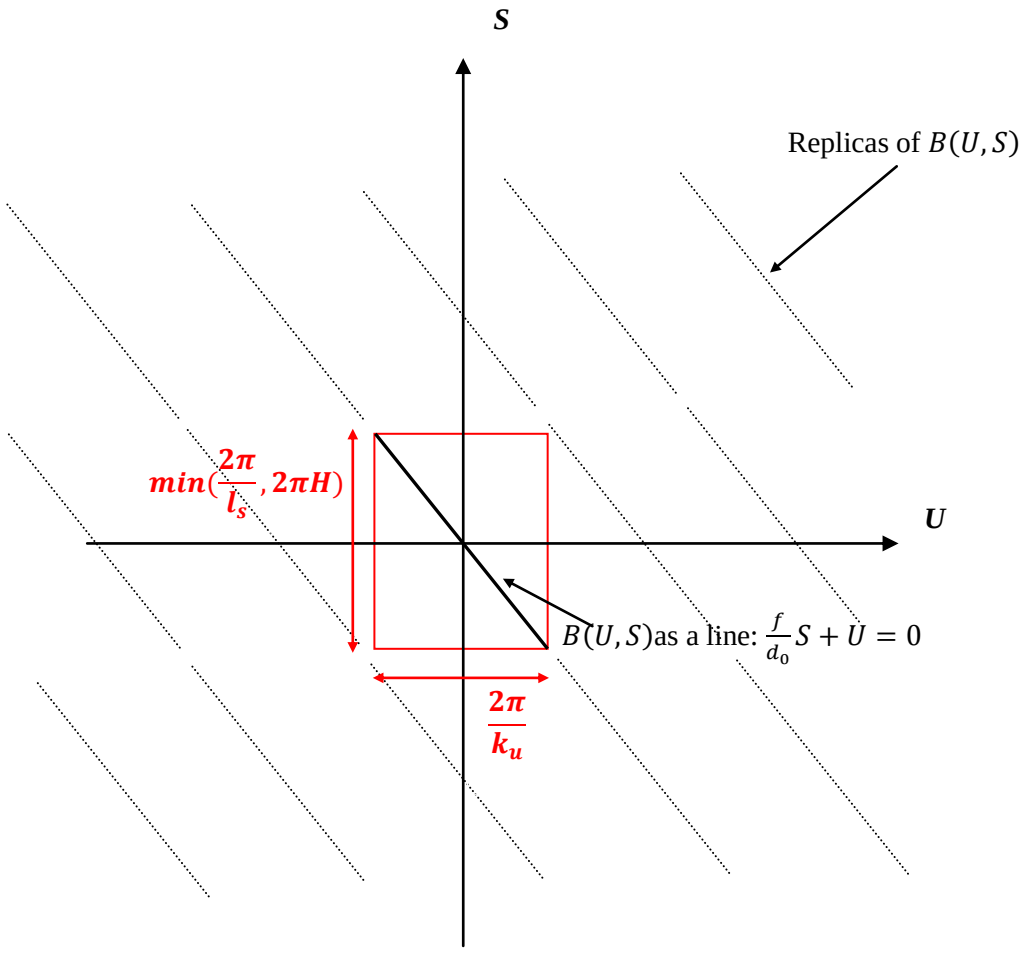


Figure 4.1. Illustration of 2D light field spectrum

Figure 4.2 illustrates a simplified 2D light field system, where a slice of the scene is bounded between d_{min} to d_{max} and its spectral support is a region bounded by 2 lines. The optimum filter for perfect reconstruction of the signal with no aliasing artefacts is shown as a blue dashed parallelogram.

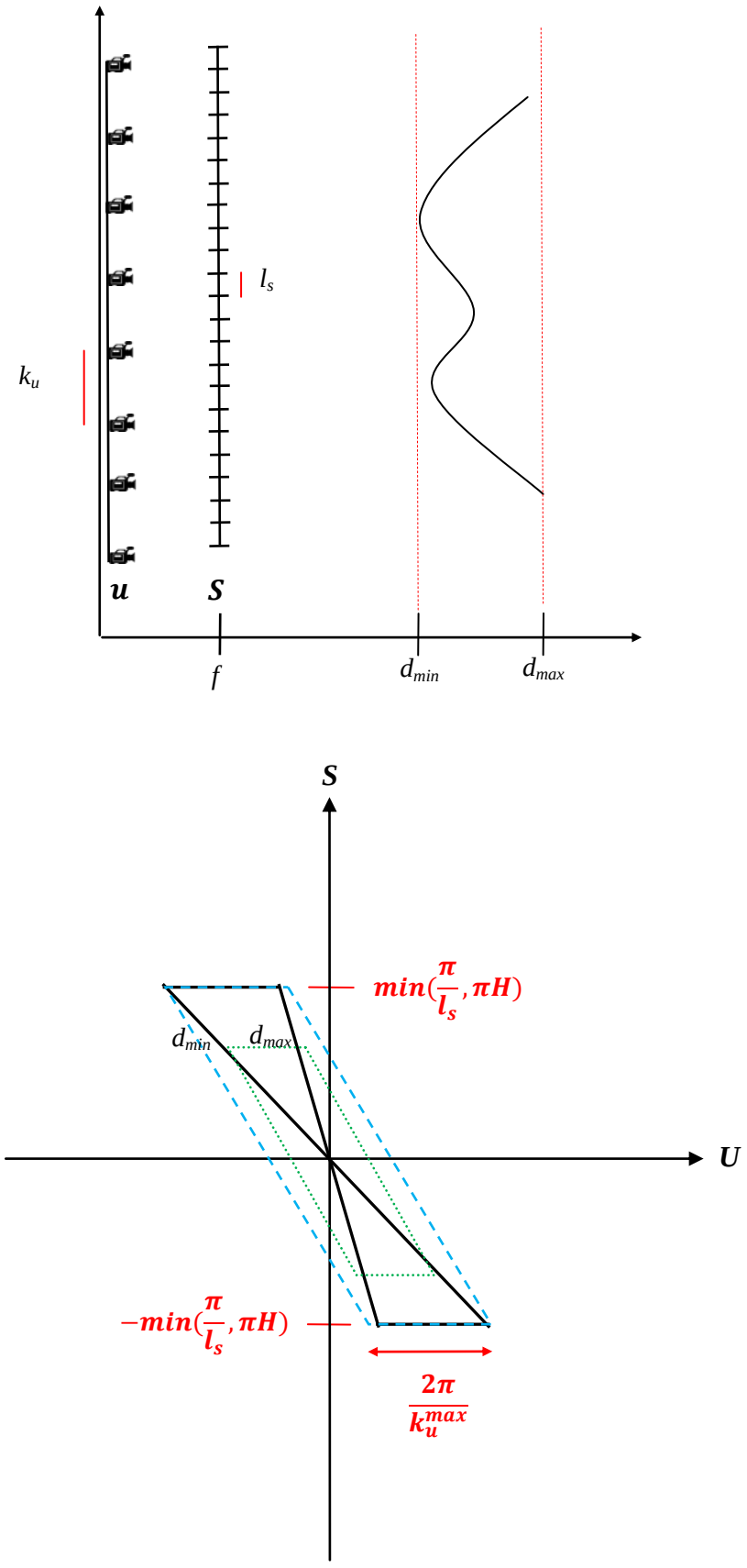


Figure 4.2. A simplified 2D light field and its spectral support

The distance between these 2 lines in U direction determines the maximum k_u, i.e., minimum sample rate. This easily can be calculated as shown in [18] :

$$\frac{2\pi}{k_u^{max}} = f.min\left(\frac{\pi}{l_s}, \pi S\right)\left(\frac{1}{d_{min}} - \frac{1}{d_{max}}\right) \rightarrow k_u^{max} = \frac{1}{f.min\left(\frac{1}{2l_s}, 2H\right)\left(\frac{1}{d_{min}} - \frac{1}{d_{max}}\right)} \tag{4.8}$$

Assume that the scene complexity, in terms of H, is not considered, the focal length $f = 1$, k is used as an abbreviation for both k_u^{max} and k_v^{max} and l for both l_s and l_t by considering a symmetric system configuration, and the depth of the scene is approximated with N_d depth layers from d_{min} to d_{max}. It is possible to derive Chai's equation [18] as:

$$k = \frac{2lN_d}{\left(\frac{1}{d_{min}} - \frac{1}{d_{max}}\right)} \tag{4.9}$$

Without going into the details, the ideal minimum sampling rate can also be calculated with geometric analysis of view interpolation. It is shown that anti-aliasing rendering is equivalent to eliminating the "double image" artefacts caused by view interpolation [32] and k can be calculated as:

$$k = \frac{l(N_d + 1)}{\left(\frac{1}{d_{min}} - \frac{1}{d_{max}}\right)} \tag{4.10}$$

Nevertheless, (4.9) and (4.10) often lead to a very high sampling rate. Thus, many practical LF systems accept some kind of under-sampling. However, to the best of our knowledge, no analytical model has yet been proposed to be able to deal with the under-sampled LF by allowing controlled amount of under-sampling for a desirable rendering quality as will be proposed in the next section.

4.3 LF Acquisition Analysis Based On ESD

In this chapter the acquisition analysis and optimization is carried out for a simplified LF-based FVV system with regular camera grid acquisition, 2-planes representation, symmetric ray selection, and by employing estimated depth maps during rendering. However, the same approach can be used to employ ESD to analyse, evaluate and optimize different LF acquisition configurations and rendering methods. As shown in previous chapter, the generic rendering method for this simplified LF system with employing 2D interpolation in camera plane uv and neighbourhood estimation in image plane st can be represented as UVDM$(d, \Delta d, k, l, |\omega|)$ and its ESD can be calculated as follows:

$$\mathrm{ESD}_{\mathrm{UVDM}(d,\Delta d,k,l,|\omega|)} = \frac{|\omega|}{\left(l(d+\Delta d) + \frac{\Delta d.\, k}{d}\left(\sqrt{|\omega|}-1\right)\right)^2} \tag{4.11}$$

While the following analysis is carried out for UV-DM method but the same approach can be generalized to UVST-DM as discussed before in chapter 3. $\mathrm{ESD}_{\mathrm{UVDM}(d,\Delta d,k,l,|\omega|)}$ predicts the rendering quality as shown in chapter 3. In the above expression, d is given by scene geometry and Δd is determined by the depth estimation method and cannot be altered by us. Changing the other three parameters could potentially improve the rendering quality. By assuming a given camera resolution, i.e., a fixed value of l, two other parameters can be tuned to compensate for the depth estimation error while maintaining the rendering quality. These parameters include k as a measure of density of cameras during acquisition and $|\omega|$ as an indicator of complexity of rendering method. ESD is proportional to $|\omega|$ and inversely proportional to k. It means higher camera density in camera grid (smaller k) and employing more rays for interpolation results in higher ESD.

4.3.1 The Relationship of ESD and Number of Rays in |ω|

Figure 4.3 shows the theoretical calculation of mean $\overline{\text{ESD}}$ for UVDM$(d, \Delta d, k, l, |\omega|)$, for a given light field system with regular camera grid with $k = 5$ and $l = 0.01$, average depth of scene $\bar{d} = 100$, relative depth map error $\frac{\Delta d}{d}$ between 0% to 20%, for three different values of $|\omega| = 4$, 16 and 32. Notice that the estimation error for depth map in most real application is around 10% to 20%.

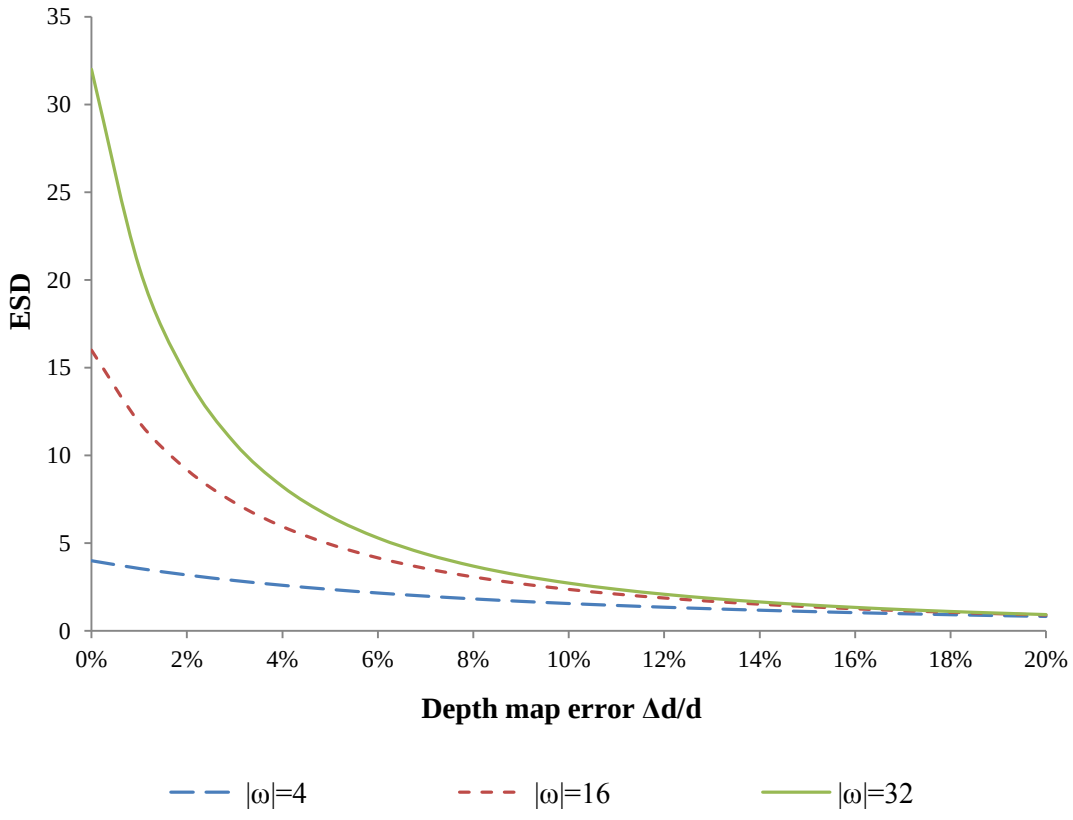


Figure 4.3. Theoretical $\overline{\text{ESD}}$ for UVDM$(d, \Delta d, k, l, |\omega|)$ for $\bar{d} = 100$, depth map with relative error $\frac{\Delta d}{d}$ in the range of $[0\%, 20\%]$, $l = 0.01$, $k = 5$, for $|\omega| = 4$, 16 and 32

As it can be observed from Figure 4.3, higher errors in depth estimation result in less ESD and subsequently less rendering quality when |ω| is fixed. The reason is that error in depth Δd increases the area A for a given |ω| and therefore decreases ESD. However, choosing more

rays for interpolation could increase the ESD and consequently rendering quality. For example, the ESD for 16 samples with errors less than 7% is still better than 4 samples with 1% error.

This analysis shows that increasing the number of rays for interpolation could compensate for the adverse effect of depth map estimation errors on ESD to some degree, at least when the depth error is not very large. Chapter 5 demonstrates the detail of ESD optimization based on |ω| which is a rendering parameter. In this chapter, fixed optimum |ω| is assumed and the effect of k, which is the density of cameras as a main parameter of regular grid LF acquisition, on ESD is investigated. Chapter 6 discusses a joint optimization of both parameters by applying a Lagrangean method to ESD.

4.3.2 The Relationship of ESD and Density of Cameras k

Figure 4.4 shows the theoretical calculation of mean $\overline{\mathrm{ESD}}$ for $\mathrm{UVDM}(d, \Delta d, k, l, |\omega|)$, for a given light field system with regular camera grid with $l = 0.01$, $|\omega| = 4$, average depth of scene $\bar{d} = 100$, relative depth map error $\frac{\Delta d}{d}$ between 0% to 20%, for four different values of $k = 5, 10, 20$ and 50.

As it can be observed from Figure 4.4, higher errors in depth estimation result in less ESD and subsequently less rendering quality when k is fixed, because area A increases for the same reason as described before. However, small k (that is, a denser camera grid), reduces A and could increase the ESD and consequently rendering quality. For example, the ESD for $k = 5$ with depth estimation error less than 17% is still better than $k = 50$ with only 2% error.

Obviously, increasing the density of cameras, i.e. decreasing the distance between the neighbouring cameras could compensate for the adverse effect of depth map estimation errors on ESD to some degree. Of course, increasing the number of cameras results in higher

acquisition cost. Thus in an LF rendering with a prior knowledge of the depth map estimation error, minimum number of cameras can be calculated in advance.

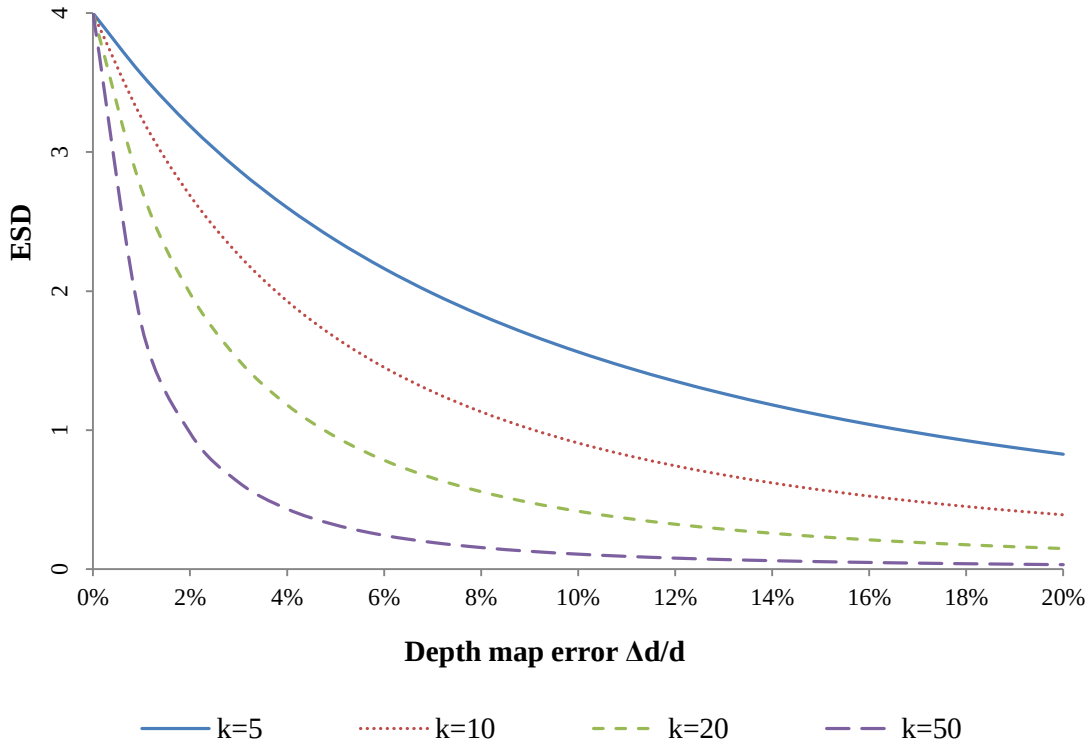


Figure 4.4. Theoretical ESD for UVDM$(d, \Delta d, k, l, |\omega|)$ at different camera densities $k = 5, 10, 20$ and 50, for $\bar{d} = 100$, depth map relative error $\frac{\Delta d}{d}$ in the range of $[0\%, 20\%]$, $l = 0.01$, and $|\omega| = 4$

4.3.3 The Proposed Method for Optimization of k

To avoid quality deterioration due to errors in depth maps, an optimum sampling rate or optimal k should be calculated. In this section, a theoretical maximum distance between neighbouring cameras is derived which can be mapped to the minimum number of required cameras.

In an ideal scenario, where there are no errors in depth map estimation and there is a depth map for each camera in the system, according to the scene reflection complexity one or

more rays would be enough for an accurate rendering. In this case,

$$\mathrm{ESD}_{\mathrm{Ideal}} = \mathrm{ESD}_{\mathrm{UVDM}(d,0,k,l,n)} = \frac{n}{(ld)^2} \text{ and } n \geq 1 \tag{4.12}$$

where $n = 1$ is for the pure Lambertian reflection scene. Higher value of n can be used to model non-Lambertian reflection.

So, the optimization problem is posed as follows: what would be the maximum k (i.e., the minimum density of cameras) for any given $\mathrm{UVDM}(d, \Delta d, k, l, |\omega|)$ with known depth map error Δd to have the same ESD as the ideal case?

$$\mathrm{ESD}_{\mathrm{UVDM}(d,\Delta d,k,l,|\omega|)} = \mathrm{ESD}_{\mathrm{Ideal}} \rightarrow \frac{|\omega|}{\left(l(d+\Delta d) + \frac{\Delta d.\, k}{d}\left(\sqrt{|\omega|} - 1\right)\right)^2} = \frac{n}{(ld)^2} \rightarrow$$

$$k = \frac{ld(d\sqrt{\frac{|\omega|}{n}} - d - \Delta d)}{\Delta d(\sqrt{|\omega|} - 1)} = \frac{l\left(\left(\sqrt{\frac{|\omega|}{n}} - 1\right)d^2 - d\Delta d\right)}{\Delta d(\sqrt{|\omega|} - 1)} \tag{4.13}$$

where $\Delta d > 0$ and $|\omega| > n(\frac{d+\Delta d}{d})^2$

Equation (4.13) shows the maximum k for a light field acquisition system to avoid quality deterioration due to errors in depth maps.

The first condition states that this equation is only valid for $\Delta d > 0$. For an ideal scenario when $\Delta d = 0$, ESD is calculated from the ideal case, shown in (4.12) and k cannot be determined. The second condition explains that for a given Δd, a minimum $|\omega|$ is required to compensate for the effect of Δd in ESD. In other words, decreasing k alone is not sufficient. For the purpose of this chapter, it is assumed that this minimum $|\omega|$ which can be mapped to a minimum required $|\Omega|$ and thus SD is always available in each point of the scene. This assumption implies that for the calculated k from (4.13), the cameras field of view is sufficient to provide enough $|\Omega|$ and thus SD in each point p.

It should be noted that ESD is a function of d, the depth of a point in the scene space (see Figure 3.4 for SD). Hence, it has different values at different points of the scene. Therefore, typically for a given scenario, (4.13) is applied to the mean $\overline{\text{ESD}}$ for the entire scene by assuming the average depth of the scene $\bar{d}$ and average error in depth $\overline{\Delta d}$ to calculate average $\overline{|k|}$. Positioning the cameras according to this $\overline{|k|}$ guarantees the scene to be sampled and rendered with average $\overline{\text{ESD}_{\text{Ideal}}}$.

If the design criteria requires the scene to be sampled and reconstructed by a minimum $\text{ESD}_{\text{Ideal}}$ instead of average $\overline{\text{ESD}_{\text{Ideal}}}$, (4.13) should be applied to all d ranging between (d_{min}, d_{min}), the minimum and maximum depths of the scene with corresponding Δd. This gives optimum k for each depth d and the minimum k can be chosen for camera density.

Figure 4.5 shows, the same system demonstrated in Figure 4.4, but this time for any given depth estimation error $\Delta d < 20\%$, k is calculated directly from (4.13) to maintain $\overline{\text{ESD}}$ at 4.00, the ideal ESD calculated for $n = 4$. Figure 4.6 shows the calculated k in such a scenario, where $|\omega|$ is calculated as follow to satisfy the condition of (4.13): $|\omega| > 4(\frac{100+20}{100})^2 > 5.76 \rightarrow |\omega| = 6$. In addition to ESD for optimum k calculated from (4.13), the ESD for fixed $k = 14.4$ is also demonstrated in Figure 4.5 for comparison.

The corresponding point for 10% error in depth estimation is highlighted in Figure 4.5 and Figure 4.6, respectively, to show the relation of these two Figures. The calculated maximum k keeps the ESD at a fixed rate 4.00 for any error in the depth map. To compensate for just 5% error in depth map as demonstrated in Figure 4.6, the camera distance k decreases by almost 6 times, i.e., 36 times higher camera density, and 15% error in depth estimation needs almost 1900 times higher camera density compared with just 1% error in depth estimation. Note that these small values of k to compensate for high error in depth only occurs when $|\omega|$ is fixed. ESD cannot be increased indefinitely by only decreasing k because the curve of ESD vs. k is saturating as k decreases: $\lim_{k\to 0}(\text{ESD}_{\text{UVDM}(d,\Delta d,k,l,|\omega|)}) = \frac{|\omega|}{(ld+l\Delta d)^2}$. Hence, in general, the

error in depth map should be compensated for with altering both $|\omega|$ and k as will be discussed in chapter 5 and 6.

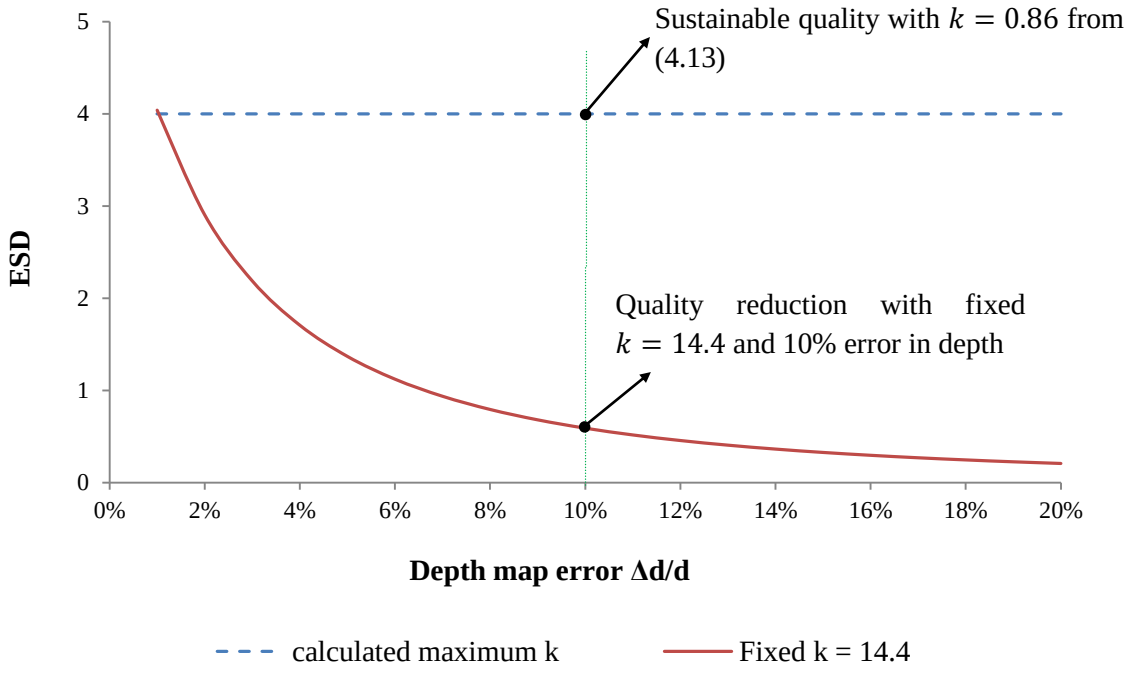


Figure 4.5. Theoretical impact of depth estimation error on rendering quality ($\overline{ESD}$) for fixed $k = 14.4$ and calculated k from (4.13)

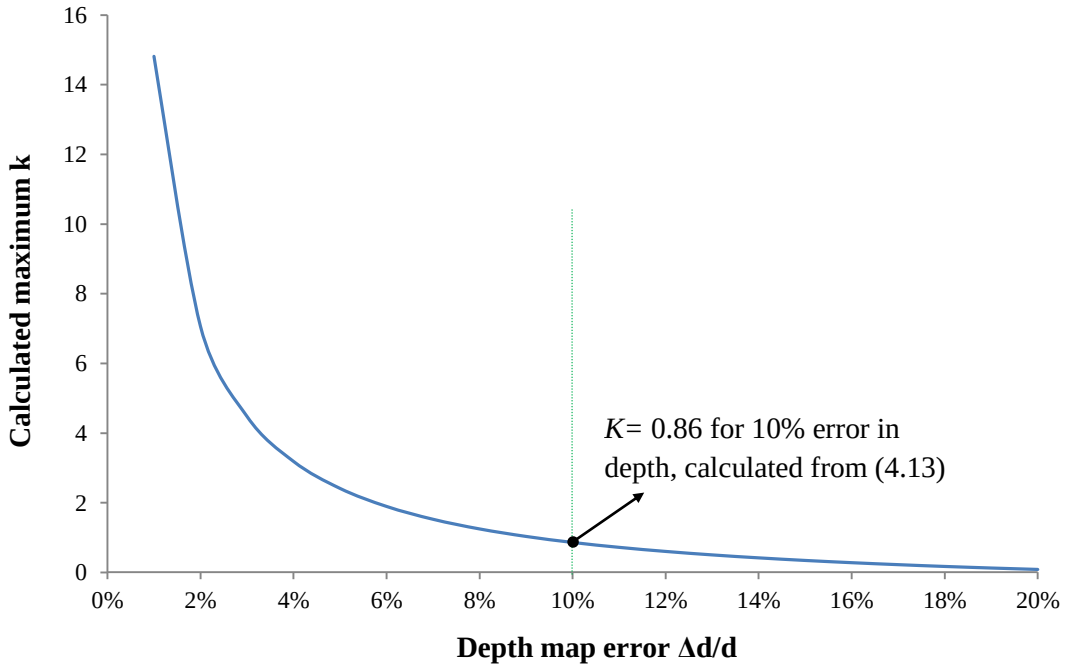


Figure 4.6. Theoretical calculation of k from (4.13) for different levels of errors to maintain the rendering quality ($\overline{ESD}$) at a constant value of 4.00

4.4 Experimental Validation

As discussed before, the main issue in quantitative analysis of LF rendering methods is the lack of ground truth data. To address this, we utilized our simulation system [96] to validate the optimization method proposed in this chapter. The details of the simulation model can be found in Appendix V. The simulator takes a 3D model of a scene and generates both reference cameras images and ground truth images. It also provides the depth maps for the following experiments. Controlled amount of depth map error is introduced to study how the rendering would be impacted when the depth map is noisy or inaccurate.

Figure 4.7 illustrates the UV-DM rendering quality for four depth map error levels $\frac{\Delta d}{d} = 5\%$, 10%, 15%, and 20% and for each error level, different k = 0.5, 1, 2, 5, and 10. Thus, 20 different combinations of UVDM$(d, \Delta d, k, l, |\omega|)$ are demonstrated. Rendering quality is reported in terms of PSNR. Four different 3D scenes were chosen and a regular camera grid based on the value of k was simulated. For each experiment, 1000 random virtual cameras were produced. Each reported PSNR is averaged among all 1000 virtual cameras for all four 3D scenes. Therefore 80,000 experiments were conducted by the simulator.

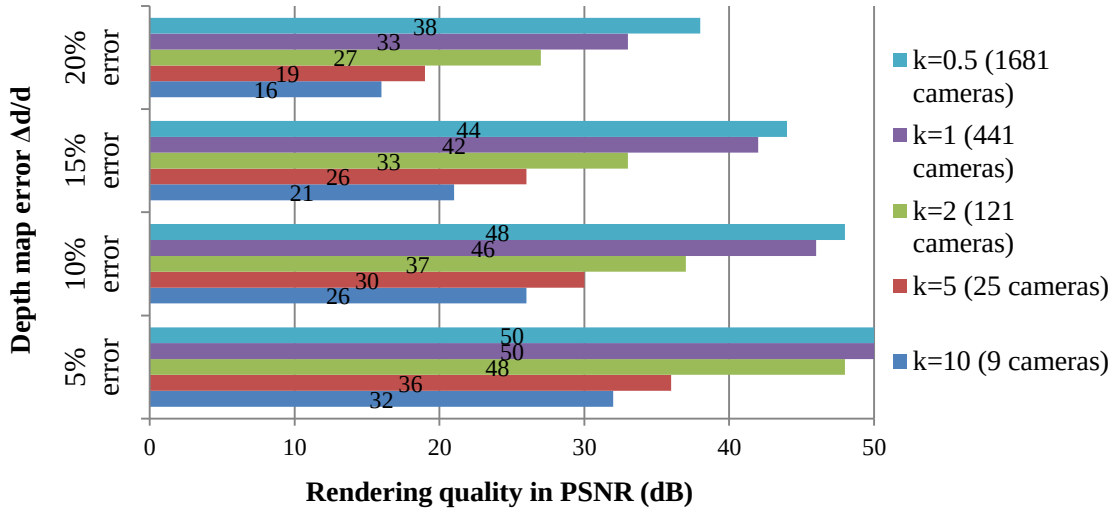


Figure 4.7. Experimental UVDM$(d, \Delta d, k, l, |\omega|)$ rendering quality in PSNR for $\frac{\Delta d}{d} = (5\%, 10\%, 15\% \text{ and } 20\%)$ and $k = (0.5, 1, 2, 5, \text{and } 10)$

As it can be seen in Figure 4.7, the same pattern expected from the proposed model is achieved, as smaller k improves the PSNR, e.g. $k = 1$ and 15% error in depth performs better than 10% error in depth with $k = 2$.

4.4.1 Rendering with Desired PSNR

Assume the desired rendering quality is given as an average PSNR value. This section shows how the proposed optimization model can be used to calculate k and amount of filtering required to produce the rendering quality at the desired PSNR value.

As shown in chapter 3, to be able to directly predict rendering PSNR from theoretical ESD, an empirical relationship between calculated ESD and rendering PSNR values has been established:

$$\mathrm{PSNR}_{\mathrm{UVDM}(d,\Delta d,k,l,|\omega|)} \cong 20\, log_{10} \frac{255}{\sqrt{Q.\,\mathrm{ESD}_{\mathrm{UVDM}(d,\Delta d,k,l,|\omega|)}{}^{P}}} \tag{4.14}$$

where $1 < Q < 15$ and $-0.9 < P < -0.2$

Equation (4.14) is employed to calculate corresponding ESD for a given PSNR value. Q and P for a given scene were approximated through experiments. Then (4.13) is applied to find the maximum k to maintain the ESD and the corresponding PSNR at a prescribed value (for instance 50 dB), as shown in Figure 4.8. Figure 4.8 also shows the average PSNR for fixed $k = 14.4$, calculated k is demonstrated in Figure 4.9.

Figure 4.8 shows that for high error rates, changing k using (4.13) results in significant improvements over the fixed camera density and can maintain the rendering quality around prescribed 50 dB. Figure 4.8 and Figure 4.9 are the experimental results corresponding to the theoretical predictions presented in Figure 4.5 and Figure 4.6.

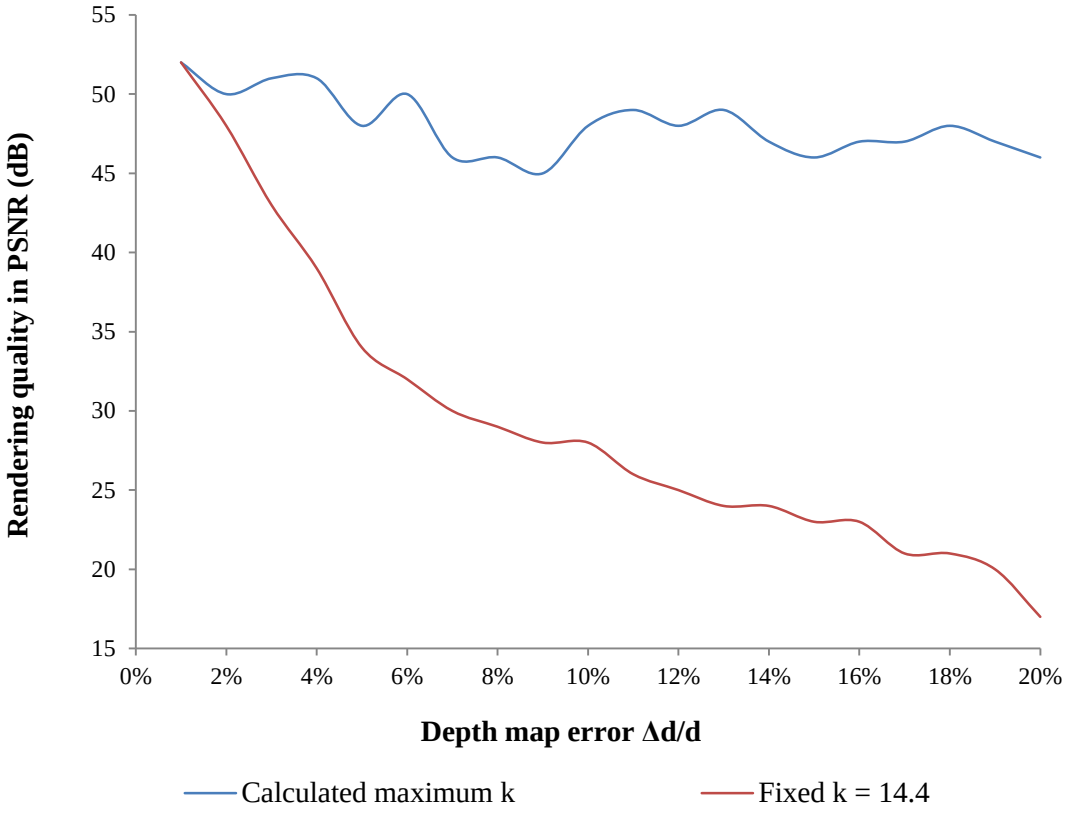


Figure 4.8. Experimental rendering quality for fixed $k = 14.4$ vs. calculated maximum k demonstrated in Figure 4.9 for different levels of depth errors

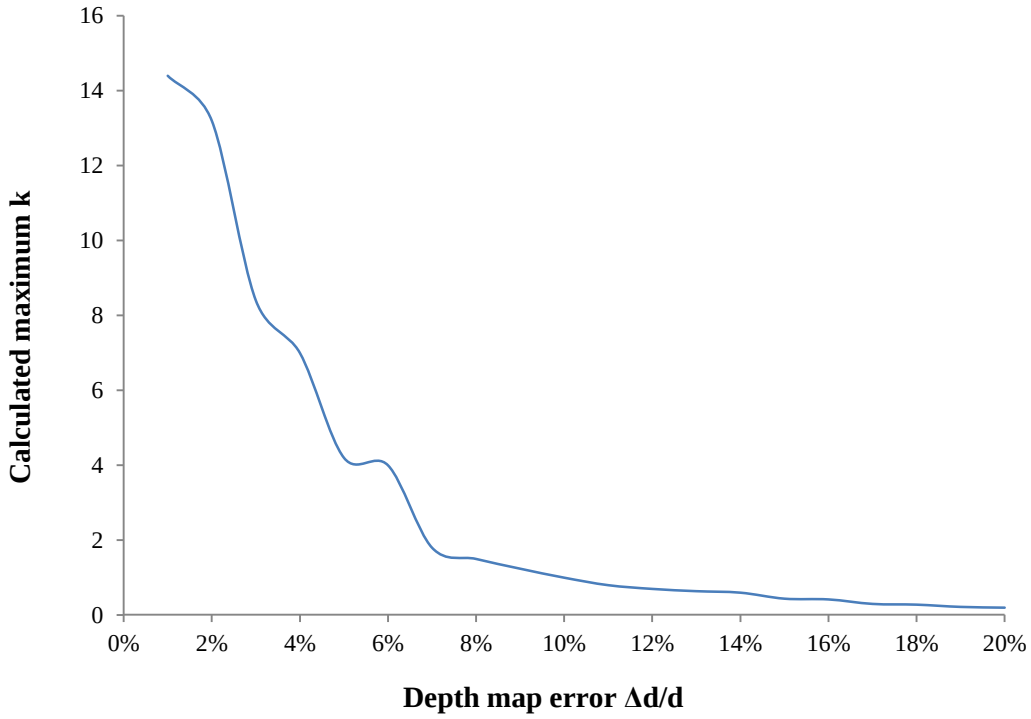


Figure 4.9. Maximum k from (4.13) and (4.14) to maintain the mean PSNR at a prescribed value of 50 dB for different levels of errors in depth estimation

4.5 Comparison with Other Methods

4.5.1 The Simplified Case of the Proposed Method

Table 4.1 shows the comparison of a simplified version of the proposed method based on ESD analysis, Chai's method based on spectral analysis and Lin's method based on geometric analysis. The simplification is aimed at matching the underlying assumptions of all models.

Table 4.1

Comparison of the proposed method with Chai's and Lin's methods

Method	*Original equation to compute k*	*Simplified equation I* [1]	*Simplified equation II* [2]
Chai's spectral analysis	$k = \frac{2lN_d}{\left(\frac{1}{d_{min}} - \frac{1}{d_{max}}\right)}$	$k = \frac{2ld_{min}d_{max}}{\overline{\Delta d}}$	$k \approx \frac{2l\bar{d}^2}{\overline{\Delta d}}$
Lin's geometric analysis	$k = \frac{l(N_d + 1)}{\left(\frac{1}{d_{min}} - \frac{1}{d_{max}}\right)}$	$k \approx \frac{ld_{min}d_{max}}{\overline{\Delta d}}$	$k \approx \frac{l\bar{d}^2}{\overline{\Delta d}}$
Proposed method based on ESD analysis	$k = \frac{l\left(\left(\sqrt{\frac{\lvert\omega\rvert}{n}} - 1\right)d^2 - d\Delta d\right)}{\Delta d(\sqrt{\lvert\omega\rvert} - 1)}$	$k = \frac{l\bar{d}^2}{\overline{\Delta d}} - l\bar{d}$	$k = \frac{l\bar{d}^2}{\overline{\Delta d}} - l\bar{d}$

[1] $\overline{\Delta d} = \frac{d_{max} - d_{min}}{N_d}$, $|\omega| = 4$ and $= 1$,
[2] $d_{min} \cong d_{max} \cong \bar{d}$

The second column of table 4.1 states the key equations of all three methods for calculating k. To compare these methods, Chai's and Lin's methods are rewritten based on Δd

instead of number of depth layers N_d. The third column, shows the modified Chai's and Lin's methods based on Δd and simplified version of the proposed method for Lambertian scene and 4 rays bilinear interpolation, i.e., $n = 1$ and $|\omega| = 4$. Fourth column illustrates another simplification when d_{min} and d_{max} are very close. As $l\bar{d}$ is typically small, it can be seen that the proposed method leads to a similar result as Lin's method, while Chai's is only different in a constant factor of 2 with other two methods. Therefore, simplified version of the proposed method is reduced to Chai's and Lin's methods, if several simplifications such as Lambertian scene, no under-sampling and 4-ray bilinear interpolation are adopted.

4.5.2 General Case Comparison

Section *A* shows that by applying several simplifications, the proposed method leads to the same results as Chai's and Lin's methods. However, for general case, the proposed method always results in less camera density. Figure 4.10 demonstrates the theoretical results of these three methods for the minimum number of cameras for an LF system with $l = 10^{-4}$, $d_{min} = 100$, $d_{max} = 600$, $n = 1$, $|\omega| = 4$, Δd in the range of $[1\% ,20\%]$, i.e., N_d in the range of $[25 ,500]$ and uv plane size is $50x50$.

Figure 4.10 shows that the proposed method always calculates much less number of cameras compared with both Chai's and Lin's methods. On average the proposed method calculated more than 5 times less than Chai's and more than 20 times less than Lin's. In Particular, for high errors in depth estimation, Chai's and Lin's results in a very high number of cameras compared with the proposed method.

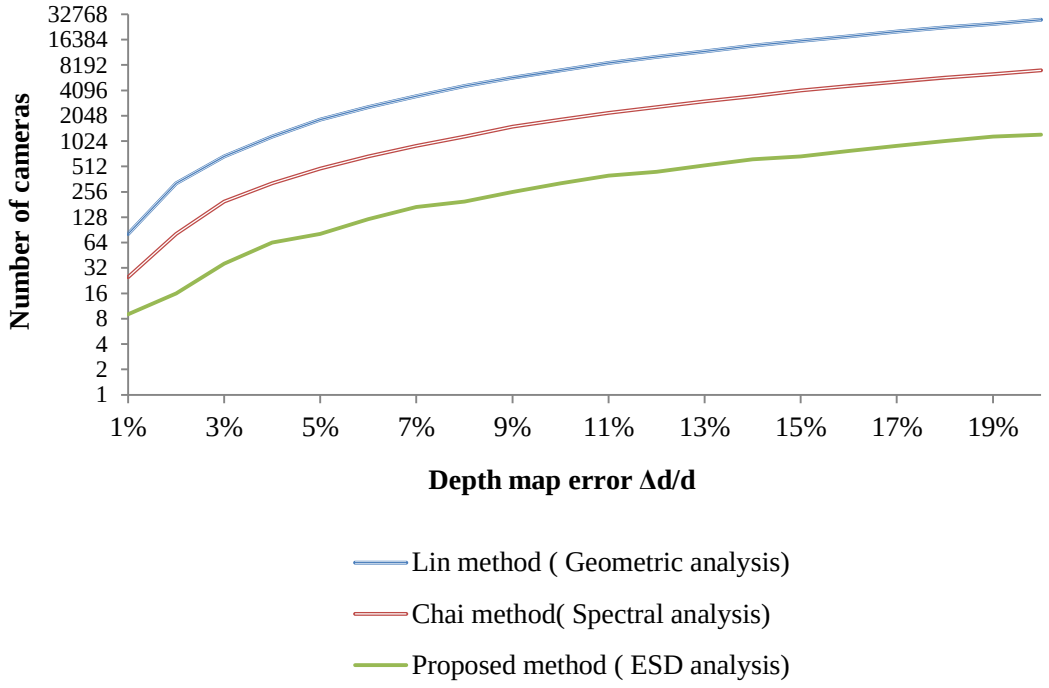


Figure 4.10. Theoretical minimum sampling rate in term of number of cameras, proposed method vs. Chai's and Lin's methods

Figure 4.11 shows the experimental rendering quality in term of PSNR for three different k calculated from these methods as demonstrated in Figure 4.10. Note that the simulation was limited to maximum 10% error in depth estimation.

As demonstrated in Figure 4.11, it is not surprising that the very small k computed by Lin's method always has the highest rendering quality followed by Chai's and the proposed method respectively. However, on average the rendering improvement of Chai's and Lin's methods over the proposed method is negligible as Chai's 5 times more and Lin's 20 times more number of cameras only result in less than 3% and 4% improvement in PSNR respectively.

This comparison shows how the proposed method is effective to calculate the minimum camera density for a desirable rendering quality with allowing controlled amount of under-sampling.

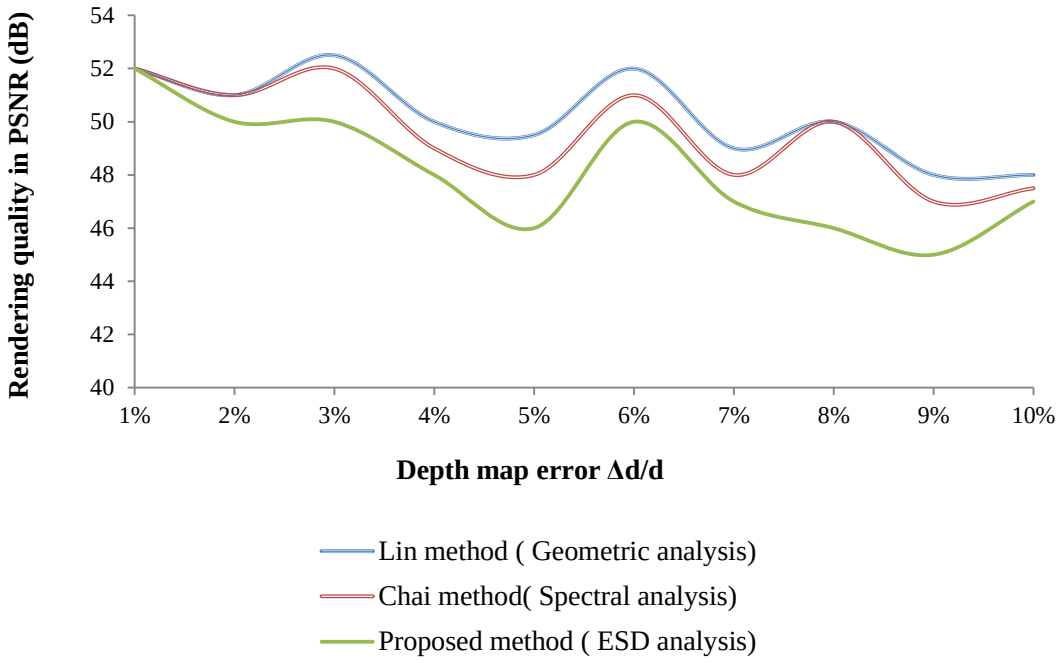


Figure 4.11. Experimental rendering quality for three different calculated number of cameras as shown in Figure 4.10.

4.6 Discussion and Conclusion

The theory of ESD is employed to optimize a regular grid LF acquisition. As a result, a method for calculating the minimum number of cameras and evaluating the impact of depth map errors on output quality for LF-based FVV systems is proposed in this chapter. It is shown that higher camera density can compensate for the adverse effect of depth map errors on the rendering quality. To employ the proposed method in LF based FVV system design, the desired rendering quality of the system in PSNR can be mapped to the corresponding ESD by employing the empirical model given as (4.14). This ESD with depth estimation error is applied to (4.13) to calculate the camera density in the grid and hence the minimum number of cameras.

Chapter 5: Rendering Optimization and Calculation of the Number of Rays in Interpolation

5.1 Summary

Light field (LF) rendering is widely used in free viewpoint video systems (FVV). Different methods have been proposed to employ depth maps to improve the rendering quality. However, estimation of depth is often error-prone. In this Chapter, a new method based on the concept of effective sampling density (ESD) is proposed for evaluating the depth-based LF rendering algorithms at different levels of errors in the depth estimation. In addition, for a given rendering quality, we provide an estimation of the rendering complexity in terms of optimum number of rays employed in interpolation algorithm so as to compensate for the adverse effect caused by errors in depth maps. The proposed method is particularly useful in designing a rendering algorithm with inaccurate knowledge of depth to achieve the required rendering quality. Both the theoretical study and numerical simulations have shown that the proposed method is reliable and accurate.

This rest of the chapter is organized as follows. Section 2 describes the proposed method by introducing LF rendering analysis based on ESD and optimization of the number of rays in interpolation by employing ESD. Experimental validation and simulation results are presented in Section 3. Section 4 concludes the chapter.

5.2 The Relationship between ESD and Number of Rays in |ω|

With the same assumptions as previous chapter (Chapter 4), the analysis for LF rendering evaluation and optimization is only given for a simplified LF-based FVV system with regular camera grid acquisition, 2-planes representation, symmetric ray selection, and by employing estimated depth maps during rendering for UV-DM method. However, the same approach can be used to employ ESD to analyse, evaluate and optimize different LF acquisition configurations and rendering methods.

In chapter 4, it is demonstrated that ESD for UVDM$(d, \Delta d, k, l, |\omega|)$ is proportional to $|\omega|$ and inversely proportional to k. It means higher camera density in camera grid (smaller k) and employing more rays for interpolation results in higher ESD. Chapter 4 demonstrated a method to optimize k as a main parameter of regular grid LF acquisition. This chapter presents the optimization of $|\omega|$, the number of rays selected by LF rendering method and employed in interpolation. $|\omega|$ is chosen as a candidate in this thesis to show how the ESD theory can be used for LF rendering evaluation and optimization. However, it is possible to generalize the approach and use ESD for any other types of rendering method evaluation and optimization.

The effect of $|\omega|$ on ESD has been previously demonstrated in chapter 4 (Figure 4.3). However the same Figure is demonstrated again here to emphasise the concept for further analysis. Figure 5.1 shows the theoretical calculation of mean $\overline{\mathrm{ESD}}$ for UVDM$(d, \Delta d, k, l, |\omega|)$, for a given light field system with regular camera grid with $k = 5$ and $l = 0.01$, average depth of scene $\bar{d} = 100$, relative depth map error $\frac{\Delta d}{d}$ between 0% to 100%, for three different values of $|\omega| = 4$, 16 and 32. Notice that the estimation error for depth map in most real application is around 10% to 20%.

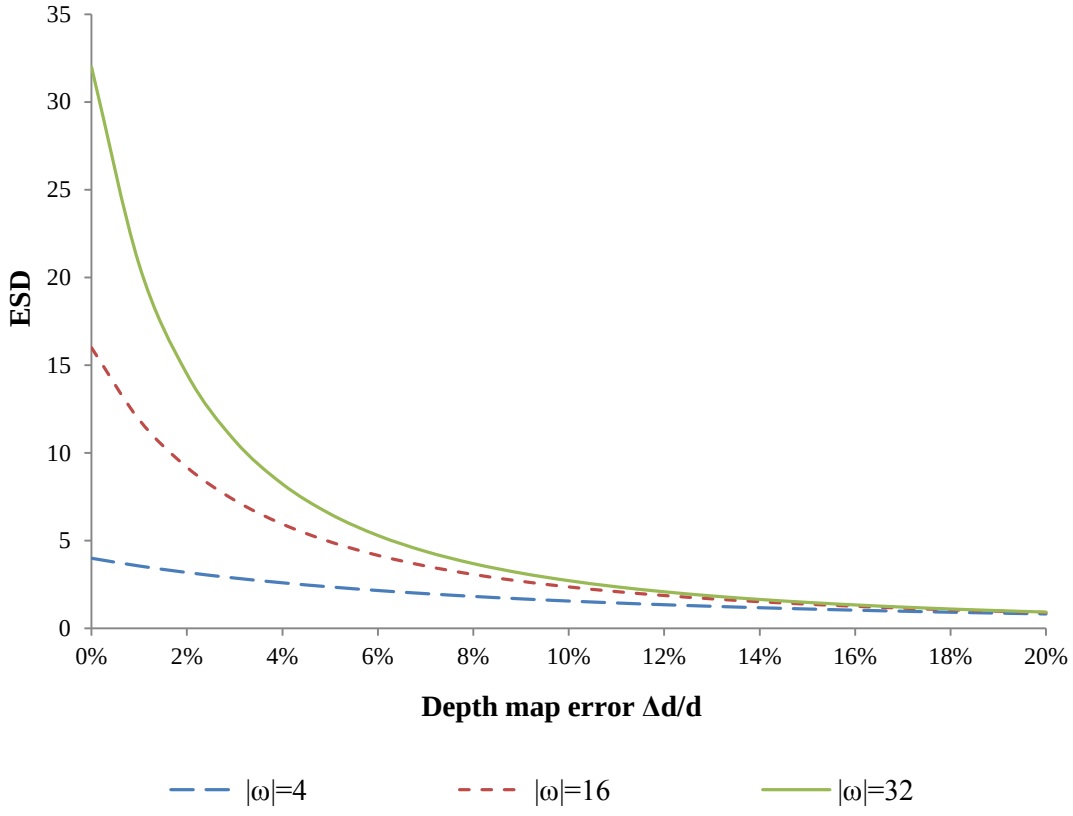


Figure 5.1. Theoretical $\overline{ESD}$ for UVDM$(d, \Delta d, k, l, |\omega|)$ for $\bar{d} = 100$, depth map with relative error $\frac{\Delta d}{d}$ in the range of $[0\%, 20\%]$, $l = 0.01$, $k = 5$, for $|\omega| = 4$, 16 and 32

As can be observed from Figure 5.1, higher errors in depth estimation result in less ESD and subsequently less rendering quality when $|\omega|$ is fixed. The reason is that error in depth Δd increases the area A for a given $|\omega|$ and therefore decreases ESD. However, choosing more rays for interpolation could increase the ESD and consequently rendering quality. For example, the ESD for 16-rays interpolation with errors less than 7% is still better than 4-rays interpolation with 1% error or ESD for 32-rays interpolation with errors less than 2% is still better than 16-rays interpolation with 1% error. However, for a very high level of errors in depth estimation, the ESDs in all three cases are declining rapidly to a very small value and consequently the rendering quality may become inadequate.

This analysis shows that increasing the number of rays for interpolation could compensate for the adverse effect of depth map estimation errors on ESD to some degree, at

least when the depth error is not very large. Of course, when more rays are employed in the interpolation, more computation is required. Thus in an LF rendering with a prior knowledge of the error bound in the depth map, the optimum number of rays |ω| could be calculated in advance to maintain the quality.

5.2.1 Optimization of |ω|

As discussed before, ESD is proportional to |ω|. On the other hand, the complexity of interpolation is increased significantly with large |ω|. Thus |ω| should be set at an optimum value to satisfy both the rendering quality and efficiency requirements. In this section, a theoretical minimum |ω| to compensate for the effect of errors in depth maps is derived. It is assumed that camera density is such that there is always enough number of rays in Ω to be used for interpolation.

Refer to chapter 4, $\mathrm{ESD}_{\mathrm{Ideal}}$ can be calculated as 5.1:

$$\mathrm{ESD}_{\mathrm{Ideal}} = \mathrm{ESD}_{\mathrm{UVDM}(d,0,k,l,n)} = \frac{n}{(ld)^2} \text{ and } n \geq 1 \tag{5.1}$$

where $n = 1$ is for the pure Lambertian reflection scene. Higher value of n can be used for non-Lambertian reflection.

Denote that this is corresponding to an ideal scenario, where there are no errors in depth map estimation and there is a depth map for each camera in the system, and depending on the complexity of reflectivity of surfaces in the scene, one or more rays would be enough for an accurate rendering.

So, the optimization problem is posed as follows: what would be the minimum |ω| (i.e., the minimum number of rays selected for interpolation by the *ray selection process M*) for any

given UVDM(d, Δd, k, l, |ω|) with known depth map error Δd to have the same ESD as the ideal case?

$$\mathrm{ESD}_{\mathrm{UVDM}(d,\Delta d,\mathrm{k,l},|\omega|)} = \mathrm{ESD}_{\mathrm{Ideal}} \rightarrow \frac{|\omega|}{\left(l(d+\Delta d)+\frac{\Delta d.k}{d}\left(\sqrt{|\omega|}-1\right)\right)^2} = \frac{n}{(ld)^2} \rightarrow$$

$$|\omega| = \left(\frac{l(d+\Delta d)-\frac{\Delta d.k}{d}}{\frac{ld}{\sqrt{n}}-\frac{\Delta d.k}{d}}\right)^2 \tag{5.2}$$

where $k < \frac{ld^2}{\Delta d\sqrt{n}}$

Equation (5.2) gives the minimum |ω| required for interpolation in rendering process to avoid quality deterioration due to errors in depth maps.

For the purpose of this chapter, it is assumed that available |Ω| and thus SD is always large enough to provide this minimum |ω| in each point of the scene.

It should be noted that ESD is a function of d, the depth of a point in the scene space. Hence, it has different values at different points of the scene. Therefore, typically for a given scenario, (5.2) is applied to the mean $\overline{\mathrm{ESD}}$ for the entire scene by assuming the average depth of the scene $\bar{d}$ and average error in depth $\overline{\Delta d}$ to calculate average $\overline{|\omega|}$. Employing $\overline{|\omega|}$ rays in interpolation, guarantees the scene to be sampled and rendered with average $\overline{\mathrm{ESD}_{\mathrm{Ideal}}}$.

If the design criteria requires the scene to be sampled and reconstructed by a minimum $\mathrm{ESD}_{\mathrm{Ideal}}$ instead of average $\overline{\mathrm{ESD}_{\mathrm{Ideal}}}$, (5.2) should be applied to all d ranging between (d_{min}, d_{max}), the minimum and maximum depths of the scene with corresponding Δd. This gives optimum |ω| for each depth d and the maximum |ω|can be chosen by *ray selection mechanism M* of a rendering method.

Figure 5.2 shows the same system demonstrated in Figure 5.1, but this time for any $\Delta d < 20\%$, |ω| is calculated directly from (5.2) to maintain $\overline{\mathrm{ESD}}$ at 4.00, the ideal ESD

calculated for $n = 4$. k is calculated as follows to satisfy the condition of (5.2): $k < \frac{0.01x100^2}{20\sqrt{4}} < 2.5 \rightarrow k = 2.2$. Figure 5.3 shows the actual number of rays |ω|, employed in interpolation in such a scenario. The corresponding point for 10% error in depth estimation is highlighted in Figure 5.2 and Figure 5.3, respectively, to show the relation of these two Figures.

Note that ESD cannot be increased indefinitely by only increasing |ω| because: a) The curve of ESD vs.|ω| is saturating as |ω| increases: $\lim_{|\omega|\rightarrow\infty}(\mathrm{ESD}_{\mathrm{UVDM}(d,\Delta d,k,l,|\omega|)}) = \left(\frac{d}{\Delta dk}\right)^2$, b) |ω| is bounded by |Ω| and cannot be increased indefinitely, i.e., ESD cannot be increased more than SD on any point of the scene because both |Ω| and thus SD are predetermined by the acquisition configuration, and c) Increasing |ω| would also increase the complexity of rendering/interpolation process significantly. Hence, in practice, the error in depth map can be compensated for by judicious alteration of *both* |ω| and k, i.e., higher rendering complexity and camera density.

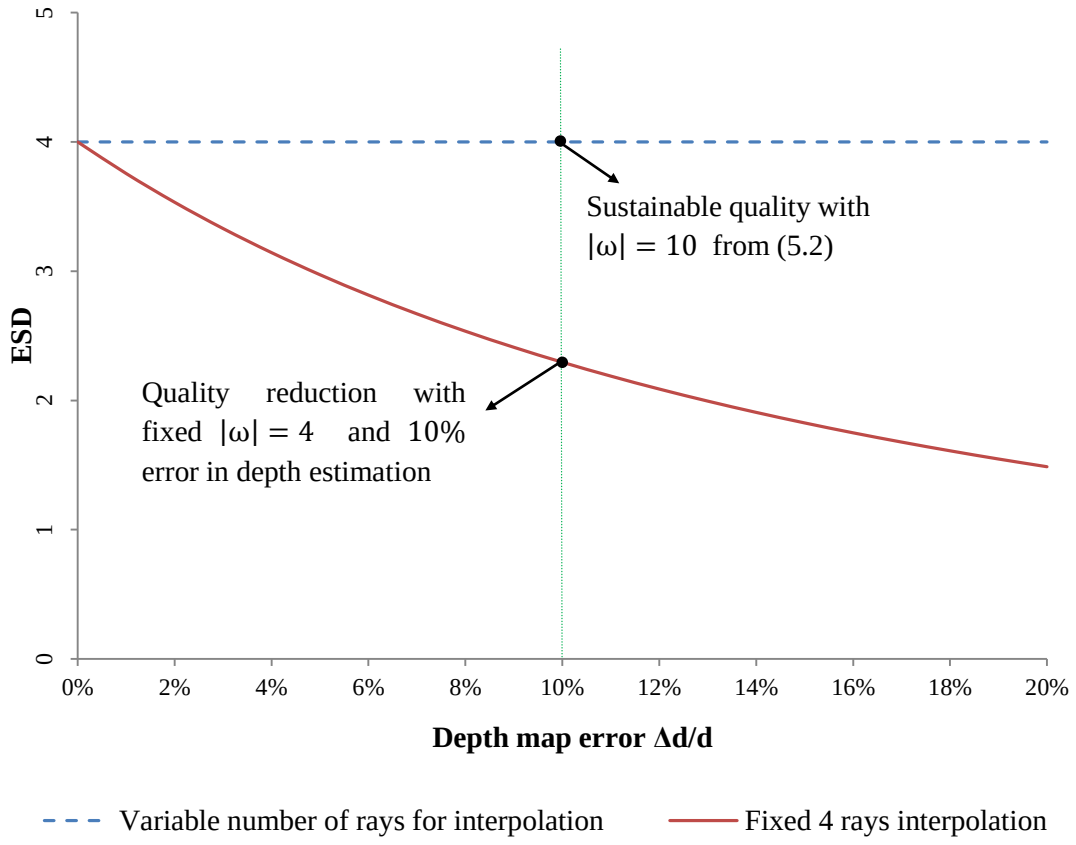


Figure 5.2 Theoretical impact of depth estimation error on rendering quality ($\overline{ESD}$) for fixed $|\omega| = 4$ and calculated $|\omega|$ from (5.2)

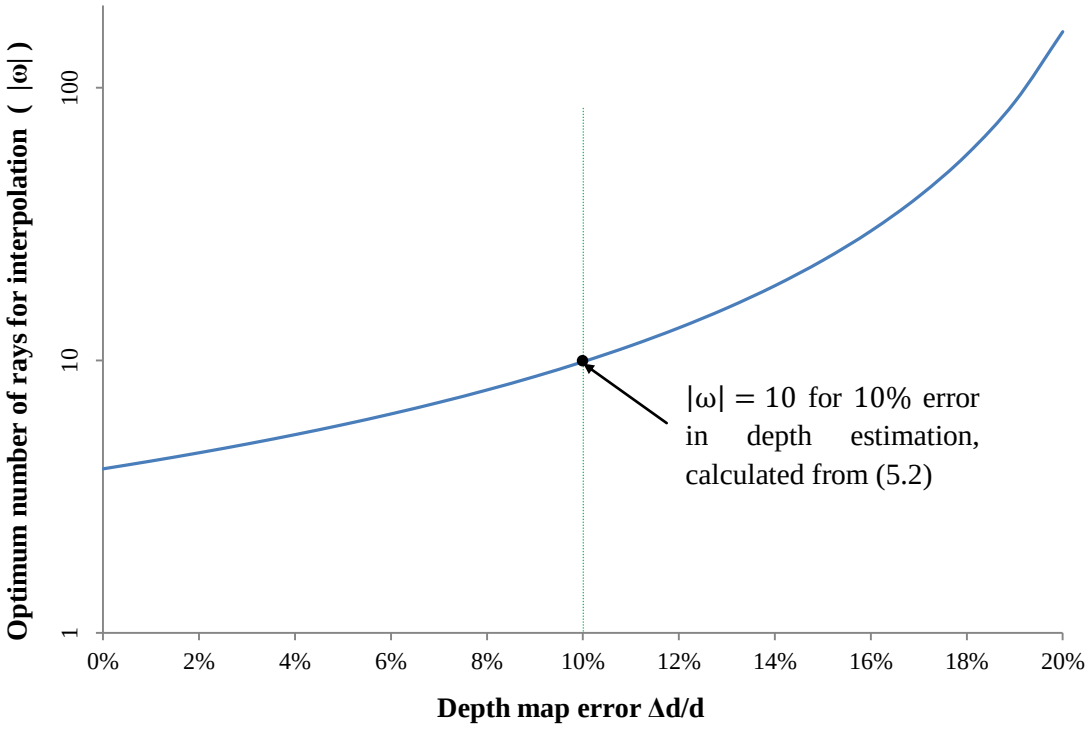


Figure 5.3. Theoretical calculation of $|\omega|$ form (5.2) for different levels of errors to maintain the rendering quality ($\overline{ESD}$) at a constant value of 4.00

5.3 Experimental Validation

The simulation system [96] is employed to validate the optimization method proposed in this chapter. The details of the simulation model can be found in Appendix V. The simulator takes a 3D model of a scene and generates both reference cameras images and ground truth images. It also provides the depth maps for the following experiments. Controlled amount of depth map error is introduced to study how the rendering would be impacted when the depth map is noisy or inaccurate.

Figure 5.4 illustrates the UV-DM rendering quality for four depth map error levels $\frac{\Delta d}{d} = 5\%, 10\%, 15\%, \text{and } 20\%$, and for each error level, different $|\omega| = 4, 9, 16, 25, \text{and } 36$. Thus, 20 different combinations of UVDM $(d, \Delta d, k, l, |\omega|)$ are demonstrated. Rendering quality is reported in terms of PSNR. Four different 3D scenes were chosen and a regular camera grid of $20x20$ was simulated as the LF acquisition component. For each experiment, 1000 random virtual cameras were produced. Each reported PSNR is averaged among 80,000 experiments for 1000 virtual cameras and four all 3D scenes.

As can be seen in Figure 5.4, the same pattern expected from the proposed model is achieved, i.e., increasing the number of rays in interpolation improves the PSNR, e.g. $|\omega| = 25$ and 15% error performs better than $|\omega| = 9$ and 10% error.

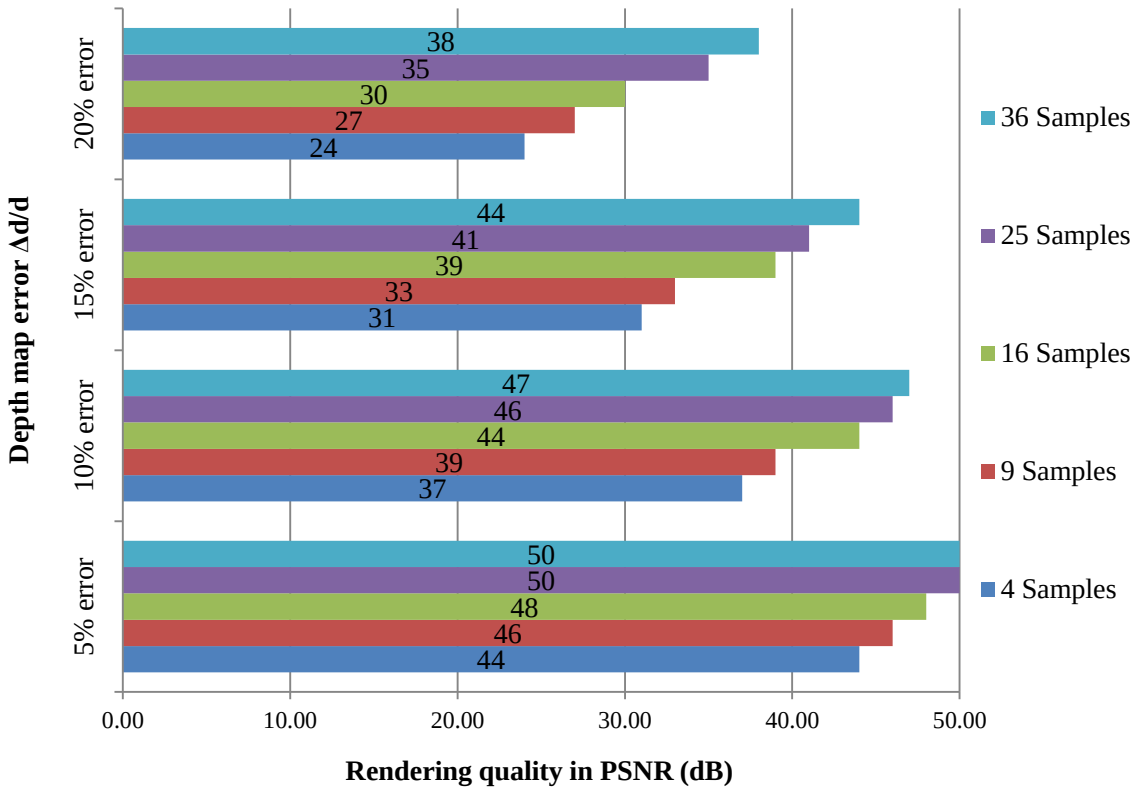


Figure 5.4. Experimental UVDM $(d, \Delta d, k, l, |\omega|)$ rendering quality in PSNR for $\frac{\Delta d}{d} = (5\%, 10\%, 15\%, \text{and } 20\%)$ and $|\omega| = (4, 9, 16, 25, \text{and } 36)$

5.3.1 Rendering with Desired PSNR

Assume the desired rendering quality is given as an average PSNR value. This section shows how the proposed optimization model can be used to calculate $|\omega|$ to produce the rendering quality at the desired PSNR value.

As shown in chapter 3, to be able to directly predict rendering PSNR from the theoretical ESD, an empirical relationship between calculated ESD and rendering PSNR values has been established:

$$\text{PSNR}_{\text{UVDM}(d,\Delta d,k,l,|\omega|)} \cong 20\, log_{10} \frac{255}{\sqrt{Q.\text{ESD}_{\text{UVDM}(d,\Delta d,k,l,|\omega|)}{}^{P}}} \tag{5.3}$$

where $1 < Q < 15$ and $-0.9 < P < -0.2$

Equation (5.3) is employed to calculate the corresponding ESD for a given PSNR value. Q and p for a given scene were approximated through experiments. Then (5.2) is applied to find the optimum number of rays $|\omega|$ to maintain the ESD and the corresponding PSNR at a prescribed value (for instance 50 dB), as shown in Figure 5.5. Figure 5.5 also shows the average PSNR for conventional fixed 4 rays interpolation, calculated number of rays $|\omega|$ is demonstrated in Figure 5.6.

Figure 5.5 shows that for high error rate, the use of optimum $|\omega|$ using (5.2) results in significant improvements over the conventional fixed 4 rays interpolation and can maintain the rendering quality around prescribed 50 dB. Figure 5.5 and Figure 5.6 are the experimental results corresponding to the theoretical predictions presented in Figure 5.2 and Figure 5.3.

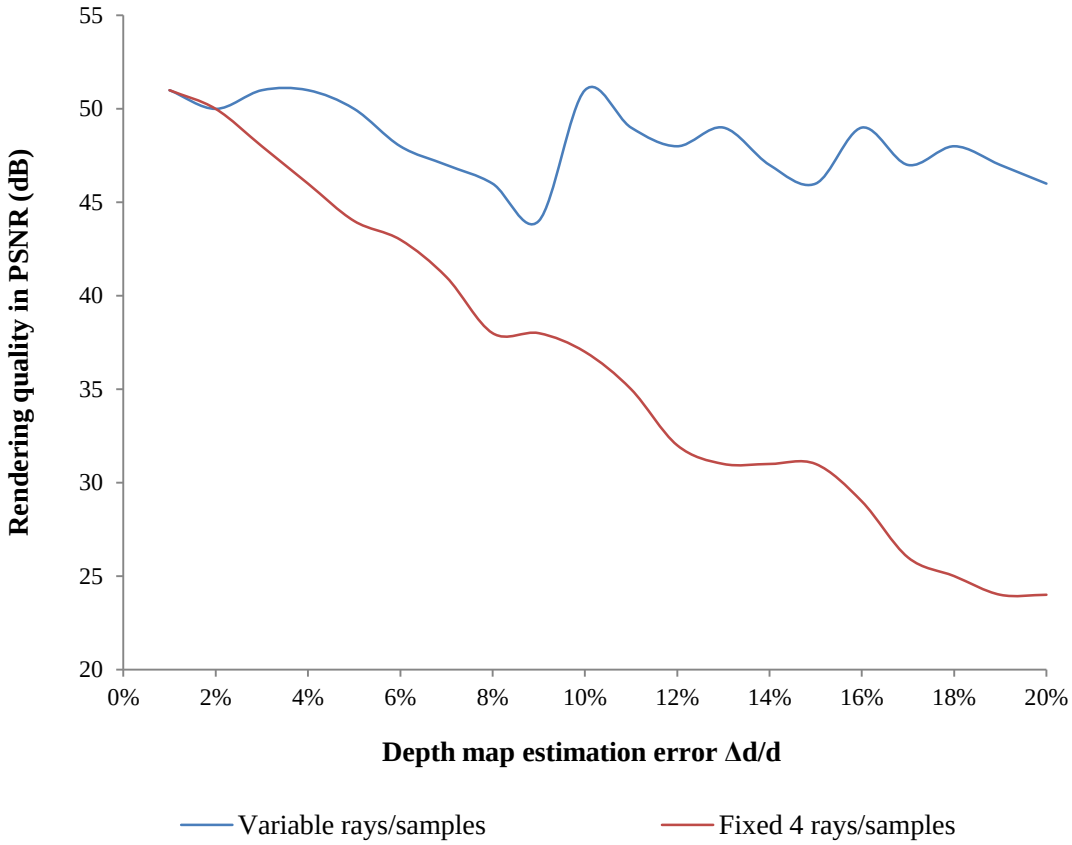


Figure 5.5. Experimental rendering quality for conventional fixed 4 rays interpolation ($|\omega| = 4$) vs. calculated optimum number of rays $|\omega|$ demonstrated in Figure 5.6 for different levels of depth errors

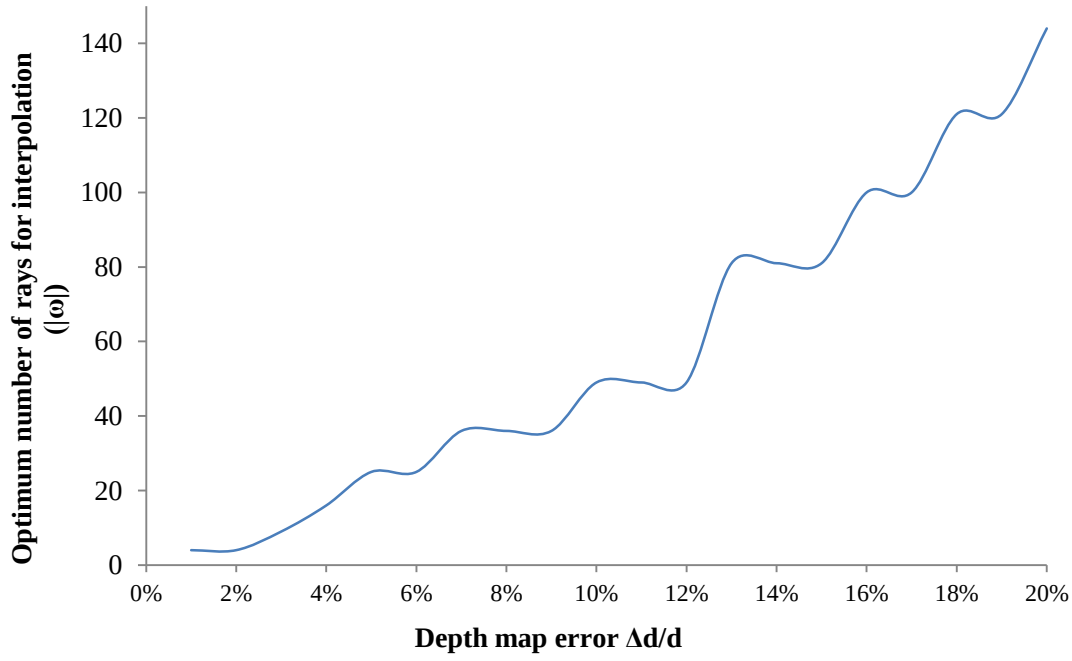


Figure 5.6. Optimum $|\omega|$ from (5.2) and (5.3) to maintain the mean PSNR at a prescribed value of 50 dB for different levels of errors in depth estimation

5.4 Discussion and Conclusion

In this Chapter a method is developed from the ESD model to calculate the optimum number of rays required for interpolation to compensate for the adverse effect of depth map errors on the rendering quality. To employ the proposed method in LF based FVV system design, the desired rendering quality of the system in PSNR can be mapped to the corresponding ESD by employing the empirical model given as (5.3). This ESD with depth estimation error is applied to (5.2) to calculate the optimum number of rays required for interpolation in rendering process.

Chapter 6: Joint Optimization of Acquisition and Rendering Subsystems by Applying Lagrangean Method to ESD

6.1 Summary

Quality of output video is an important usability objective in Free Viewpoint Video (FVV) systems. As shown in previous chapters, the density of a camera grid for acquisition on one hand and the complexity of the interpolation/rendering algorithm on the other hand directly influence the FVV output video quality. As shown before ESD is an analytically tractable metric that can be used to predict and evaluate the FVV video quality for a given acquisition and rendering. To increase ESD and hence improve the video quality, two parameters can be altered: (i) the density of cameras in the acquisition grid; and/or (ii) the density of rays within the interpolation area employed during rendering. While in chapters 4 and 5 individual optimizations of these parameters are demonstrated, in this chapter, we present a method to optimize these parameters jointly for a target output video quality using ESD. Study of a system with a regular camera grid has shown that the number of cameras can be reduced by 8 times if 32 rays, instead of 8 rays, are employed during rendering to achieve the similar rendering quality for a typical 20% error in depth estimation.

The rest of this chapter is organized as follows. Section 2 presents an overview of the problem discussed in this chapter. Section 3 reviews chapter 4 and 5 and how the system can be optimized individually with respect to the camera density and number of rays employed during interpolation. Section 4 describes the proposed joint optimization methods by applying Lagrangean method to ESD in a realistic context that error in depth maps is inevitable. Section 5 concludes the chapter.

6.2 Overview of the Problem

Let's consider the situation depicted in Figure 6.1. The unknown ray r intersects the scene at point p. To estimate r, the *ray selection process* will choose a number of rays captured by cameras within the interpolation neighbourhood of r denoted by A. In this Figure, the SD in vicinity of r is shown to be $4/A$ (because four rays, R_1, R_2, R_3, R_4 in this neighbourhood are captured by cameras in the acquisition grid). However, the ray selection process may only select the rays from the closest cameras to r in the uv (camera) plane. In this case only R_1 and R_2 are selected for interpolation, which reduces the effective sampling density, ESD, to $2/A$. If higher quality output is required, then the ESD has to increase. One way to do this is to keep the same *ray selection process* and reduce the spacing between the cameras k in the acquisition grid. This will result in higher acquisition cost. The second approach is to use more advanced algorithms for *ray selection* and select more rays in A (in this example R_3 and R_4).

Of course the above discussion assumes that the depth of point p is exactly known. As shown in chapter 3, ESD is also affected by the inaccuracy in estimation of depth, denoted by Δd, because the *ray selection process* may select rays that are not the closest captured rays to r. This has the effect of increasing the area of interpolation A and reducing ESD.

Chapter 4 proposed an individual optimization for camera density k, by focusing on the first approach and chapter 5 proposed an individual optimization for the number of rays selected for interpolation $|\omega|$, by considering the second approach. However, as demonstrated in chapters 4 and 5, individual optimizations of k and $|\omega|$ have limitations. In addition, by analyzing the conditions of equations (4.13) and (5.2), it is clear that individual optimizations of these two parameters are strongly inter-related, i.e., individual optimization based on k requires a minimum $|\omega|$ and vice versa. Hence, the optimization of an LF-based FVV system should utilize both approaches to give an optimum output quality. In other words, acquisition component and rendering component should be optimized jointly for a target output quality.

This chapter studies the trade off among acquisition camera density, ray selection, depth error and rendering quality using the concept of ESD and presents methods to optimize these parameters jointly for a system with a desired output quality in terms of ESD or Peak-to-Signal Noise Ratio (PSNR).

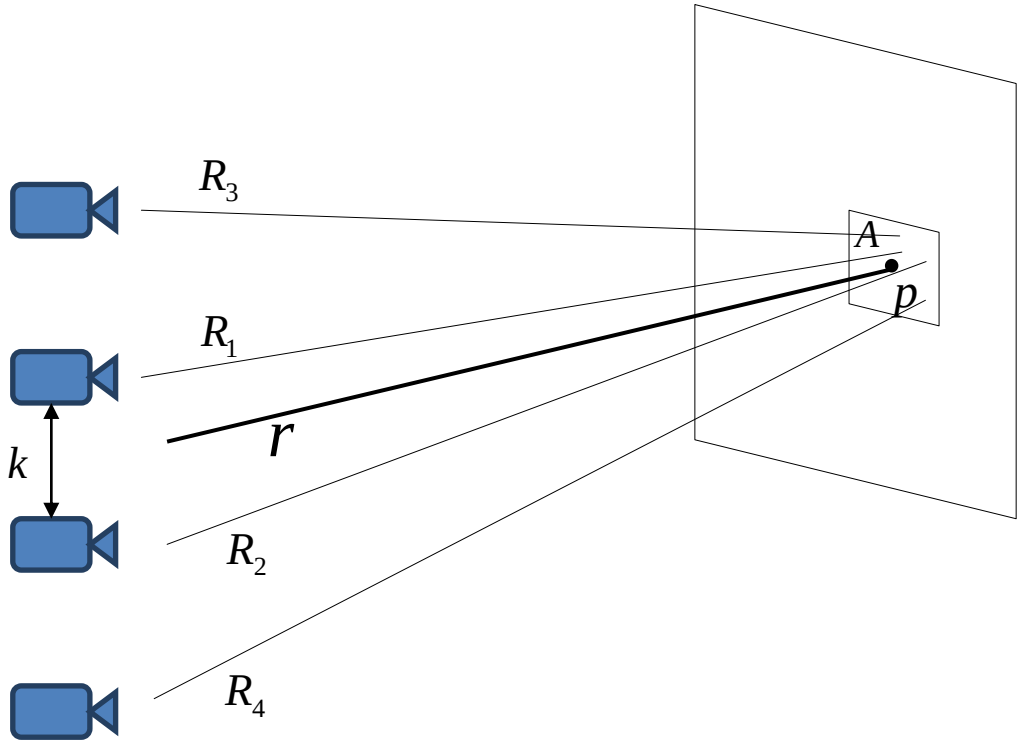
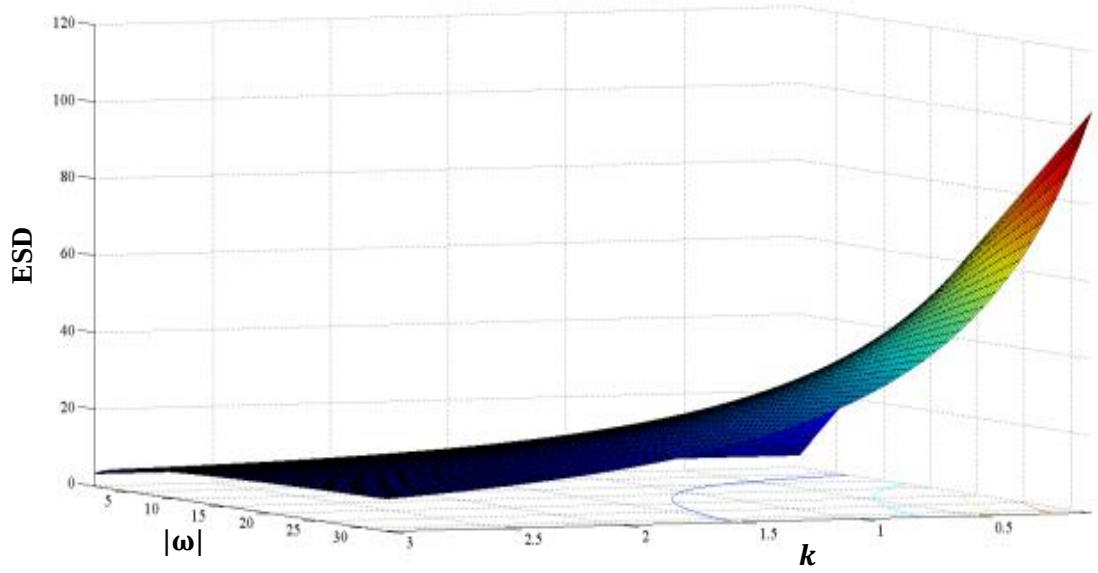


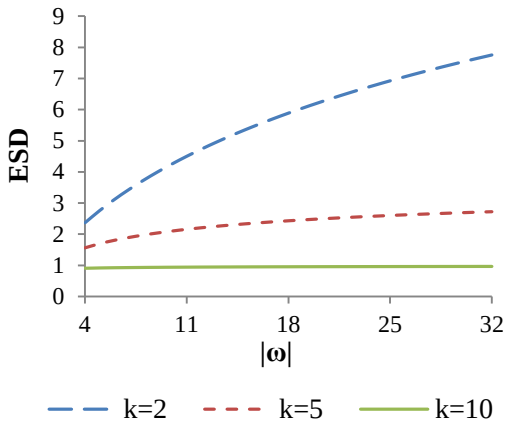
Figure 6.1. The effect of acquisition camera density and ray selection process on ESD

6.2.1 Relation between ESD, Depth Estimation Error Δd, k and $|\omega|$

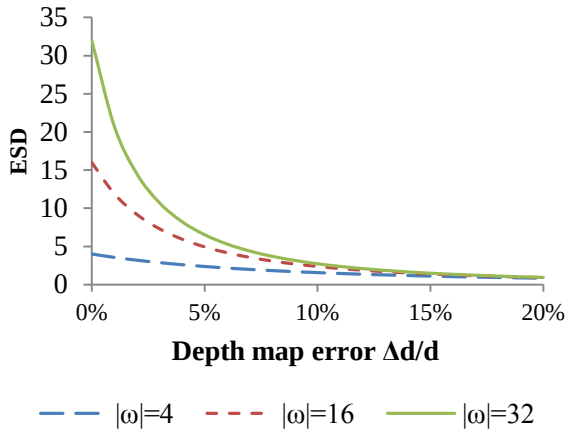
Figure 6.2 shows how ESD can be influenced by k, $|\omega|$, and Δd. Figure 6.2.a shows a 3D surface of ESD for different k and $|\omega|$. It is assumed that $\bar{d} = 100$, $\overline{\Delta d} = 10$ and $l = 0.01$. Figure 6.2.c and Figure 6.2.e demonstrate the effect of depth estimation error $\overline{\Delta d}$ (in the range of $[0\%, 20\%]$) on ESD and how k and $|\omega|$ can compensate for Δd to some extent. Notice that in each of these figures, one parameter is fixed and the effect of the other parameter on ESD is shown. Figures 6.2.b and 6.2.d illustrate a few 2D slices of the ESD surface demonstrated in 6.2.a in $|\omega|$ and k directions, respectively.



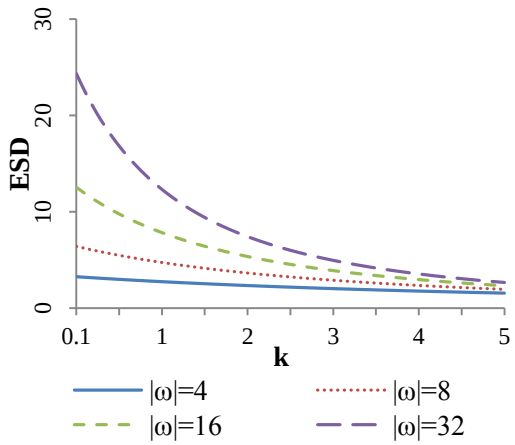
a) ESD 3D surface for different k and |ω|, for $\bar{d} = 100$, $\overline{\Delta d} = 10$ and $l = 0.01$



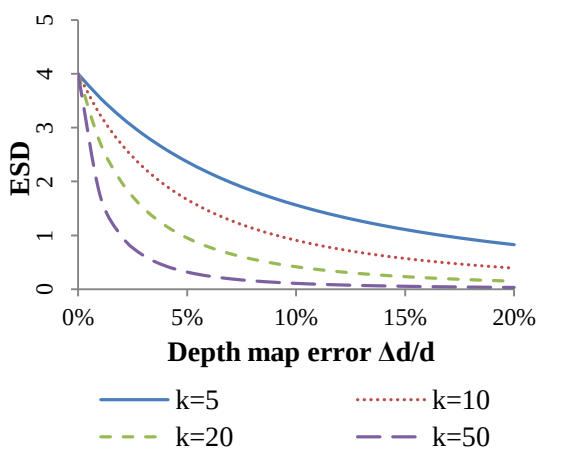
b) ESD for $\bar{d} = 100$, $\overline{\Delta d} = 10$, $l = 0.01$, |ω| in the range of $[4, 32]$, for $k = 2, 5$ and 10



c) ESD for $\bar{d} = 100$, $\overline{\Delta d}$ in the range of $[0\%, 20\%]$, $l = 0.01$, $k = 5$, for $|\omega| = 4, 16$ and 32



d) ESD for $\bar{d} = 100$, $\overline{\Delta d} = 10$, $l = 0.01$, k in the range of $[0.1, 5]$, for $|\omega| = 4, 8, 16$ and 32



e) ESD for $\bar{d} = 100$, $\overline{\Delta d}$ in the range of $[0\%, 20\%]$, $l = 0.01$, $|\omega| = 4$, for $k = 5, 10, 20$ and 50

Figure 6.2. Analysis of ESD based on k, |ω|, and $\overline{\Delta d}$

6.3 Review of the System Optimization based on Individual Variables

This section demonstrates a summary of the results from chapter 4 and 5 on optimization with respect to the camera density k and number of rays chosen for rendering $|\omega|$ individually.

6.3.1 Overview of Optimization of Camera Density k

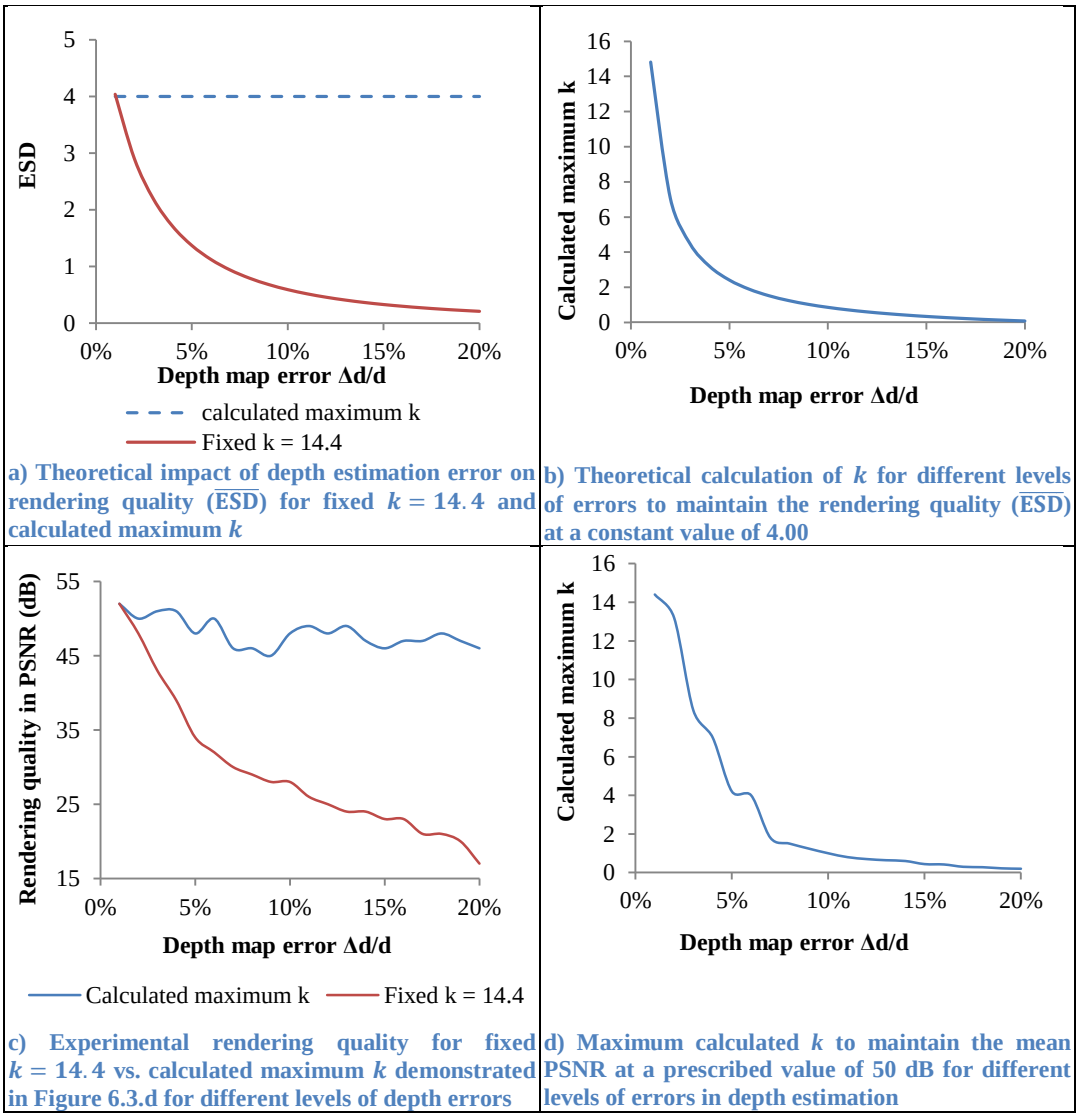
Recall from chapter 4 that the equation to calculate the maximum k for $\mathrm{ESD}_{\mathrm{UVDM}(d,\Delta d,k,l,|\omega|)}$ to avoid quality deterioration due to errors in depth maps is obtained as follows:

$$k = \frac{l\left(\left(\sqrt{\frac{|\omega|}{n}}-1\right)d^2 - d\Delta d\right)}{\Delta d\left(\sqrt{|\omega|}-1\right)} \tag{6.1}$$

where $\Delta d > 0, |\omega| > n\left(\frac{d+\Delta d}{d}\right)^2$ and n refers to scene reflection complexity ($n = 1$ is for a pure Lambertian reflection and higher value of n could be used to model non-Lambertian reflections).

Figure 6.3 demonstrates a summary of theoretical expectations and experimental results presented in chapter 4 (Illustrated before as Figures 4.5, 4.6, 4.8 and 4.9). Figure 6.3.a and Figure 6.3.b show the theoretical expectations for this optimization model. For any given depth estimation error $\Delta d \leq 20\%$, k is calculated directly from (6.1) to maintain $\overline{\mathrm{ESD}}$ at 4.00, the ideal ESD calculated for $n = 4$ and $\Delta d = 0$ ($\mathrm{ESD}_{\mathrm{Ideal}} = \mathrm{ESD}_{\mathrm{UVDM}(d,0,k,l,n)} = \frac{n}{(ld)^2}$ and $n \geq 1$). Figure 6.3.a demonstrates the ESD for fixed $k = 14.4$ and optimum k calculated from (6.1). Figure 6.3.b shows the calculated k in such a scenario. Figures 6.3.c and 6.3.d are the experimental results corresponding to the theoretical predictions presented in Figures 6.3.a and

6.3.b. As it can be seen the rendering PSNR is maintained at a prescribed value (for instance 50 dB) with calculated k in contrast with the average PSNR for fixed $k = 14.4$. Figure 6.3 shows that for high error rates, changing k using (6.1) results in significant improvements over the fixed camera density and can maintain the rendering quality around the prescribed 50 dB.



a) Theoretical impact of depth estimation error on rendering quality ($\overline{ESD}$) for fixed $k = 14.4$ and calculated maximum k

b) Theoretical calculation of k for different levels of errors to maintain the rendering quality ($\overline{ESD}$) at a constant value of 4.00

c) Experimental rendering quality for fixed $k = 14.4$ vs. calculated maximum k demonstrated in Figure 6.3.d for different levels of depth errors

d) Maximum calculated k to maintain the mean PSNR at a prescribed value of 50 dB for different levels of errors in depth estimation

Figure 6.3. Summary of theoretical and experimental optimization of k (camera density) based on ESD

6.3.2 Overview of Optimization with Respect to |ω|

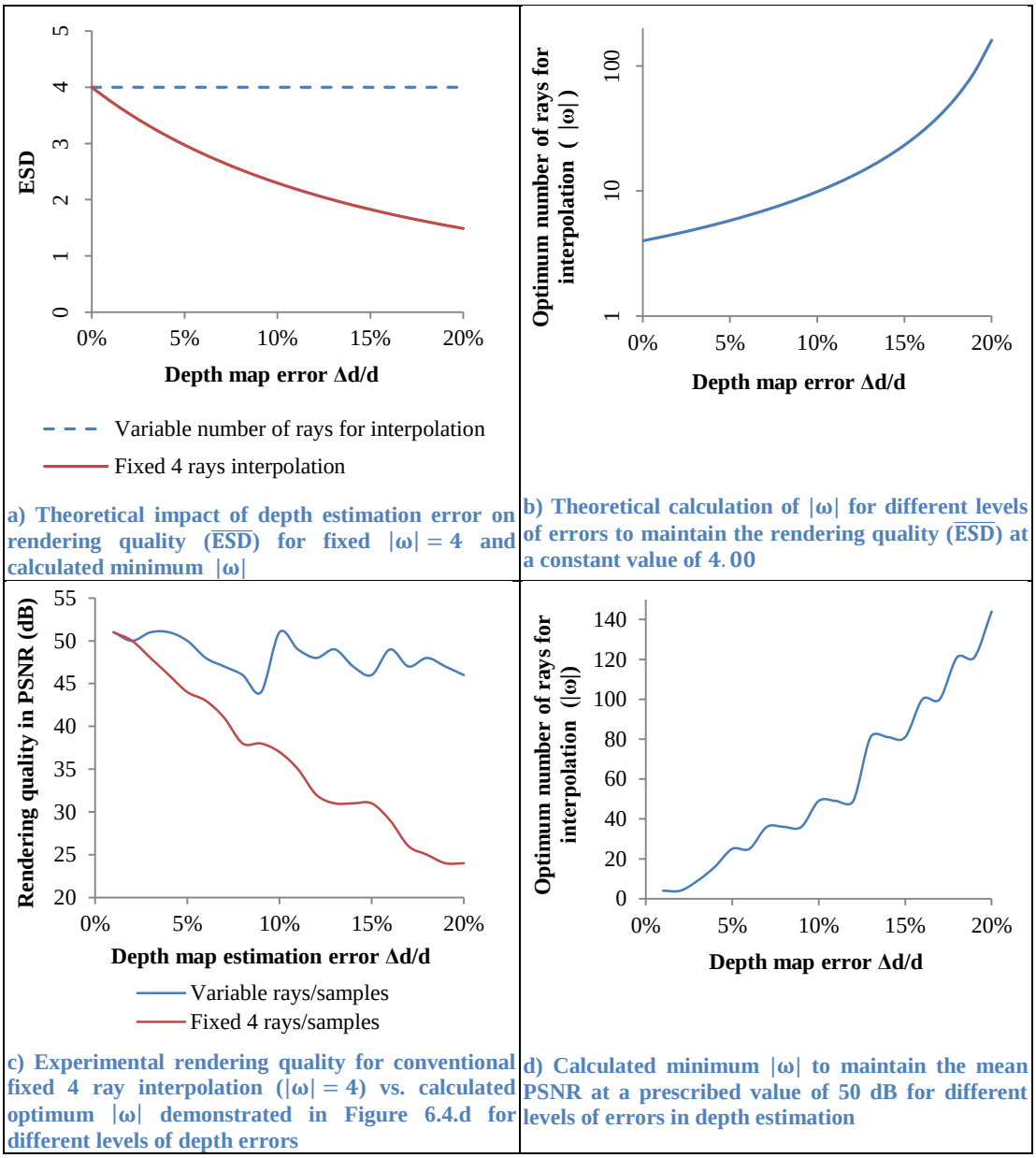
In reference to chapter 5, the optimum |ω| for $\mathrm{ESD}_{\mathrm{UVDM}(d,\Delta d,k,l,|\omega|)}$ to avoid quality deterioration due to errors in depth maps can be derived as follows:

$$|\omega| = \left(\frac{l(d+\Delta d) - \frac{\Delta d.k}{d}}{\frac{ld}{\sqrt{n}} - \frac{\Delta d.k}{d}} \right)^2 \tag{6.2}$$

where $k < \frac{ld^2}{\Delta d\sqrt{n}}$

For the purpose of the optimization, it is assumed that available |Ω| and thus SD is always large enough to provide this minimum |ω| at each point of the scene.

Figure 6.4 demonstrates a summary of theoretical expectations and experimental results presented in chapter 5 (Illustrated before as Figures 5.2, 5.3, 5.5 and 5.6). Figures 6.4.a and 6.4.b show the theoretical expectations for this optimization model. For any given depth estimation error $\Delta d < 20\%$, |ω| is calculated directly from (6.2) to maintain $\overline{\mathrm{ESD}}$ at 4.00, the ideal ESD calculated for $n = 4$. Figure 6.4.a demonstrates the ESD for fixed 4 ray interpolation and for optimum number of rays calculated from (6.2). Figure 6.4.b shows the calculated number of rays |ω|, employed in interpolation in such a scenario. Figures 6.4.c and 6.4.d are the experimental results corresponding to the theoretical predictions presented in Figures 6.4.a and 6.4.b. As it can be seen, the rendering PSNR is maintained at a prescribed value (for instance 50 dB) with calculated optimum number of rays |ω| in contrast with the average PSNR for conventional fixed 4 ray interpolation. Figure 6.4 shows that for high level of error in depth, the use of optimum |ω| using (6.2) results in significant improvements over the conventional fixed 4 ray interpolation and can maintain the rendering quality around the prescribed 50 dB.



a) Theoretical impact of depth estimation error on rendering quality ($\overline{ESD}$) for fixed $|\omega| = 4$ and calculated minimum $|\omega|$

b) Theoretical calculation of $|\omega|$ for different levels of errors to maintain the rendering quality ($\overline{ESD}$) at a constant value of 4.00

c) Experimental rendering quality for conventional fixed 4 ray interpolation ($|\omega| = 4$) vs. calculated optimum $|\omega|$ demonstrated in Figure 6.4.d for different levels of depth errors

d) Calculated minimum $|\omega|$ to maintain the mean PSNR at a prescribed value of 50 dB for different levels of errors in depth estimation

Figure 6.4. Summary of theoretical and experimental optimization of $|\omega|$ (number of rays employed in interpolation) based on ESD

6.3.3 Limitations of Individual Optimizations

As discussed in previous chapters, the above optimization equations are applied to the mean $\overline{\text{ESD}}$ for the entire scene by assuming the average depth of the scene $\bar{d}$ and average error in depth estimation $\overline{\Delta d}$ to calculate average $\overline{|k|}$ or $\overline{|\omega|}$. If the design criteria requires the scene to be sampled and reconstructed by a minimum $\text{ESD}_{\text{Ideal}}$ instead of average $\overline{\text{ESD}_{\text{Ideal}}}$,

optimization equations should be applied to all d ranging between (d_{min}, d_{max}) , the minimum and maximum depths of the scene with corresponding Δd.

6.3.3.1 Limitations of optimization based on $|\omega|$

Note that ESD cannot be increased indefinitely by only increasing $|\omega|$ because:

a) The curve of ESD vs. $|\omega|$ is saturating as $|\omega|$ increases: $\lim_{|\omega|\to\infty}(\mathrm{ESD}_{\mathrm{UVDM}(d,\Delta d,k,l,|\omega|)}) = \left(\frac{d}{\Delta d k}\right)^2$,

b) $|\omega|$ is bounded by $|\Omega|$ and cannot be increased indefinitely, i.e., ESD cannot be increased to over SD at any point of the scene because both $|\Omega|$ and thus SD are predetermined by the acquisition configuration, and

c) Increasing $|\omega|$ would also increase the complexity of interpolation process significantly.

6.3.3.2 Limitations of Optimization based on k

ESD cannot be increased indefinitely by only decreasing k because the curve of ESD vs. k is saturating as k decreases: $\lim_{k\to 0}(\mathrm{ESD}_{\mathrm{UVDM}(d,\Delta d,k,l,|\omega|)}) = \frac{|\omega|}{(ld+l\Delta d)^2}$.

Hence, in practice, the error in depth map can be compensated for by judicious alteration of *both* $|\omega|$ and k, i.e., rendering complexity and camera density. This demonstrates the desirability of a joint optimization model to calculate *both* $|\omega|$ and k. Next section provides an analytical ESD joint optimization model based on the Lagrangean method.

6.4 Joint optimization of |ω| and k

The gradient of $\mathrm{ESD}_{\mathrm{UVDM}(d,\Delta d,k,l,|\omega|)}$ can be expressed as:

$$\nabla\left(\mathrm{ESD}_{\mathrm{UVDM}(d,\Delta d,k,l,|\omega|)}\right) = \left(\frac{\partial \mathrm{ESD}_{\mathrm{UVDM}(d,\Delta d,k,l,|\omega|)}}{\partial |\omega|}, \frac{\partial \mathrm{ESD}_{\mathrm{UVDM}(d,\Delta d,k,l,|\omega|)}}{\partial k}\right)$$

$$= \left(\frac{\left(1 - \frac{\Delta d.k\sqrt{|\omega|}}{d\left(l(d+\Delta d) + \frac{\Delta d.k}{d}\left(\sqrt{|\omega|}-1\right)\right)}\right)}{\left(l(d+\Delta d) + \frac{\Delta d.k}{d}\left(\sqrt{|\omega|}-1\right)\right)^2}, \frac{-2\Delta d|\omega|\left(\sqrt{|\omega|}-1\right)}{d\left(l(d+\Delta d) + \frac{\Delta d.k}{d}\left(\sqrt{|\omega|}-1\right)\right)^3} \right) \quad (6.3)$$

To optimize $\mathrm{ESD}_{\mathrm{UVDM}(d,\Delta d,k,l,|\omega|)}$ with respect to both $|\omega|$ and k, a Lagrangean optimization method can be used. Assume that the minimum required rendering quality as $\overline{\mathrm{ESD}}$ is given as T. In individual optimizations of k and $|\omega|$, it was assumed that $T = \frac{n}{(ld)^2}$,the ideal ESD with no error in depth estimation ($\Delta d = 0$) by employing n rays during interpolation, where n refers to the scene reflection complexity ($n = 1$ is for a pure Lambertian reflection and higher value of n could be used to represent non-Lambertian reflections). But in general we can assign any value to T based on the design requirements. The optimization problem can then be formulated as:

$$\mathrm{ESD}_{\mathrm{UVDM}(d,\Delta d,\mathrm{k},\mathrm{l},|\omega|)} \geq T \Rightarrow \frac{|\omega|}{\left(l(d+\Delta d) + \frac{\Delta d.k}{d}\left(\sqrt{|\omega|}-1\right)\right)^2} \geq T \quad (6.4)$$

Note that the acquisition cost is directly determined by the number of cameras. For a 2D regular camera grid, this cost is proportional to $\frac{1}{k^2}$ where k determines the camera density. The cost function to express the combined rendering and acquisition cost can be defined as:

$$C(k,|\omega|) = \frac{C_k}{k^2} + C_\omega|\omega| \tag{6.5}$$

where C_k refers to the cost associated with each camera in camera grid and C_ω is the rendering computational cost associated with $|\omega|$, the number of rays employed in interpolation, i.e., C_k is a hardware configuration cost and C_ω is a software complexity cost. Note that, in this thesis a linear cost function is used. However, the same optimization method can be extended to other cost functions.

This optimization problem can be solved by a Lagrangean method by rewriting the optimization problem with the help of Lagrange multipliers as:

$$\begin{aligned}\Lambda(k,|\omega|,\lambda) &= C(k,|\omega|) + \lambda\left(\mathrm{ESD}_{\mathrm{UVDM}(d,\Delta d,\mathrm{k,l},|\omega|)} - T\right)\\ &= \frac{C_k}{k^2} + C_\omega|\omega| + \lambda\left(\frac{|\omega|}{\left(l(d+\Delta\mathrm{d}) + \frac{\Delta\mathrm{d.\,k}}{\mathrm{d}}\left(\sqrt{|\omega|}-1\right)\right)^2} - T\right)\end{aligned} \tag{6.6}$$

The optimum k and $|\omega|$ should satisfy (6.7),

$$\nabla_{k,|\omega|,\lambda}\Lambda(k,|\omega|,\lambda) = 0 \Rightarrow \begin{cases}\dfrac{\partial\Lambda(k,|\omega|,\lambda)}{\partial k} = 0\\ \dfrac{\partial\Lambda(k,|\omega|,\lambda)}{\partial|\omega|} = 0\\ \dfrac{\partial\Lambda(k,|\omega|,\lambda)}{\partial\lambda} = 0\end{cases} \tag{6.7}$$

By expanding these equations we have:

$$\frac{\partial\Lambda(k,|\omega|,\lambda)}{\partial k} = \frac{-2C_k}{k^3} + \frac{-2\lambda\Delta d|\omega|\left(\sqrt{|\omega|}-1\right)}{d\left(l(d+\Delta d) + \frac{\Delta d.\,k}{d}\left(\sqrt{|\omega|}-1\right)\right)^3} = 0 \tag{6.8}$$

$$\frac{\partial \Lambda\,(k,|\omega|,\lambda)}{\partial|\omega|} = C_\omega + \frac{\lambda\left(1 - \frac{\Delta d.\,k\sqrt{|\omega|}}{d\left(l(d+\Delta d) + \frac{\Delta d.\,k}{d}\left(\sqrt{|\omega|}-1\right)\right)}\right)}{\left(l(d+\Delta d) + \frac{\Delta d.\,k}{d}\left(\sqrt{|\omega|}-1\right)\right)^2} = 0 \tag{6.9}$$

$$\frac{\partial \Lambda\,(k,|\omega|,\lambda)}{\partial\lambda} = \frac{|\omega|}{\left(l(d+\Delta\mathrm{d}) + \frac{\Delta\mathrm{d.\,k}}{\mathrm{d}}\left(\sqrt{|\omega|}-1\right)\right)^2} - T = 0 \tag{6.10}$$

By rewriting (6.8) and (6.9) based on λ and substituting k from (6.10) after an extensive calculation, these equations can be analytically solved. Due to the limitations of the space, details of analytical derivations can be found in the appendix IV. The analytical derivations show that by introducing an auxiliary variable $X = \sqrt{|\omega|}$, $|\omega|$ can be computed from (6.11) which is a quintic equation (polynomial equation of order 5). Analytical methods for solving a quintic equation by using radicals are available under specific criteria, such as Arthur Cayley method [105]. Alternatively well-known Newton-Raphson numerical method [106] can be used. Note that only real roots $|\omega| > 1$ are acceptable.

$$\alpha_5 X^5 + \alpha_4 X^4 + \alpha_3 X^3 + \alpha_2 X^2 + \alpha_1 X + \alpha_0 = 0 \tag{6.11}$$

where,

$$\begin{cases} \alpha_5 = (-C_\omega d^3 \Delta d) \\ \alpha_4 = \left(3C_\omega l d^3 \Delta d\sqrt{T}(d+\Delta\mathrm{d}) + C_\omega d^3 \Delta d\right) \\ \alpha_3 = \left(-3C_\omega l d^3 \Delta d\sqrt{T}(d+\Delta d) - 3C_\omega l^2 d^3 \Delta d T(d+\Delta d)^2\right) \\ \alpha_2 = \begin{pmatrix} 3C_\omega l^2 d^3 \Delta d T(d+\Delta d)^2 - C_k d\Delta d^3 T + \\ C_\omega l^3 d^3 \Delta d T\sqrt{T}(d+\Delta d)^3 + C_k l d\Delta d^3 T\sqrt{T}(d+\Delta d) \end{pmatrix} \\ \alpha_1 = \begin{pmatrix} -C_\omega l^3 d^3 \Delta d T\sqrt{T}(d+\Delta d)^3 - \\ 2C_k l d\Delta d^3 T\sqrt{T}(d+\Delta d) + 2C_k d\Delta d^3 T \end{pmatrix} \\ \alpha_0 = \left(-C_k d\Delta d^3 T + C_k l d\Delta d^3 T\sqrt{T}(d+\Delta d)\right) \end{cases}$$

After calculating $|\omega|$, k can be calculated from $|\omega|$ as:

$$k = \frac{\sqrt{|\omega|}d - ld\sqrt{T}(d + \Delta d)}{\Delta d\sqrt{T}\left(\sqrt{|\omega|} - 1\right)} \tag{6.12}$$

where $|\omega| > \left(l\sqrt{T}(d + \Delta d)\right)^2$

Note that any calculated $|\omega|$ from (6.11) should satisfy the condition of (6.12). If none of the roots of (6.11) can satisfy this condition, a minimum $|\omega|$ from this condition should be calculated as $|\omega|$ instead of using (6.11).

6.4.1 Discussion

To demonstrate how the proposed method can be used, let us start with two examples. Assume a given FVV system with $l = 0.01$, $\bar{d} = 100$, $\frac{\Delta d}{d} = 10\%$, $C_k = 10$, $C_\omega = 1$, and $T = 4$. Real roots of (6.11) are $X = \{1, 2.878, 0.0458\}$. As $|\omega| > 1$, first and third roots are not acceptable and by rounding the $|\omega|$, we have $|\omega| = 8$. k can be calculated from (6.12) for $|\omega| = 8$, which results in $k = 1.7185$ and the minimum cost function $C(k, |\omega|) = 11.39$. For the same system but by assuming $C_k = C_\omega = 1$, we have $|\omega| = 6$, $k = 0.8606$, and the minimum cost function $C(k, |\omega|) = 7.35$. The second example results in a higher number of cameras but less interpolation complexity as expected. Both examples result in $\overline{\mathrm{ESD}} = 4.00$ as desired.

Figure 6.5 shows the cost function C for different values of $|\omega|$ for these two examples. Figure 6.6 demonstrates the calculated k from (6.12) for each $|\omega|$ in Figure 6.5. Note that $|\omega| > 5$ to satisfy the condition of (6.12).

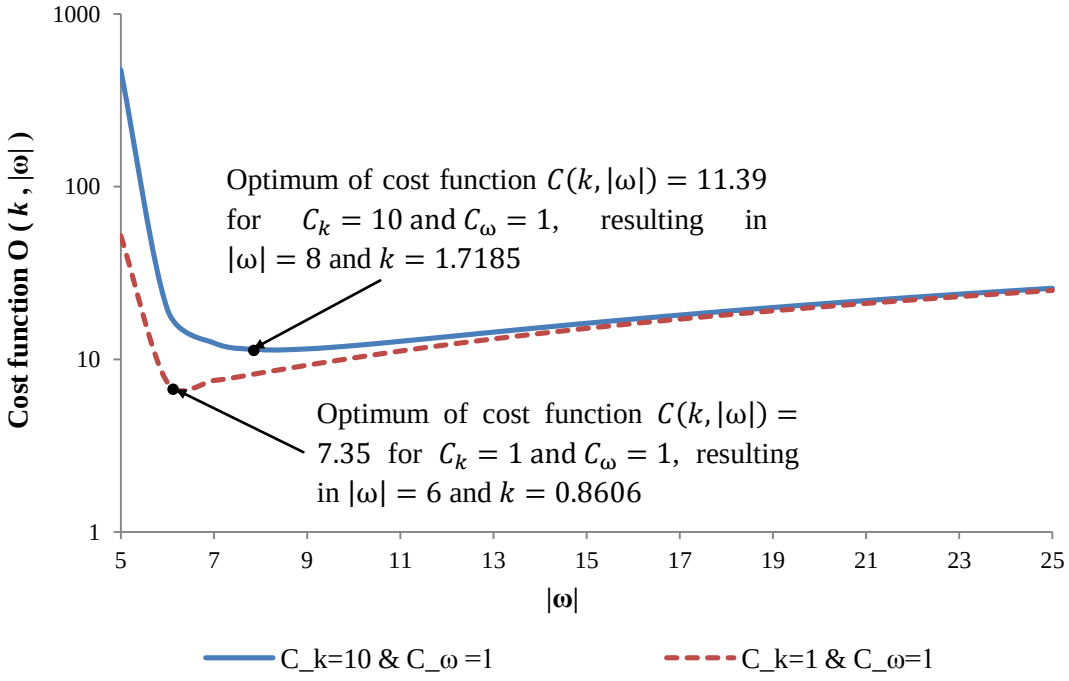


Figure 6.5. Cost function $C(k, |\omega|)$ vs. $|\omega|$ for two test cases

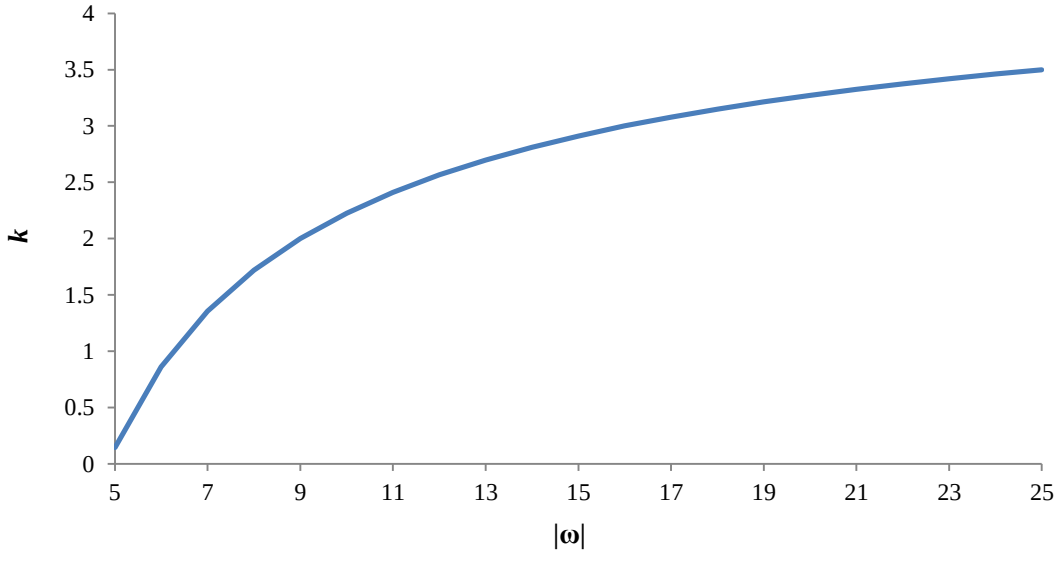


Figure 6.6. Calculated k from (6.12) for each $|\omega|$ shown in Figure 6.5.

Figure 6.7.a demonstrates the joint optimization of both $|\omega|$ and k for different values of $\frac{C_k}{C_\omega}$ from 0 to 500 for a system with $l = 0.01, \bar{d} = 100, \frac{\Delta d}{d} = 20\%,$ and $T = 4$. $\frac{C_k}{C_\omega} = 0$ occurs when $C_k = 0$, i.e., no cost is assumed for adding cameras. This results in a very high number of cameras and small $|\omega|$ as expected. High values of $\frac{C_k}{C_\omega}$ are associated with high cost of adding cameras and hence results in high $|\omega|$ but small number of cameras. Note that in Figure 6.7 the left vertical axis is based on $\frac{1}{k^2}$ and not k to express the camera grid density. The steps in the curves are due to the fact that $|\omega|$ must be an integer. Figure 6.7.b shows a region of 6.7.a in which $\frac{C_k}{C_\omega} \leq 10$.

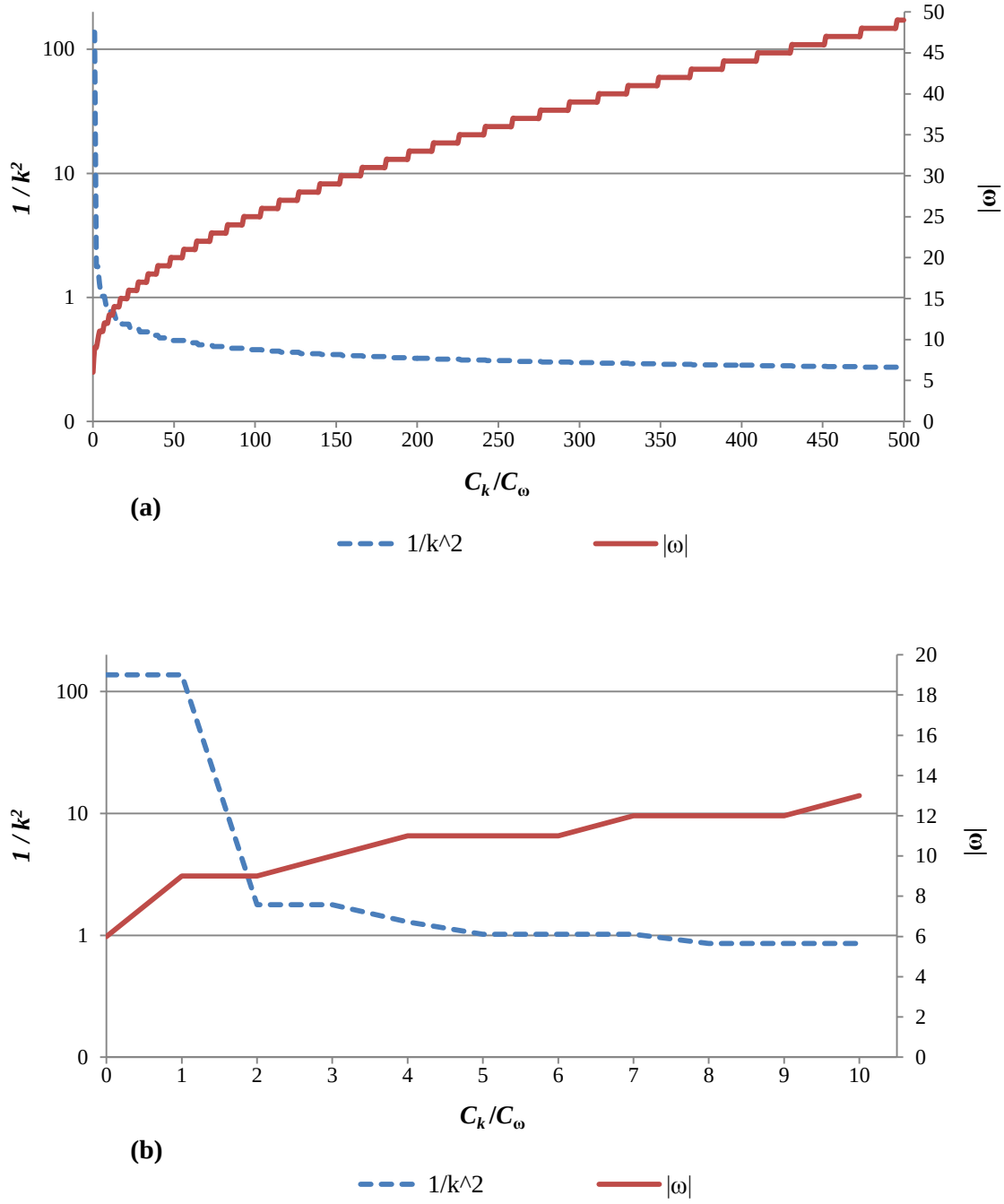


Figure 6.7. Joint optimization of $|\omega|$ and k for different values of $\frac{C_k}{C_\omega}$

a) $\frac{C_k}{C_\omega} \leq 500$ and b) $\frac{C_k}{C_\omega} \leq 10$

The main application of the proposed joint optimization method is to calculate the optimum trade-off among camera density and rendering complexity for a desired output video quality. Key advantages of this trade-off is in minimizing the cost of acquisition (by

significantly reducing the number of the cameras in the gird) and compensate it by increasing the rendering complexity up to a permitted level based on constraints on computational load (e.g., the acceptable delay introduced by rendering). Figure 6.8 demonstrates an LF-based FVV system with $l = 0.01$, $\bar{d} = 100$, and the camera grid size of $50x50$ units. For each depth estimation error $\frac{\Delta d}{d}$ in the range of $[1\%, 20\%]$, three scenarios of the proposed optimization method are shown for $|\omega| = 8, |\omega| = 16, \text{and } |\omega| = 32$. The vertical axis illustrates the calculated number of cameras for each scenario.

As can be seen from Figure 6.8, employing higher density of rays for interpolation can significantly decrease the number of cameras. For example, for 15% error in depth estimation, employing $|\omega| = 32$ rays for interpolation requires 484 cameras, compared with 729 cameras for 16 and 2809 cameras for 32 ray interpolation.

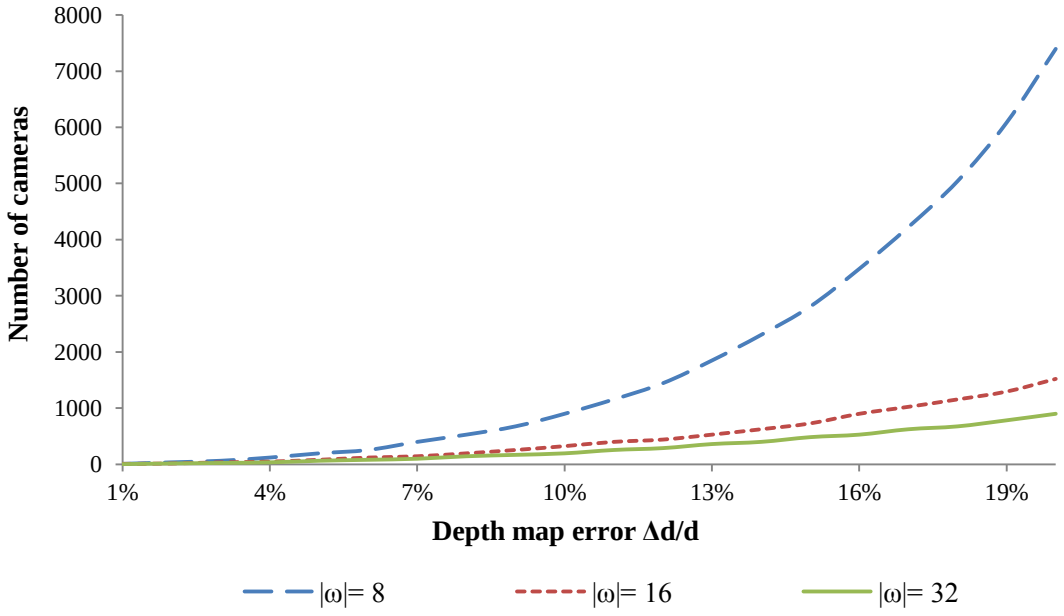


Figure 6.8. Calculated number of cameras in a camera grid for three different optimizations with $|\omega| = 8, |\omega| = 16, \text{and } |\omega| = 32$

6.4.2 Experimental Validation

For experimental validation, the same simulation system and empirical model introduced in chapter 2-5 is employed. Figure 6.9 illustrates the UV-DM rendering quality for fixed $k = 5$ and different $|\omega| = 4, 9, 16, 25, \text{and } 36$, for depth map error levels $\frac{\Delta d}{d}$ in the range of 5% to 20%. Rendering quality is reported in terms of PSNR. As can be seen in Figure 6.9, the same pattern expected from the proposed model is achieved, i.e., increasing the number of rays in interpolation improves the PSNR, e.g. $|\omega| = 25$ and 15% error performs better than $|\omega| = 9$ and 10% error.

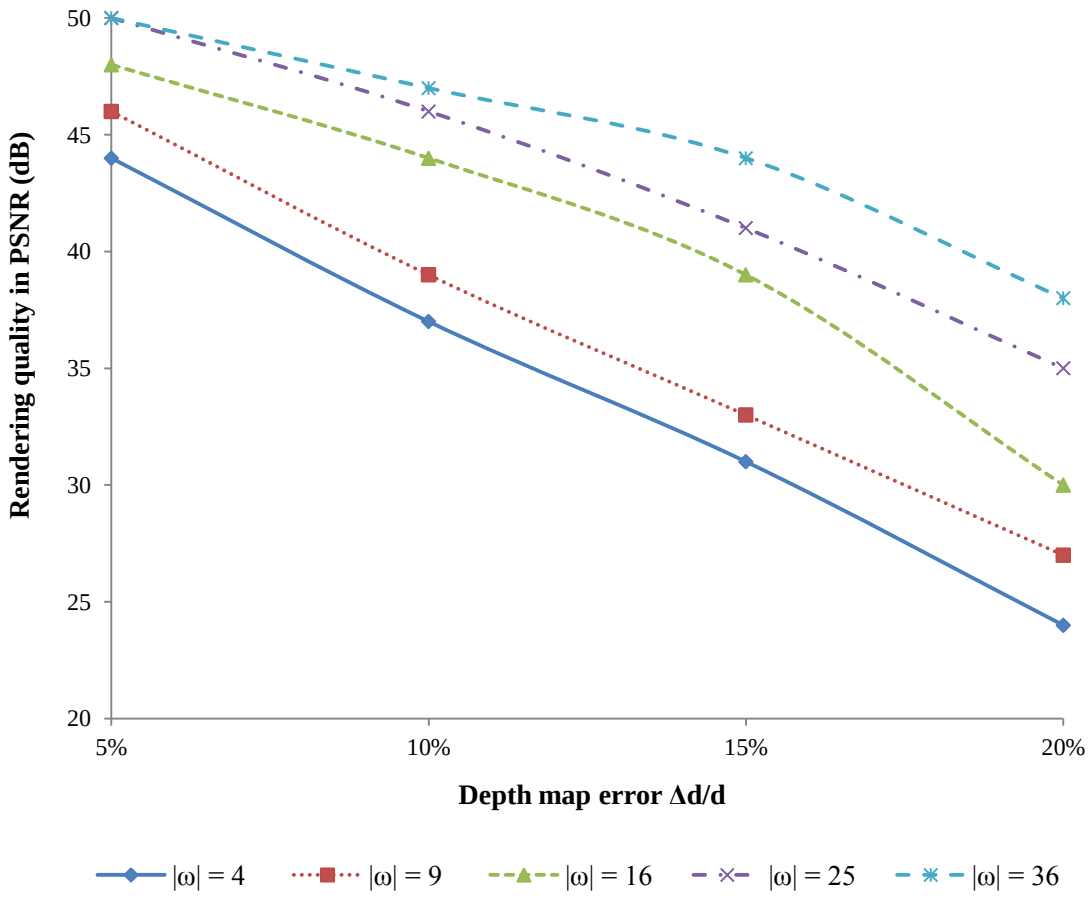


Figure 6.9. Experimental UVDM $(d, \Delta d, k, l, |\omega|)$ rendering quality in PSNR for $\frac{\Delta d}{d}$ in the range of $[5, 20]$, and $|\omega| = (4, 9, 16, 25, \text{and } 36)$

Figure 6.10 illustrates the UV-DM rendering quality for fixed $|\omega| = 4$ and different k = 0.5, 1, 2, 5, and 10 (and corresponding number of cameras), for depth map error levels $\frac{\Delta d}{d}$ in the range of 5% to 20%. As it can be seen in Figure 6.10, the same pattern expected from the proposed model is achieved, as smaller k improves the PSNR, e.g. $k = 1$ and 15% error in depth performs better than 10% error in depth with $k = 2$. Note that the data demonstrated in Figure 6.9 and Figure 6.10 are produced by 400,000 experiments, conducted by the simulator.

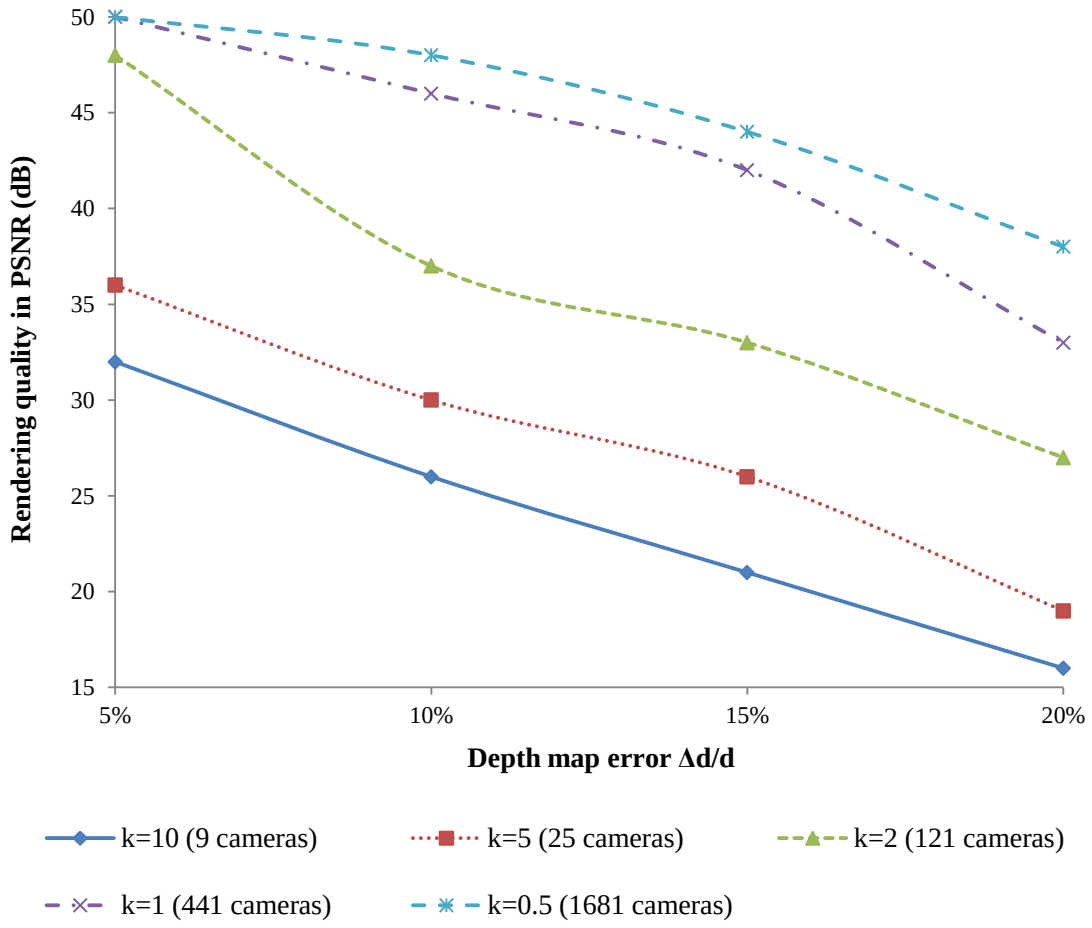


Figure 6.10. Experimental UVDM$(d, \Delta d, k, l, |\omega|)$ rendering quality in PSNR for $\frac{\Delta d}{d}$ in the range of $[5, 20]$, and $k = (0.5, 1, 2, 5, \text{and } 10)$

6.5 Conclusion

A Lagrangean joint optimization method of the (i) density of rays used for interpolation and (ii) the camera density for a light field based free viewpoint video system is presented in this chapter. It is shown that the error in depth maps can be compensated for by judicious alteration of *both* variables, to maintain the rendering quality at a desired level. Theoretical analysis based on effective sampling density (ESD) is validated through extensive experiments.

Chapter 7: Non-Uniform/Irregular Acquisition based on the Scene Complexity Variations

7.1 Summary

Acquisition of a free viewpoint video (FVV) system is theoretically considered as a problem of plenoptic signal sampling. It is typically performed by using a regular camera grid. While a regular acquisition itself results in non-uniform sampling density, this non-uniformity does not match the scene complexity and frequency variations. This Chapter shows how to superimpose the ESD theory with the scene complexity and proposes an irregular acquisition method for optimum non-uniform LF acquisition corresponding to the variations of the scene complexity. Specifically, scene complexity is measured through analyzing DCT coefficients of reference images of the scene, describing the frequency behavior of the plenoptic signal over the scene space. An optimization model is formulated to calculate the optimum configurations of the acquisition cameras including positions and orientations. The theoretical analysis and numerical simulations demonstrate that the rendered video quality can be significantly improved (around 20% in mean PSNR) by employing the proposed irregular acquisition compared with the regular camera grid.

The rest of this chapter is organized as follows. Section 2 presents an overview of how the theory of ESD can be superimposed by scene complexity. Section 3 describes the proposed non-uniform/irregular acquisition optimization model. Section 4 demonstrates the experimental validation for the proposed optimization model. Section 5 concludes the chapter.

7.2 Effective Sampling Density and Scene Complexity

Denote from chapter 3 that LF sampling density (SD) is defined as the number of acquired rays per unit area of the scene space and Effective Sampling Density (ESD) as the number of rays per unit area of the scene that have been acquired and employed in the rendering process to estimate an unknown ray [16, 30]. Clearly SD is an acquisition parameter representing the upper bound of ESD which is also dependent on the rendering method.

As shown in chapter 3, the output quality of an LF-based FVV system is determined by three factors:

1. The ESD in interpolation area A, the vicinity of the point p (the intersection of unknown ray r with scene), which is mathematically derived in chapter 3 for a given acquisition configuration and rendering method.

2. The scene complexity in area A, which could be measured in terms of its spatial frequency components.

3. The accuracy and effectiveness of the interpolation function F employed for the estimation of the unknown ray r.

Note that this is a micro analysis of quality assessment for reconstruction of only one unknown ray r. For a macro analysis of the output quality of LF-based FVV systems; typically, $\overline{\text{ESD}}$, the average of the ESD for a given acquisition and rendering components is calculated for the whole scene and the average of output quality is calculated empirically from this $\overline{\text{ESD}}$. The evaluation, comparison and optimization of the acquisition components and rendering algorithms which have been carried out in chapters 3 to 6 are based on this assumption.

In addition, note that chapters 3 to 6 assumed a fixed scene complexity and a given interpolation algorithm for the analysis, evaluation and optimization. These assumptions were necessary to make ESD the exclusive objective indicator of the output quality.

In contrast, this chapter shows that scene complexity can be included in the proposed ESD theory. This will result in an optimization model, formulated for optimum non-uniform/irregular LF acquisition. Before demonstrating the optimization model, let's reproduce the relation of output quality and scene complexity from the chapter 3 experiments. Figure 7.1 shows the mean PSNR for different rendering methods, categorized based on the complexity of the scene. As can be seen, more complex scenes result in reduced output quality. This can be explained due to fixed ESD for different scenes with different complexities in term of higher spatial frequency components. Nevertheless, ESD provides the right prediction on the relative performance amongst the various methods.

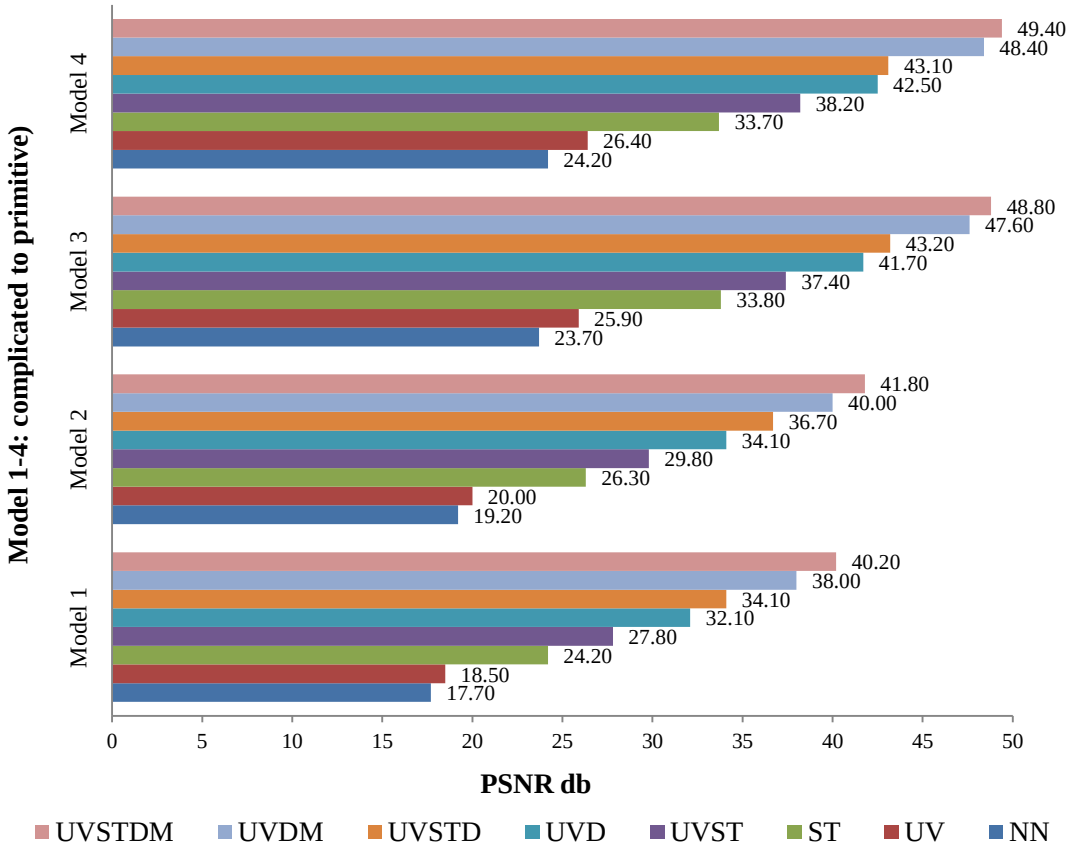


Figure 7.1. Output quality and scene complexity

In general, experiments demonstrate that the relation between ESD, scene complexity, and output quality can be illustrated as Figure 7.2. Output quality in Figure 7.2 is demonstrated

as normalized *LF signal reconstruction accuracy* which is a parameter between 0 and 1, with 1 representing perfect LF signal reconstruction. In practise, value 0 for *reconstruction accuracy* is corresponding to very low output quality in PSNR (e.g., less than 10) and value 1 to very high quality in PSNR (e.g., higher than 50). In addition, different levels of scene complexities are represented by parameter h in Figure 7.2. As will be discussed later, in this thesis h is calculated from spatial frequency variations of the scene, particularly from the values of the high frequency DCT coefficients. As can be seen from Figure 7.2, the *LF signal reconstruction accuracy* vs. ESD can be regarded as a set of utility functions $U_h(ESD)$ based on the given h. The higher the scene complexity, more ESD would be required for a given reconstruction fidelity.

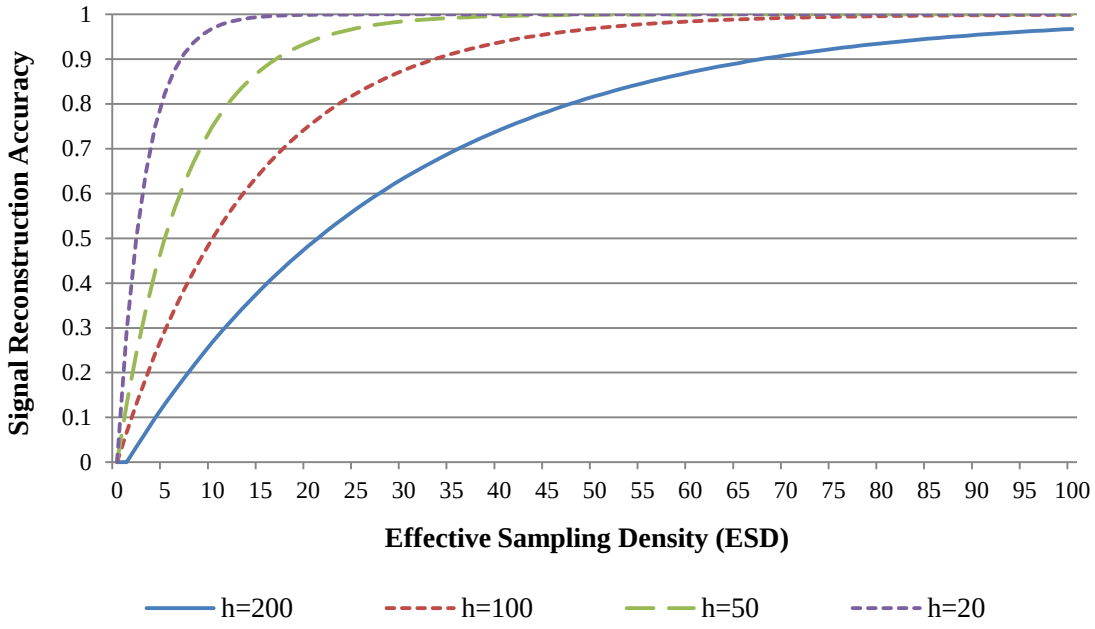


Figure 7.2. LF signal reconstruction accuracy vs. amounts of under-sampling, in term of ESD as a set of utility functions U_h(ESD) for several scenes with different complexities h

Each acquisition configuration and rendering method results in an ESD pattern, which varies in the scene space. Assume that the scene could be partitioned into a number of smaller

3D regions or blocks, each having a fixed average complexity h. Then, the aim of the optimization problem could be to find the optimum acquisition configuration which provides the minimum required ESD for all blocks. Section 4 proposes a mathematical formulation for this optimization problem. Let's first review several assumptions used in the optimization model.

7.2.1 Simplifications Applied to the Proposed Optimization Model

While the proposed optimization model proposed in next section is generic, several simplifications will be applied for the initial solution as well as validation. First, the rendering method is assumed to select all available rays in each point of the scene, i.e., $|\omega| = |\Omega|$ for the reconstruction of each unknown ray r. This simplification assumes unlimited computational resources for *rendering process* and compromises the influences of rendering component in output quality, hence restricts the optimization to acquisition component. However, the future work on the optimization model requires a joint optimization on both acquisition and rendering components similar to what has been proposed in chapter 6 but this time with the addition of scene complexity. This assumption results in ESD to be equal to SD in any point of the scene space and thus ESD can be substituted with SD in the optimization model.

Another simplification is to limit the camera positions to a plane, i.e., irregular camera grid. In other word, while the optimization model is proposed in 3D space, the simplified solution is given for 2D space.

Final simplification is for representing the scene complexity. While the depth maps are available and the scene can be discretized into small 3D blocks and for each 3D block, the frequency variations of depth and texture computed to estimate average complexity h, but we approximate the scene complexity with 2D *scene complexity maps* produced from 2D reference images. This approximation method will be discussed later in subsection 7.4.2.

7.2.2 SD Pattern in Scene Space

Please recall from chapter 3 that even a regular camera grid results in a non-uniform SD in the scene space. However, this non-uniformity is not related to the spatial variations of scene complexity.

As discussed in subsection 7.3.1, the optimization model will be simplified to SD instead of ESD. Figure 7.3 shows the non-uniformity of SD as contour maps for an initial regular camera grid of $30x30$ with $k = 2$ at different depths, $d = 30$, 60, and 90, camera field of view of 30°, image resolution of $100x100$ pixels ($l = 0.005358984$) and ideal area $A = (ld)^2$, i.e., LF system resolution.

While Figure 7.3 illustrates the SD pattern for a regular acquisition, it is possible to conclude that by changing the acquisition pattern, the SD pattern would change and therefore, the proposed optimization model is required to map this SD pattern to the scene complexity, i.e., provide the minimum required SD for all 3D discretized blocks of the scene.

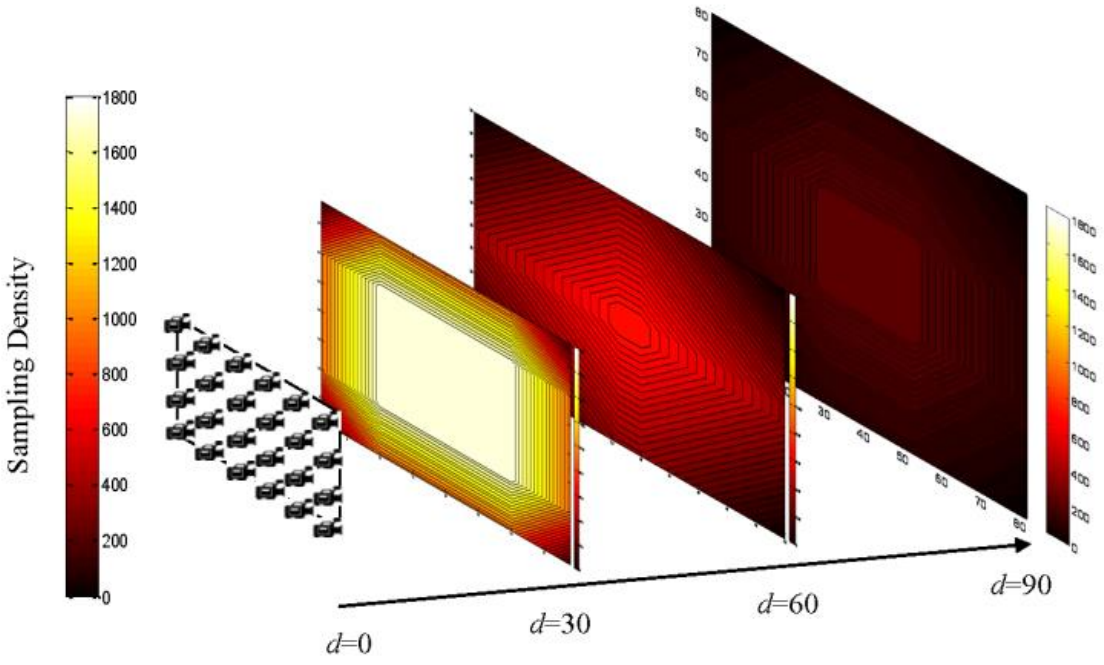


Figure 7.3. SD pattern in scene space as contour maps at different depths

7.3 Proposed Acquisition Optimization Model

This section proposes the optimization model for non-uniform/irregular LF acquisition. Initially the generic optimization will be formulated. Subsequently, the simplifications from subsection 7.3.1 are applied one by one to the optimization model to derive an initial solution for validation. In general, the cameras should be arranged such that the proportion of the scene with higher complexity is sampled by more cameras.

Assume the scene is discretized with m small 3D blocks, $\{p_1, p_2, \dots p_m\}$. Each block p_i has a fixed volume with equal length ε_i in X, Y and Z directions, 3D position of the center of block in the scene space (x_i, y_i, z_i), and a weight h_i indicating the complexity of the scene that may be measured based on the localized highest frequency components of p_i . Each block p_i is represented as $p_i(x_i, y_i, z_i, \varepsilon_i, h_i)$. In addition, assume there are n cameras with known parameters in the acquisition component, $G_n = \{c_1, c_2, \dots c_n\}$. Each camera c_i is described as $c_i(x_i, y_i, z_i, \alpha_i, \beta_i, \theta_i, f_i)$, where x, y, and z are the camera's position and α, β and θ are the camera's orientation in X, Y and Z directions respectively and f refers to the camera's internal parameters, e.g., horizontal and vertical fields of view. Acquisition $G_n = \{c_1, c_2, \dots c_n\}$ and rendering algorithm R can be mathematically employed to compute the ESD pattern in scene space. $ESD_R(G_j, p_i)$ gives the ESD at point p_i for a given acquisition G_j of j cameras and rendering method/algorithm R. Let $U_{h_i}(ESD_R(G_j, p_i))$ denote the rendering accuracy for each block p_i as demonstrated in Figure 7.2. An optimal configuration of the cameras is expected to provide the minimum ESD required for each block of complexity h_i that results in an acceptable output quality as specified by the utility function for all blocks of the scene. To improve the chances of obtaining a feasible solution, we define $W(h_i)$ as the weight of h_i based on U_{h_i} shape. W in the simplest form can be $W(h_i) = h_i$ which means there is a linear relation between required ESD and h_i. For a given number of cameras, an objective function $O(G_n)$ can be formulated as:

$$O(G_n) = \sum_{i=1}^{m} \left(U_{h_i}\left(ESD_R(G_n, p_i)\right).W(h_i) \right) \tag{7.1}$$

The optimization problem is to find an acquisition G_n which makes the $O(G_n)$ maximum. Note that G_n is a vector with $6n$ elements (n cameras and 6 degrees of freedom for each camera). For a simplified acquisition component, i.e., a 2D camera grid, z_i and θ_i are fixed and the vector G_n has $4n$ elements. Due to acquisition limitations the constraints over G_n may be expressed as:

$$\begin{cases} x_{i_{min}} \le x_i \le x_{i_{max}} \\ y_{i_{min}} \le y_i \le y_{i_{max}} \\ z_{i_{min}} \le z_i \le z_{i_{max}} \\ \alpha_{i_{min}} \le \alpha_i \le \alpha_{i_{max}} \\ \beta_{i_{min}} \le \beta_i \le \beta_{i_{max}} \\ \theta_{i_{min}} \le \theta_i \le \theta_{i_{max}} \end{cases} \tag{7.2}$$

The core of the optimization problem defined in (7.1) is the U_{h_i} utility functions definition. As shown in Figure 7.2, U_{h_i} has a diminishing marginal utility. U_{h_i} can be iteratively defined by its marginal utility $\delta_{U_{h_i}}$ as:

$$U_{h_i}(i) = U_{h_i}(i-1) + \delta_{U_{h_i}}(i-1) \; and \; U_{h_i}(0) = 0 \tag{7.3}$$

Note that for a continuous approximation of U_{h_i} , $\delta_{U_{h_i}} \approx \frac{dU_{h_i}}{dESD}$.

To iteratively define utility functions U_{h_i} based on (7.3), $\delta_{U_{h_i}}$ can be approximated by a linear or logarithmic diminishing behavior as:

$$\delta_{U_{h_i}}(j) = U_{h_i}(1) - \left(\frac{(j-\beta_1)}{\beta_2}\right) U_{h_i}(1) \text{ and } \beta_1 \ge 1, \beta_2 \ge U_{h_i}(1) \tag{7.4}$$

Note that by assuming $\beta_2 = U_{h_i}(1)$, (7.4) is simplified to a linear marginal utility. From (7.3) and (7.4), it can be shown that:

$$U_{h_i}\left(ESD_R(G_j, p_i)\right) = \sum_{j=1}^{ESD_R(G_j,p_i)} \delta_{U_{h_i}}(j) \tag{7.5}$$

To simplify (7.5) let's assume all blocks have equal size $\varepsilon_i = 1$. $E(G_j, p_i)$ gives the number of cameras that are able to see/capture p_i for a given acquisition $G_j = \{c_1, c_2, \dots c_j\}$. In addition let's assume the rendering algorithm selects all available rays in each point of the scene for interpolation, then:

$$ESD_R(G_j, p_i) = SD(G_j, p_i) = \frac{E(G_j, p_i)}{{\varepsilon_i}^2} = E(G_j, p_i) \tag{7.6}$$

By applying (7.4) and (7.6) to (7.5) and assuming $W(h_i) = \frac{h_i}{U_{h_i}(1)}$ for all blocks, (7.1) can be rewritten as:

$$O(G_n) = \sum_{i=1}^{m} \sum_{j=1}^{E(G_n, p_i)} \left(h_i - \left(\frac{(j - \beta_1)}{\beta_2} \right) h_i \right) \tag{7.7}$$

To compute (7.7), it is easier to rewrite it based on cameras in acquisition component G rather than scene blocks. $Q(c_i)$ gives a set of scene blocks that can be seen/captured by camera c_i :

$$O(G_n) = \sum_{i=1}^{n} \sum_{p_j \in Q(c_i)} \left(\delta_{U_{h_j}} \left(E(G_i, p_j) \right) \frac{h_j}{U_{h_j}(1)} \right) \rightarrow$$

$$O(G_n) = \sum_{i=1}^{n} \sum_{p_j \in Q(c_i)} \left(h_j - \left(\frac{\left(E(G_i, p_j) - \beta_1 \right)}{\beta_2} \right) h_j \right) \tag{7.8}$$

To assess the problem complexity, consider that an exhaustive search method to maximize (7.8) would require all possible configurations of G_n to be examined. To reduce the size of the search space a simplified 2D camera grid UV can be assumed. Let UV plane size be $u.v$, the discretization size, i.e., the minimum space required for a camera as du and dv *in* U an V directions, and O_u and O_v possible number of discrete orientations in U and V directions, then there are $N = \frac{u.v.O_u.O_v}{du.dv}$ possibilities for positioning the cameras. Let a new set $S = \{s_1, s_2, \dots s_N\}$

all potential cameras in UV plane. Clearly, $G_n \subset S$. The total number of possible G_n is $\frac{\left(\frac{u.v}{du.dv}\right)!O_u.O_v}{n!\left(\left(\frac{u.v}{du.dv}\right)-n\right)!}$ which represents the size of the search space.

In this thesis we adopt a dynamic programming solution to obtain G_n effectively for reasonable n and N. The boundary condition for calculating $G_1$is:

$G_1 = \{c_1\}, O(G_1)$is maximum when:

$c_1 = s_k$ and k is the index of the maximum in:

$$\max_{1 \le r \le N} \sum_{p_j \in Q(s_r)} (h_j) \tag{7.9}$$

The general rule for calculating G_{i+1} can be written as:

Assume $O(G_i)$is maximum,

$O(G_{i+1})$is maximum when:

$G_{i+1} = G_i \cup \{c_{i+1}\}$ and

$c_{i+1} = s_k$ and k is the index of the maximum in:

$$\max_{\substack{1 \le r \le N, \\ no\ camera\ in\ s_r}} \sum_{p_j \in Q(s_r)} \left(h_j - \left(\frac{\left(E\left(G_i \cup \{s_r\}, p_j\right) - \beta_1\right)}{\beta_2} \right) h_j \right) \tag{7.10}$$

From (7.9) and (7.10), it is straightforward to write an algorithm to obtain the optimum G_n. The dynamic programming solution also can store all h_j in a data structure and in each step of calculating the optimum G_n only update the affected subset of h_j with $\delta_{U_{h_i}}$.

7.3.1 Computing the h_i Complexity Factors

Different methods can be used to obtain h_i for each block based on the FVV design requirements. This chapter assumes that h_i is be determined by the highest frequency components of the block computed by applying DCT transform.

After applying DCT to each block p_i, different techniques can be utilized. One can calculate the norm or sum of energies for all AC coefficients and put it in h_i . Another technique is to only calculate the norm or sum of energies for high frequency components such as $4x4$ or $16x16$ bottom-right coefficients of the DCT. Experiments show that it is a good practice to apply a normalization process to the raw weights from the previous step.

7.3.2 Optimization with no Scene Geometric Information

In many LF based FVV systems, little or no information is available about the scene geometry. In such cases, it would not be possible to ascertain the requisite information about blocks p_i such as x_i, y_i, z_i, and our knowledge about the scene is limited to n reference images captured with current acquisition cameras.

This subsection introduces a modified 2D version of the optimization problem defined in (7.1) to (7.10) by replacing p with approximated 2D scene blocks $\tilde{p}_i(\tilde{x}_i, \tilde{y}_i, \tilde{h}_i)$, where $\tilde{x}_i, \tilde{y}_i$ refer to the position of the projected scene over image plane ST. The 2D projection transform which maps p to $\tilde{p}$ can be a typical 2D image projection transform or a more complex transform by averaging the disparity of each block of the scene among all reference images to minimize the errors of approximation of $(\tilde{x}_i, \tilde{y}_i)$ from (x_i, y_i).

Scene complexity map is defined as a 2D representation of all $\tilde{p}_i(\tilde{x}_i, \tilde{y}_i, \tilde{h}_i)$. Assume that the cameras in 2D camera grid (G_n) are in an initial state and it is desired to change the

cameras positions and orientations based on the current scene. To generate the *complexity map*, images of the scene are first captured using initial camera configuration, subsequently this *complexity map* is employed in the modified 2D optimization to obtain the optimal configuration of acquisition cameras. In the following, two methods are described for the generation of the *complexity map*.

Method I: Apply an image mosaicing algorithm to all reference images to generate a panoramic view of the whole scene. Alternatively, a virtual camera can be located in the middle of UV camera plane with a very large field of view which covers the whole scene and LF rendering itself can be used to calculate the view. Subsequently, this image is discretized into 2D scene blocks $\tilde{p}_i$ and DCT transform is applied to each 2D block $\tilde{p}_i$. Weights $\tilde{h}_i$ can be calculated from these 2D DCT coefficients with the same methods proposed in subsection 7.4.1.

Method II: In this method, the reference images captured by each camera are first discretized into blocks. The DCT transform then is applied to these image blocks and a localized *complexity map* is generated for each camera. Subsequently a "DCT mosaicing" algorithm is applied to all of these localized *complexity maps* to generate the final *scene complexity map*. The main difference between the "DCT mosaicing" and "image mosaicing" is that they treat the overlapping parts differently, where the proposed "DCT mosaicing" computes the average or maximum energy of DCT coefficients for compositing stage.

In practice, the scene complexity map could change in time and, hence, a dynamic acquisition optimization would be desired to control movable and steerable cameras. This extension is left for future publications.

7.4 Experimental Validation

The same FVV simulation system discussed in chapter 3 to 6 was utilized to validate the proposed optimization model. Four new 3D scenes were chosen for validation including:

bonny, dragon, a scene combined from bonny, dragon and two textured cubes and a room, in order of complexities, respectively. The 2D version of the optimization algorithm using reference images to quantify the scene complexity is reported in the following experiments.

First regular camera grids with $n = 49$ and 169 were generated for these scenes. For each experiment, 1000 random virtual cameras were produced. Rendering quality was calculated in terms of averaged PSNR among all 1000 virtual cameras for all four 3D scenes by employing UVST-DM rendering method. Consequently, the proposed optimization model was applied to each scene and a new irregular camera grid was obtained. The same virtual cameras were produced and the rendering quality for new irregular grid was compared with the initial regular grid.

Figure 7.4 demonstrates the summary of output quality for all four scenes for initial regular and optimized irregular $7x7$ and $13x13$ camera grids.

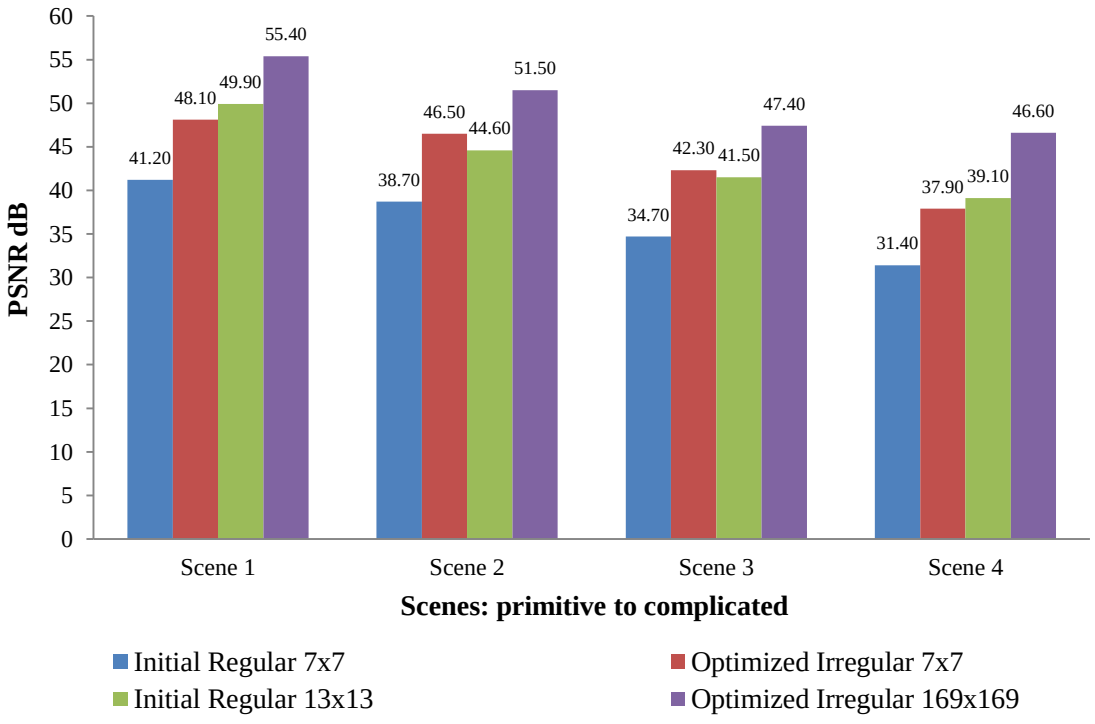
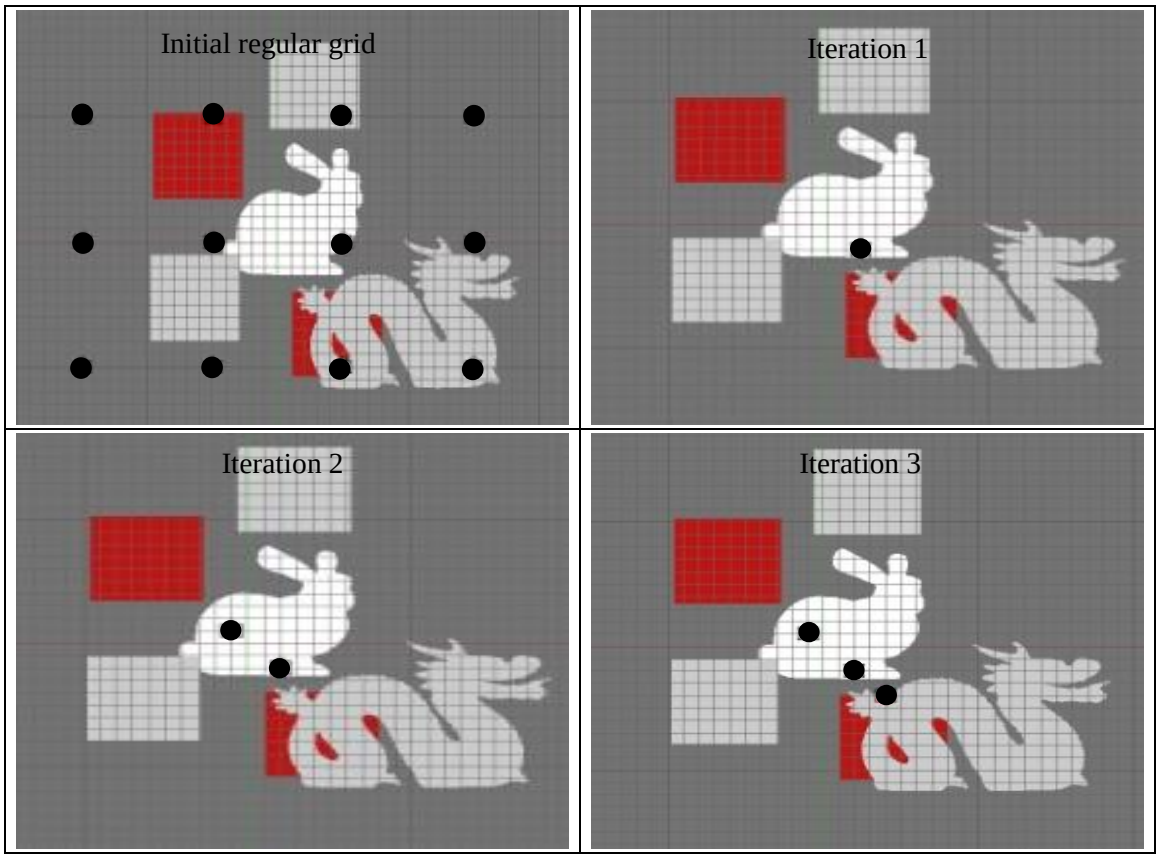


Figure 7.4. Output quality for four scenes for initial regular and optimized irregular camera grids

In general, Figure 7.4 shows about 20% improvement for $7x7$ and 15% improvement for $13x13$ camera grids in the PSNR values by using the proposed optimization model.

For the visualization of the output only two of the four scenes are selected in this chapter due to limited space, the bonny, dragon and two textured cubes scene for validation and subjective evaluation and the complex room scene for demonstration of the application of the proposed model.

Figure 7.5 illustrates the initial regular camera grid and steps/iterations of 2D optimization algorithm (7.9) and (7.10) for creating an optimum irregular camera grid for the first scene.



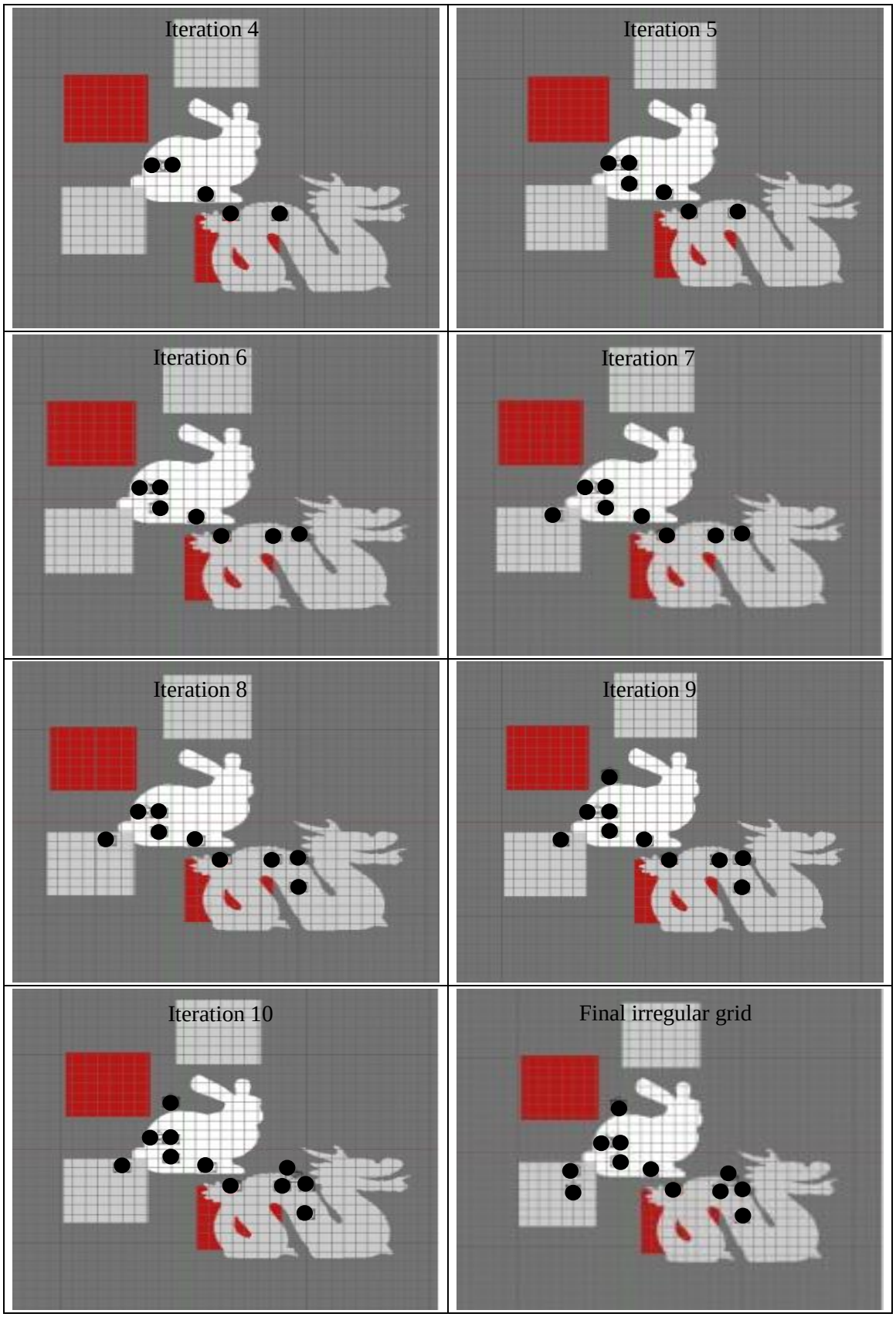
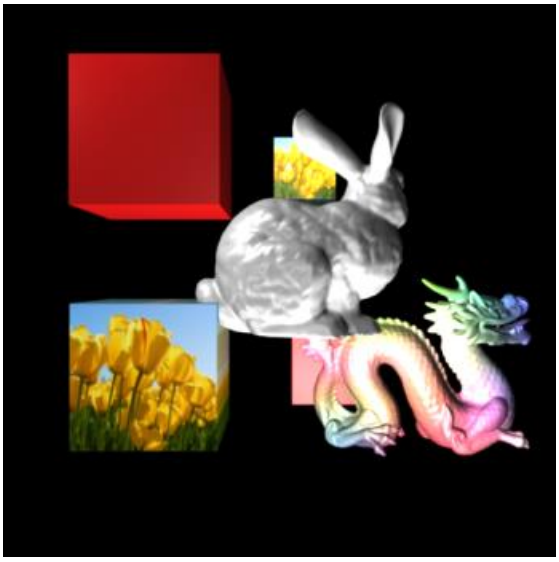


Figure 7.5. Initial regular camera grid and steps/iterations of the optimization algorithm generating an optimum irregular camera grid

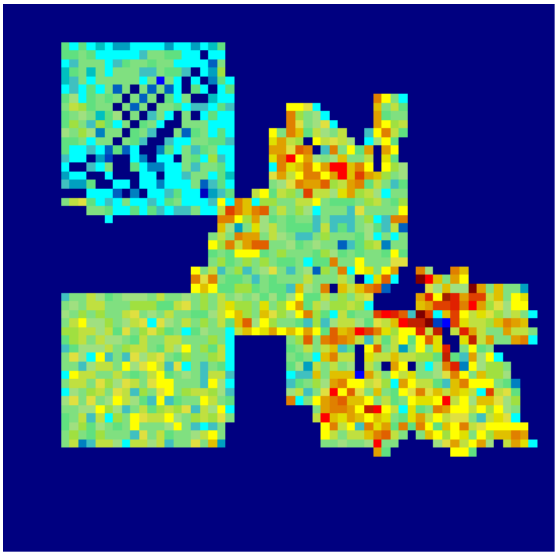
Figure 7.6.a demonstrates a complete view of first scene. Figures 7.6.b and 7.6.c show generated *complexity map* for this scene from *method I* and *method II* as described in section 4.

Similarly, Figure 7.7.a demonstrates a complete view of the second scene. Figures 7.7.b and 7.7.c show generated *complexity map* for this scene from *method I* and *method II*.

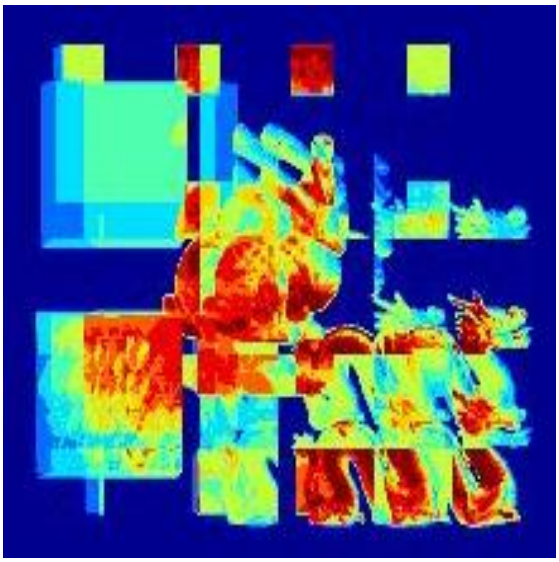
Figure 7.8.a illustrates a sample image out of 1000 ground truth images generated with simulator. Figures 7.8.b and 7.8.c show the rendered output for the same virtual camera by employing a regular and an optimum irregular acquisition respectively. While subjective comparison shows the rendering improvement for irregular grid, the objective comparison also shows average 15% to 20% improvement in average PSNR among all 1000 virtual cameras as demonstrated before in Figure 7.4.



(a)

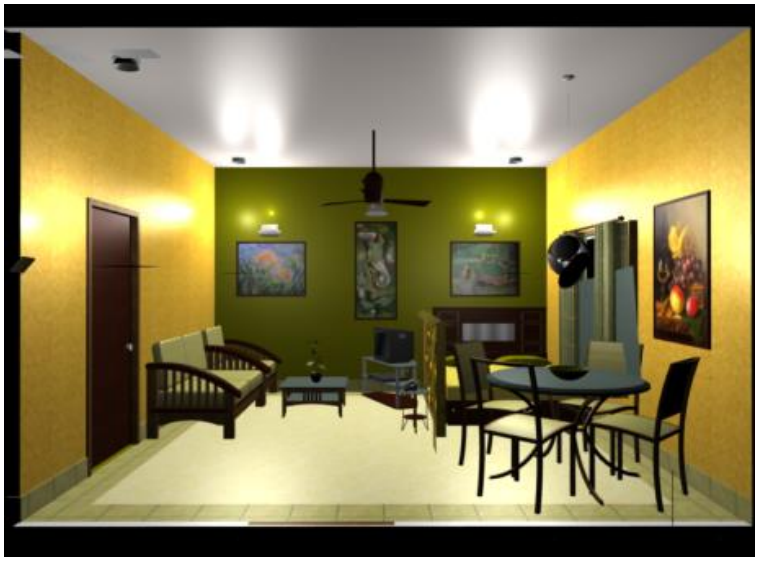


(b)

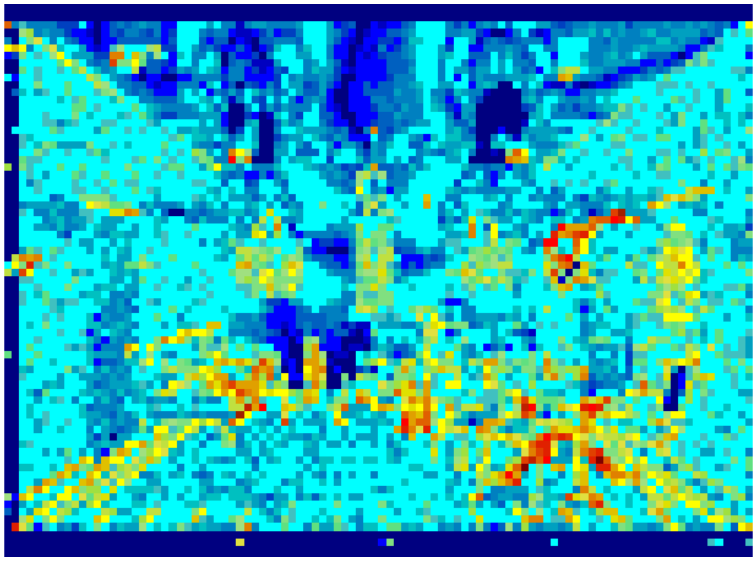


(c)

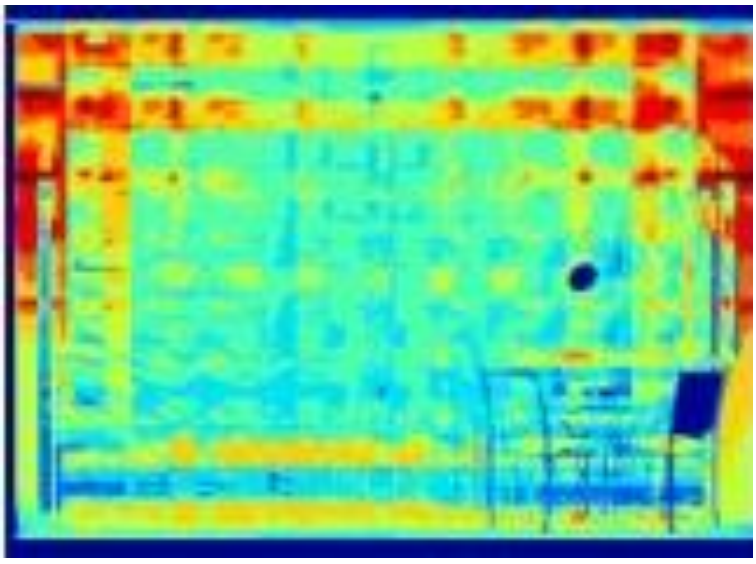
Figure 7.6. a) Scene I; Normalized complexity map from b) method I and c) method II



(a)

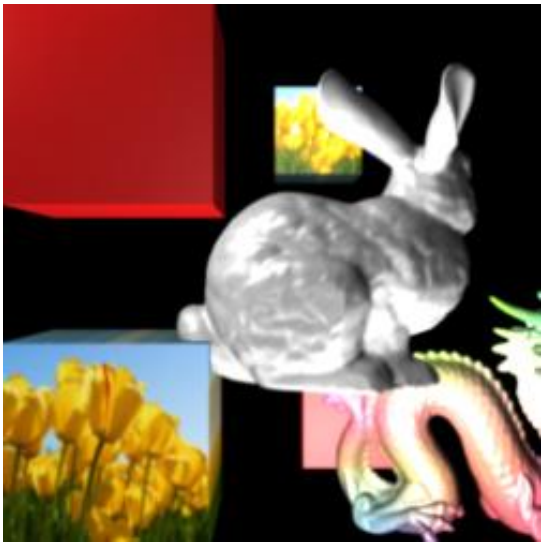


(b)

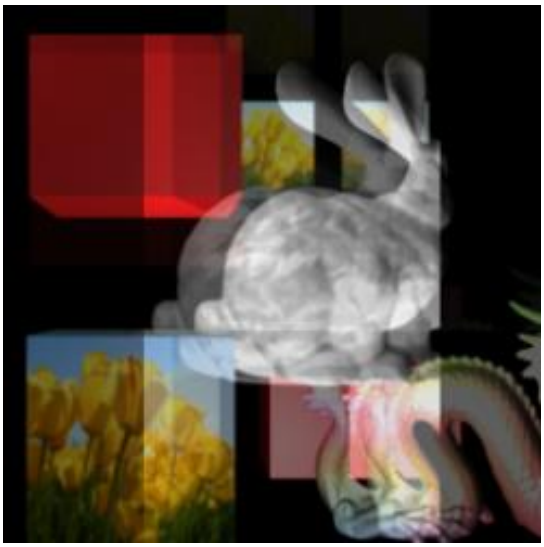


(c)

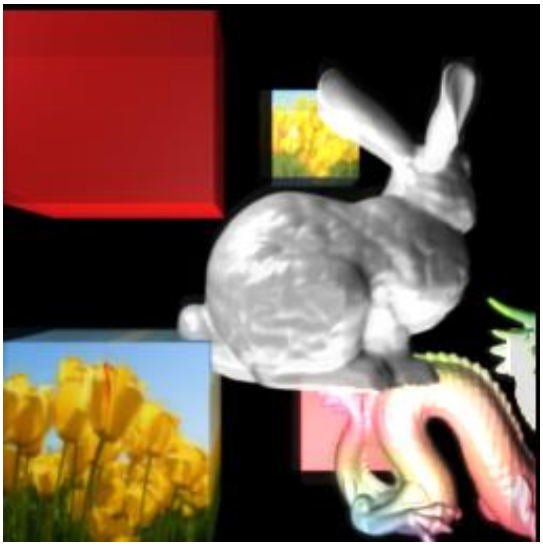
Figure 7.7. a) Scene II; Normalized complexity map from b) method I and c) method II



(a)



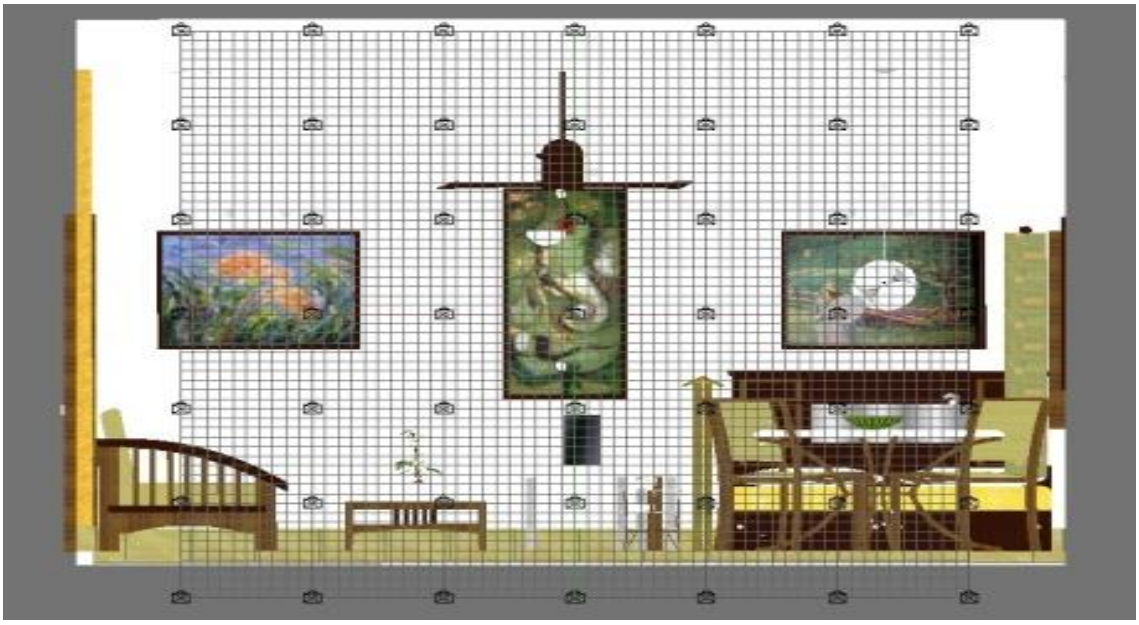
(b)



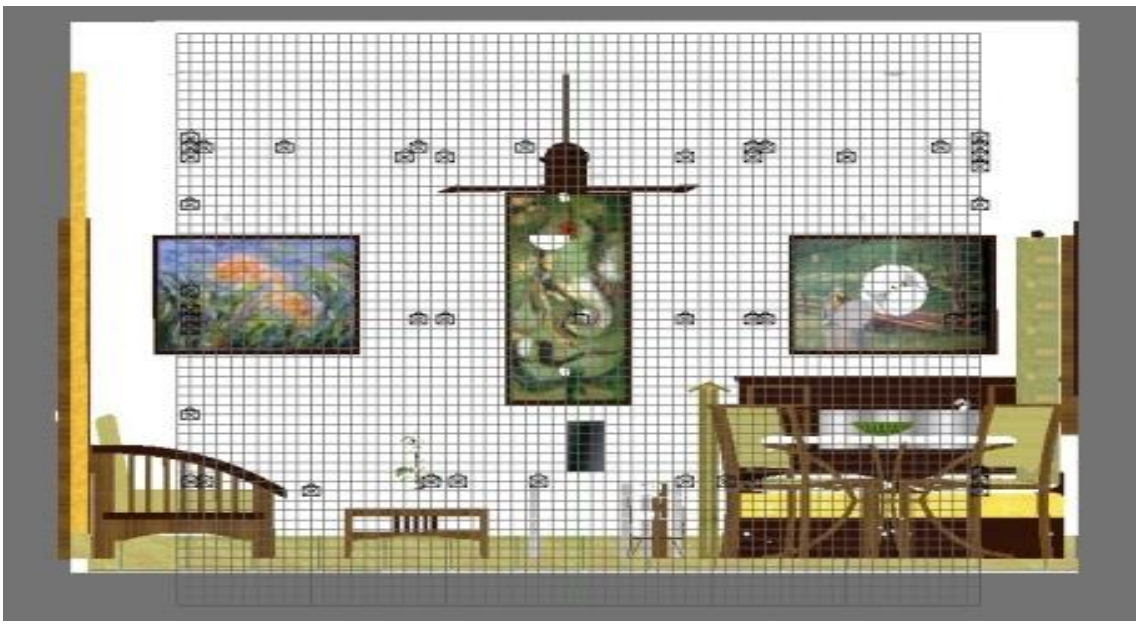
(c)

Figure 7.8. A sample of a) ground truth from simulator, b) rendered with regular grid, c) rendered with irregular grid, out of 1000 images generated, with average rendering improvment of 15% to 20% in PSNR

Figures 7.9.a and 7.10.a show regular camera grids with 49 and 169 cameras capturing the room scene respectively. Optimum irregular camera grids based on proposed optimization with 49 and 169 cameras are illustrated in 7.9.b and 7.10.b respectively. Note that the objective output quality improvements have been demonstrated in Figure 7.4 before.

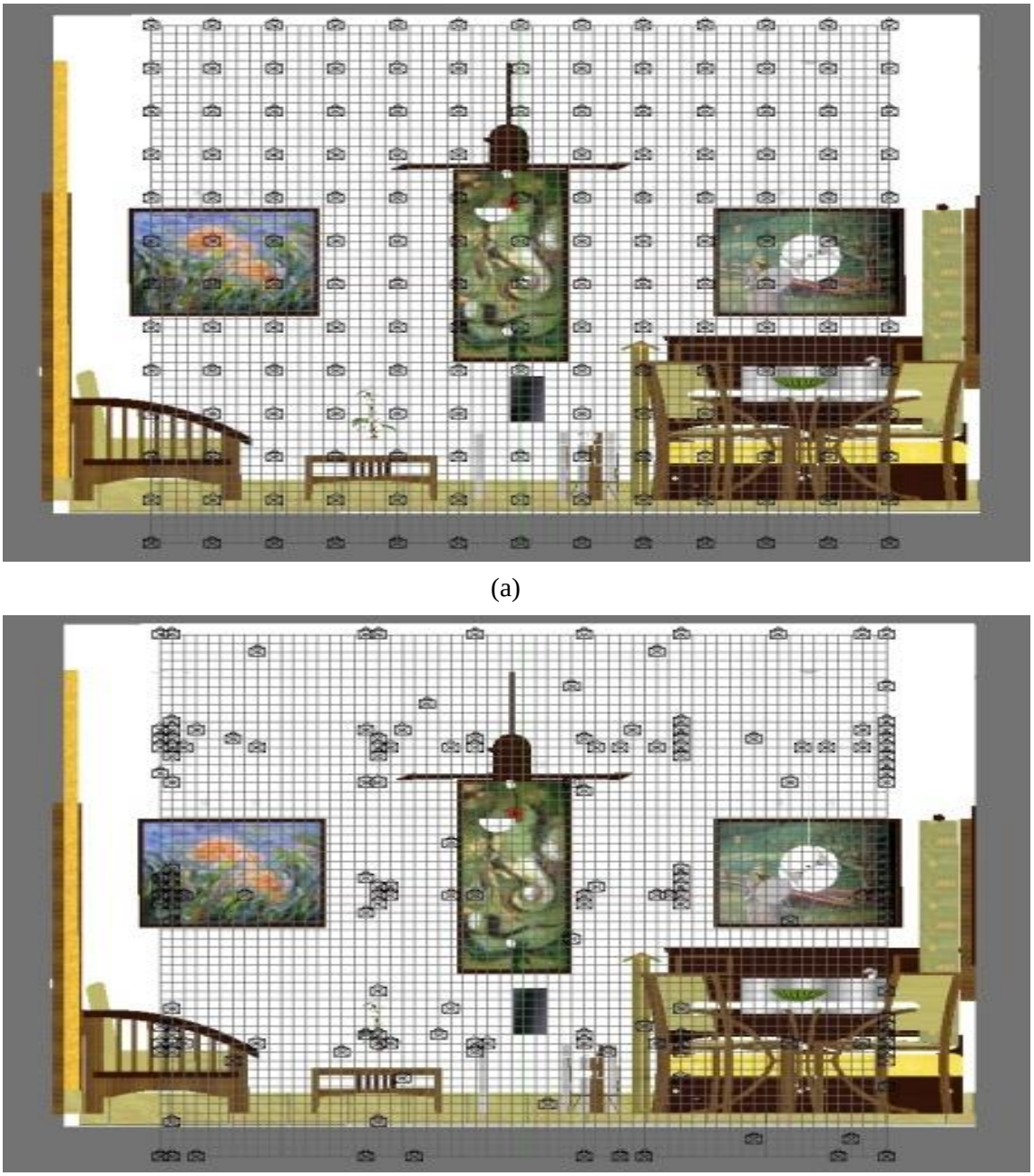


(a)



(b)

Figure 7.9. a) Regular camera grid with 49 ($7x7$) cameras; b) optimum irregular camera grid based on the proposed optimization model for 49 cameras



(a)

(b)

Figure 7.10. a) Regular camera grid with 169 (13*x*13) cameras; b) optimum irregular camera grid based on the proposed optimization model for 169 cameras

7.5 Discussion and Conclusion

By superimposing the scene complexity in term of scene special frequency variations to ESD theory, an optimization model is proposed in this chapter to calculate the configurations of cameras in an irregular acquisition of an LF based FVV system. Theoretical analysis and numerical simulation have demonstrated that the rendered video quality is significantly improved by employing the proposed irregular acquisition.

Chapter 8: Conclusions and Future Works

The main contribution of this research is to propose a theory for analytical comparison, evaluation, and optimization of the LF acquisition and rendering components for LF-based FVV systems, particularly in the more realistic *under-sampled* LF and *approximated depth information* with errors in depth map. The core of the proposed theory is the concept of effective sampling density (ESD). ESD is shown to be an analytically tractable metric that represents the combined impact of the imperfections of LF acquisition and rendering and can quantify their impacts on the final video quality separately as well as jointly.

Chapter 3 shows that different LF rendering methods and LF acquisition configurations can be theoretically evaluated and compared using the proposed ESD. Eight well-known rendering methods with different acquisition configurations have been analyzed through ESD and simulation. The results have shown that ESD is an effective indicator of output distortion and quality that can be obtained directly form system parameters and that takes into consideration both acquisition and rendering. In addition, an empirical relationship between the theoretical ESD and achievable rendering quality (in PSNR) has been established which allows direct prediction of the overall video quality without the actual implementation of the system. Furthermore, a subjective quality assessment has confirmed that ESD is highly correlated with the perceived output quality.

Chapter 4 discusses the use of ESD for LF acquisition optimization. The acquisition optimization is simplified to regular camera grid acquisition and the number of cameras required to capture the scene which is an essential problem in a practical LF based FVV system. The ESD analysis is applied to under-sampled LF under realistic conditions (non-Lambertian reflections and occlusions) and rendering with complex interpolations. As a result of this analysis, a method for calculating the minimum number of cameras and evaluating the impact of

depth map errors on output quality for LF-based FVV systems is proposed in this chapter. It is shown that higher camera density can compensate for the adverse effect of depth map errors on the output quality. To employ the proposed method in LF based FVV system design, the desired rendering quality of the system in PSNR can be mapped to the corresponding ESD by employing the empirical model given as (4.14). This ESD with depth estimation error can be applied to (4.13) to calculate the camera density in the grid and hence the minimum number of cameras. Theoretical and numerical results showed that the resulting number of cameras is significantly lower than what was reported in the previous studies with only a few percent reduction in the rendering quality. Moreover, it was shown that the previous methods are special cases of the one derived from ESD theory.

Chapter 5 discusses the problem of LF rendering optimization by utilizing ESD theory. In particular, the optimization is applied to the depth-based LF rendering algorithms to provide an estimation of the rendering complexity in terms of optimum number of rays employed in interpolation algorithm so as to compensate for the adverse effect caused by errors in depth maps for a given rendering quality. To employ the proposed method in LF based FVV system design, the desired rendering quality of the system in PSNR can be mapped to the corresponding ESD by employing the empirical model given as (5.3). This ESD with depth estimation error is applied to (5.2) to calculate the optimum number of rays required for interpolation in rendering process. The proposed method is particularly useful in designing a rendering algorithm with inaccurate knowledge of depth to achieve the required rendering quality. Both the theoretical study and numerical simulations have shown that the proposed method was reliable and accurate.

Chapter 6 demonstrates a joint optimization of both LF acquisition and LF rendering to achieve a desired output quality. A Lagrangean joint optimization method of ESD for the (i) the density of cameras in the acquisition grid and (ii) the density of rays within the interpolation area was proposed in this chapter and after an extensive calculation, an analytical solution to the Lagrangean optimization was given as (6.11) and (6.12). In particular, this chapter studied the

tradeoff among acquisition camera density, ray selection, depth error and rendering quality and it is shown that the error in depth maps can be compensated for by judicious alteration of *both* variables, to maintain the rendering quality at a desired level. Employing the proposed method on a regular grid camera system has shown that the number of cameras can be reduced by 8 times if 32 rays, instead of 8 rays, are employed during rendering to achieve the similar rendering quality for a typical 20% error in depth estimation.

Chapter 7 proposes a non-uniform/irregular LF acquisition by superimposing the scene complexity in terms of scene spatial frequency variations to ESD theory. It was shown that while a regular acquisition itself results in non-uniform sampling density, this non-uniformity does not match the scene complexity and frequency variations. Subsequently, an optimization model (7.1) to (7.10) was proposed to compute the optimum configurations of the acquisition cameras including positions and orientations, i.e., optimum non-uniform/irregular LF acquisition, corresponding to the variations of the scene complexity. Specifically, scene complexity was measured through analyzing DCT coefficients of reference images of the scene, describing the frequency behavior of the plenoptic signal over the scene space. The theoretical analysis and numerical simulations demonstrated that the rendered video quality was significantly improved (around 20% in mean PSNR) by employing the proposed irregular acquisition compared with the regular camera grid.

The thesis also provides a number of appendices to cover the extensive mathematics. Appendix I gives an example for further demonstration of ω, Ω, and Θ sets. Appendix II gives details of SD calculation for a regular camera grid. Appendix III illustrates the details of ESD calculation for a regular camera grid. Appendix IV gives the analytical solution to Lagrangean optimization of ESD. Appendix V gives a brief overview on the proposed quantitative analysis of LF systems and simulation system used for validation of the proposed ESD theory.

8.1 Future Works

While the fundamentals of ESD theory has been studied and reported in this thesis, the theory has a long way to go. We are working on extending the ESD theory and applying it to more problems related to LF-based FVV systems evaluation, comparison, and optimization and will report the results in future publications. In particular following problems should be investigated with higher priority:

a) Chapter 3 shows the ESD derivations of eight rendering methods, for regular camera grid acquisition and 2-planes representation. While this was sufficient to introduce the theory and its application but the derivation of ESD for other well-known rendering methods and for more complicated acquisition architectures/topologies and LF representations should be carried out.

b) For the experimental validation and objective rendering quality assessment, basic PSNR metric was chosen throughout the thesis. The extension of this to other metrics which are more suitable for video quality assessment is very important. In addition, the subjective quality assessment should also be done. Regardless, the relationship between ESD and those experimental assessments should be empirically established similar to the one demonstrated for PSNR and ESD.

c) More research is required on the empirical model demonstrated for the relationship of the theoretical ESD and experimental output quality. This can be done by using better curve fitting methods as well as experiments with more scenes and in particular real scenes.

d) Throughout the thesis the LF analysis, evaluations, comparisons, and optimizations of ESD are demonstrated for one unknown ray r and the area of interpolation A at the vicinity of p the intersection point of r with the scene. This is a micro analysis of quality assessment for reconstruction of only on unknown ray r based on ESD. In this thesis as mentioned several times, the macro analysis of quality assessment of the system was given with an average of ESD

for a scene ($\overline{ESD}$). It was argued that sampling the scene with this calculated $\overline{ESD}$ guarantees the corresponding average output quality for the system. While this simplification was shown to be sufficient for uncomplicated scenes but a generic formulation of macro analysis of ESD is required for future works. This generic macro analysis can be formulated by performing the integral of ESD over the scene.

e) The acquisition optimization proposed in chapter 4 was simplified to a regular grid acquisition and calculation of the number of cameras for the gird. Extension of this optimization to other acquisition architectures and configuration, parameterizations, and acquisition parameters are required.

f) The rendering optimization proposed in chapter 5 was simplified to a regular grid acquisition and calculation of the number of rays employed for interpolation for UV-DM rendering method. Extension of this optimization to the other rendering methods, and other rendering parameters are required.

g) The interpolation method F was assumed to be fixed for all the methods proposed in this thesis. However, as speculated before in chapter 3, the efficacy of the interpolation method has a direct impact on the rendering quality. A mathematical analysis of various interpolation methods and superimposing the ESD theory by the efficacy of the interpolation method is also a very important problem for the future works.

h) While in chapter 7, the ESD theory was superimposed by the scene complexity, but the analysis demonstrated in chapter 4, 5, and 6 are assumed that the scene complexity is fixed. The extensions of the optimization methods proposed in those chapters by superimposing scene complexity is another important problem to be considered in future works.

i) The extension of the joint optimization method proposed in chapter 6 for other types of acquisition and rendering components is also an important problem for future works.

j) While the optimization method proposed in chapter 7 was generic, but several simplifications have been applied to the optimization model to solve it including substituting ESD by SD (i.e., not considering the rendering method), 2D version of the optimization for a irregular grid, and using the approximated *scene complexity maps*. The extensions of the proposed optimization for ESD (instead of SD), 3D acquisition topology, and calculating the scene complexity by assuming the depth information are all significant problems, required to be performed in future.

k) While the simulation model proposed in chapter 8 for quantitative analysis of LF-based FVV systems and validation of the proposed ESD theory was very effective but applying the theory to real scenes is very important in next stages. We already designed and implemented a camera grid in our laboratory and trying to apply the results of the ESD theory to the real scenes.

l) The same as any other signal processing applications, pre-filtering and post-filtering of LF signal are important problems. The optimum LF filtering to avoid anti-aliasing has been investigated before in several researches as discussed in literature review chapter. While the relation between optimum filtering and ESD was discussed in spectral analysis of LF in chapter 4, but the optimum amount and the type of filtering are among the important problems that required to be answered by using ESD theory in future.

m) As discussed before, LF can be considered as a discrete synthetic aperture. Employing more rays for interpolation would increases the size of the aperture and consequently decreases the depth of field. On the other hand, employing less number of rays would reduce the aperture size and results in increasing the depth of field. This size of the aperture is shown to be directly related to the LF filtering. While Chapter 5 proposed a method to calculate the optimum number of rays for interpolation but the optimum size of the discrete synthetic aperture for an optimum ray filtering is also a very important problem and we are working on this for a future publication.

n) The *LF compression and transmission component* was not the focus of this thesis and has not been investigated here. However, future extension of ESD theory requires a full analysis over LF compression and transmission methods and the relation between ESD and the required bandwidth. Hence, for a desired output quality, and system parameters the calculated ESD can be mapped to the minimum bandwidth required for the signal transmission.

Appendices

9.1 Appendix I: Demonstration of ω, Ω, and Θ sets

Appendix I describes an example for further demonstration of ω, Ω, and Θ sets in an LF-based FVV system. Denote from chapter 3 that Θ refers to a set of all rays captured by an LF acquisition component. Ω is a subset of rays from Θ that intersects the scene at A, the area of interpolation at vicinity of p, the intersection point of unknown ray r with the scene. All of the rays in Ω can be potentially employed in *interpolation process* to estimate r. However, practically, due to lack of knowledge about the scene geometry, as well as computational limitations only ω a subset of rays from Ω is selected by the *selection process* of the rendering method. Clearly, ω ⊆ Ω ⊆ θ. Figure 9.1 illustrates a sample 2D light field with 8 cameras and 8 pixels per image. It shows 64 rays in Θ, and r the unknown ray intersecting the scene on point p and A the area of interpolation souranding p.

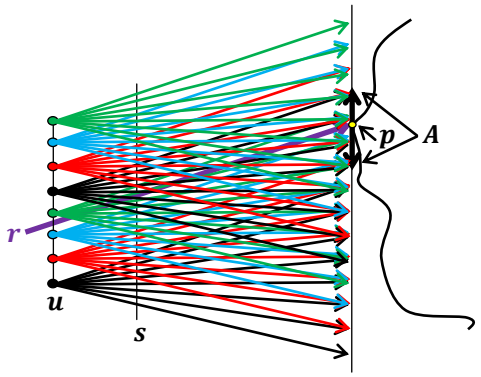


Figure 9.1. A sample 2D light field with 8 cameras and 8 pixels per image. It shows 64 rays in Θ, and r the unknown ray intersecting the scene on point p and A the area of interpolation souranding p

Figure 9.2 illustrates the same 2D light field with 22 rays in Ω, all the captured rays in the LF system that flowing through A, the area of interpolation. These 22 rays potentially could

be employed in a rendering method to estimate r. However, practically, just a subset of them would be used for interpolation as discussed before.

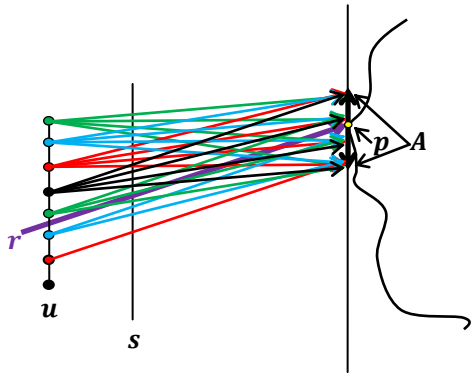


Figure 9.2. The same 2D light field as Figure 9.1 with 22 rays in Ω, all the rays in the system that flowing through A, the area of interpolation

Figure 9.3 shows the same 2D light field with 7 rays in ω , a subset of rays from Ω selected by an imaginary rendering method to be employed for interpolation of unknown ray r. B_1 and B_2 are the boundary rays souranding all other rays in ω which can be employed to calculate the area of interpolation A. We will be shown in Appendix III, effective sampling density (ESD) could be calculated by deriving the line equations for boundary rays B_1 and B_2 and calculating A.

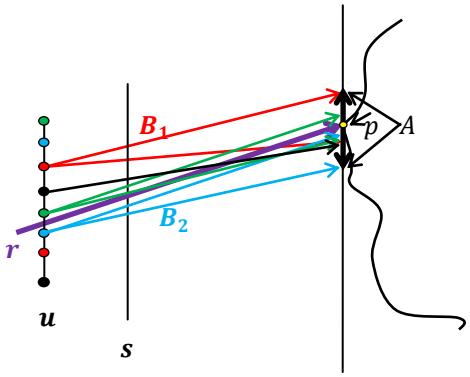


Figure 9.3. The same 2D light field as Figure 9.2 with 7 rays in ω, a subset of rays in Ω selected by rendering method to be employed in interpolation of unknown ray r. B_1 and B_2 are the boundary rays souranding all other rays in ω creating the area of interpolation A.

9.2 Appendix II: Details of SD (Sampling Density) Calculation for a Regular Camera Grid LF Acquisition

This appendix demonstrates an approach for calculating the SD for an LF-based FVV system. The main problems in calculating the SD are to determine the area of interpolation A and to calculate the number of rays in Ω for that A for a given acquisition configuration. The area of interpolation A, itself is determined by the *selection process* of the rendering component and the amount of error in depth estimation. In this appendix, the ideal interpolation area by assuming no errors in depth estimation is assumed as $A = (ld)^2$, which is the system resolution as discussed before. In addition, to calculate Ω for this ideal A, a regular camera grid for LF acquisition is considered. Clearly, SD can be calculated in any points of the scene from: $\text{SD} = \frac{|\Omega|}{A}$. The proposed approach can be extended to any acquisition architecture and configuration and for any size of interpolation area A. Subsequently, SD can be illustrated as a contour chart throughout the scene space.

The problem of calculating $|\Omega|$ is corresponding to calculating the number of cameras able to see a specific part of the scene determined by A. To calculate the number of cameras, let's start with a basic demonstration of a pinhole camera in an LF system in Figure 9.4. Camera vertical angle of view, pixel length, image length, and the length of camera's field of view in depth d are demonstrated in Figure 9.4

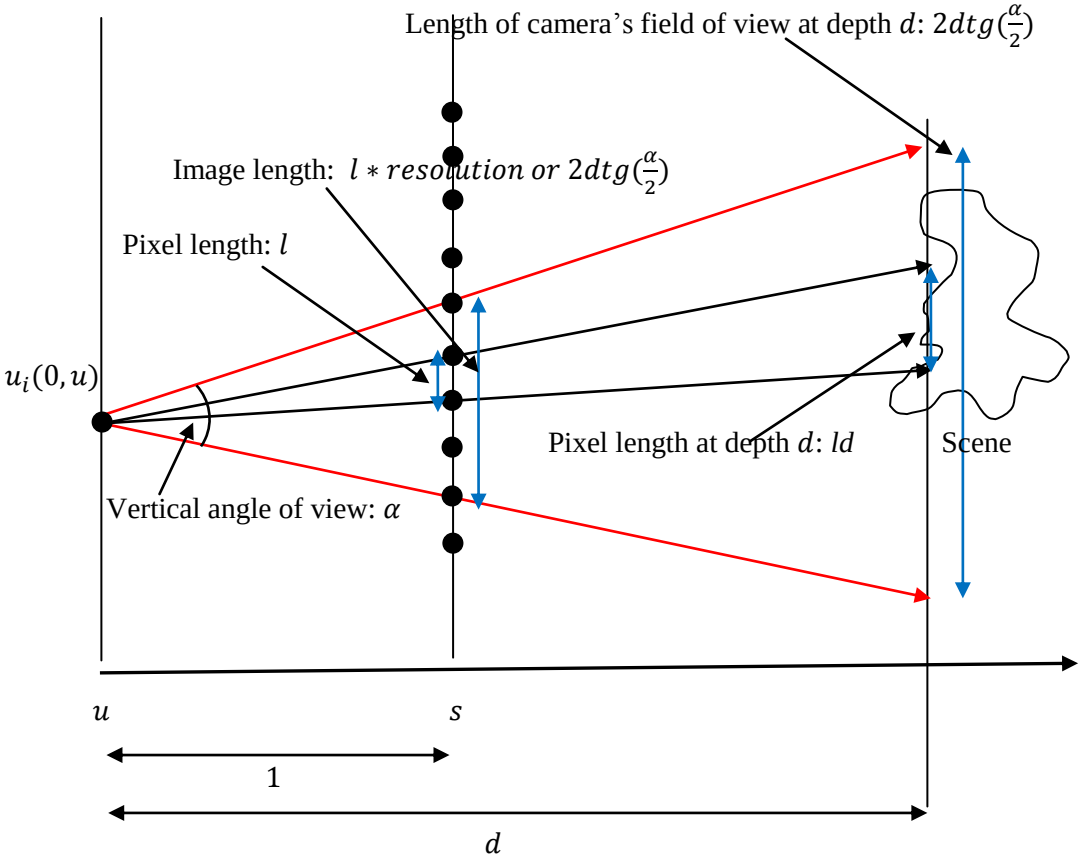


Figure 9.4. A pinhole camera in an LF acquisition and the basic geometric concepts

Figure 9.5 illustrates a row of cameras in u and how these cameras field of view overlap. The distance between two adjacent cameras is k which makes imaginary discrete line segments in the scene space with the same length k. Each line segment can be seen by a number of cameras which can be calculated based on the overlaps of the field of views. The same concept can be easily extended to a regular camera grid uv and rectangular segments.

As it can be seen from Figure 9.5, there is a pattern for number of cameras can see the segments as depth d. The very first and very last segments could only be seen by one camera but intermediate sections could be seen by three cameras in this example. In general scenario, there is still a pattern for the number of cameras seeing each segment at depth *d*. This pattern is incremental starting from one at the borders of the LF system field of view, incrementing to a fixed maximum in the centre as: $\{1,2,3,4,\dots,max-1,max,max,\dots max,max-1,\dots 4,3,2,1\}$.

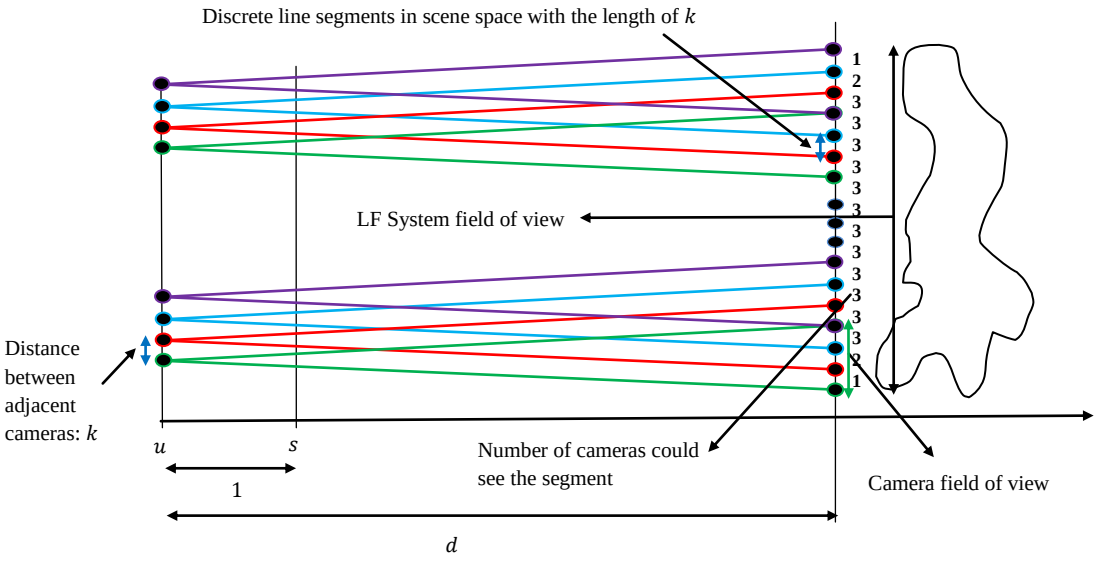


Figure 9.5. A sample 2D light field system showing the scene space discretization and the number of cameras able to see each discrete line segment at depth d

Following equations demonstrate how to calculate the geometric parameters demonstrated in Figure 9.4 and 9.5. FOV stands for "field of view" in these equations.

$$FOV_{camera} = 2dtg\left(\frac{\alpha}{2}\right) \tag{9.1}$$

and α is the camera vertical or horizontal angle of view.

$$FOV_{LF\ system} = (n-1)k + 2dtg\left(\frac{\alpha}{2}\right) \tag{9.2}$$

and n is the number of cameras.

$$number\ of\ segments = \frac{FOV_{LF\ system}}{k} = \frac{(n-1)k+2dtg\left(\frac{\alpha}{2}\right)}{k} \tag{9.3}$$

And the resolution of the LF system at depth d can be calculated as:

$$FOV_{LF\ system}\ resolution = \min\left(n.camera\_resolution, \frac{FOV_{LF\ system}}{ld}\right) \tag{9.4}$$

$camera\_resolution$ is the resolution of the cameras and ld is the length of the pixel at depth d. For depth d close to st image plane not all of the LF system FOV could be seen by

cameras and there are some dark regions. In this scenario, $n.camera\_resolution$ gives the system resolution in (9.4). Nevertheless, further from this initial depth, overlapping among cameras is accrued and the second part of (9.4) gives the resolution. Please note that the system resolution after overlapping would reduce from original $n.camera\_resolution$.

To compute the pattern of $\{1,2,3,4,\ldots,max-1,max,max,\ldots max,max-1,\ldots 4,3,2,1\}$ shown in Figure 9.5, parameter max should be calculated. Note that max refers to the maximum number of cameras able to see a segment in central zone of FOV at depth *d*.

$$max = \text{maximum}\left(\left[\frac{FOV_{camera}}{k}\right], n\right) = \text{maximum}\left(\left[\frac{2dtg\left(\frac{\alpha}{2}\right)}{k}\right], n\right) \tag{9.5}$$

and n is the number of cameras.

Figure 9.6 shows a sample 2D light field with a regular row of 30 cameras with $k = 2$, camera vertical angle of view is 30^o, camera resolution is 100 pixels which makes pixel length $l = 0.005358984$, and d is between [2,100]. The scene space is discretized with line segments with $length = 2$ (equal to camera distance as discussed before) and $width = 1$.

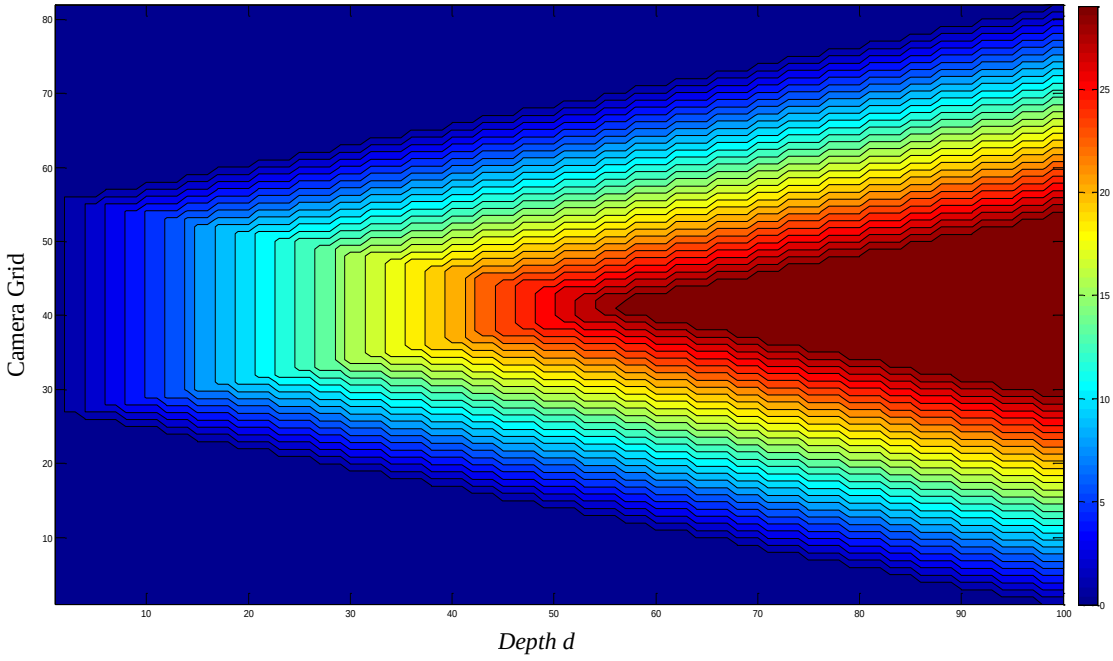


Figure 9.6. Number of cameras able to see a discrete segment in the scene space for a sample 2D light field

Note that the scene space is already discretized based on the pixel size which is a natural phenomena in digital photography. To align the discretization process based on the line segments with the underneath natural pixel discretization, it is assumed in Figure 9.6 that in each depth d, each segment has integer number of pixels to have a unambiguous contour. Without this simplification, there are some aliasing in the boarders of each segment, where there is a pixel shared between two adjacent line segments. This assumption could be mathematically explained as:

$$For\ each\ discrete\ value\ of\ d, i = \frac{k}{ld}\ \ and\ i\ is\ an\ integer \tag{9.6}$$

Also it is assumed that there are always several pixels in each segment, i.e., the discretization based on the line segment is coarser than pixel discretization. Mathematically,

$$For\ each\ discrete\ value\ of\ d, k >> ld\ \ \rightarrow\ \ d \ll \frac{k}{l} \tag{9.7}$$

If the condition of (9.7) is not satisfied, i.e., pixel length is larger than the segment length for large d> $\frac{k}{l}$, the calculation and contour visualization should be carried out based on the pixel length ld rather than segment length k. Without losing generality, this case is not demonstrated here.

The number of rays in |Ω| can be calculated from the number of cameras demonstrated in Figure 9.6 for each segment. It is know from (9.6) that there are $\frac{k}{ld}$ pixels in each segment, hence there are $\frac{k}{ld}$ rays from each camera intersect that given segment. Thus,

$$|\Omega|\ for\ given\ segment = number\ of\ cameras\ seeing\ the\ segment.\frac{K}{ld} \tag{9.8}$$

By assuming the area of interpolation A to be equal to the segment length, i.e., $A = k$, SD for each segment can be calculated as:

$$SD\ at\ given\ segment = \frac{|\Omega|\ for\ given\ segment}{A} =$$

$$\frac{number\ of\ cameras\ seeing\ the\ segment.\frac{k}{ld}}{k} = \frac{number\ of\ cameras\ seeing\ the\ segment}{ld} \tag{9.9}$$

Equation (9.9) demonstrates that SD can be calculated from a nonlinear transform of number of cameras seeing each segment. By applying this nonlinear transform to Figure 9.6, a contour of SD in scene space is demonstrated in Figure 9.7.

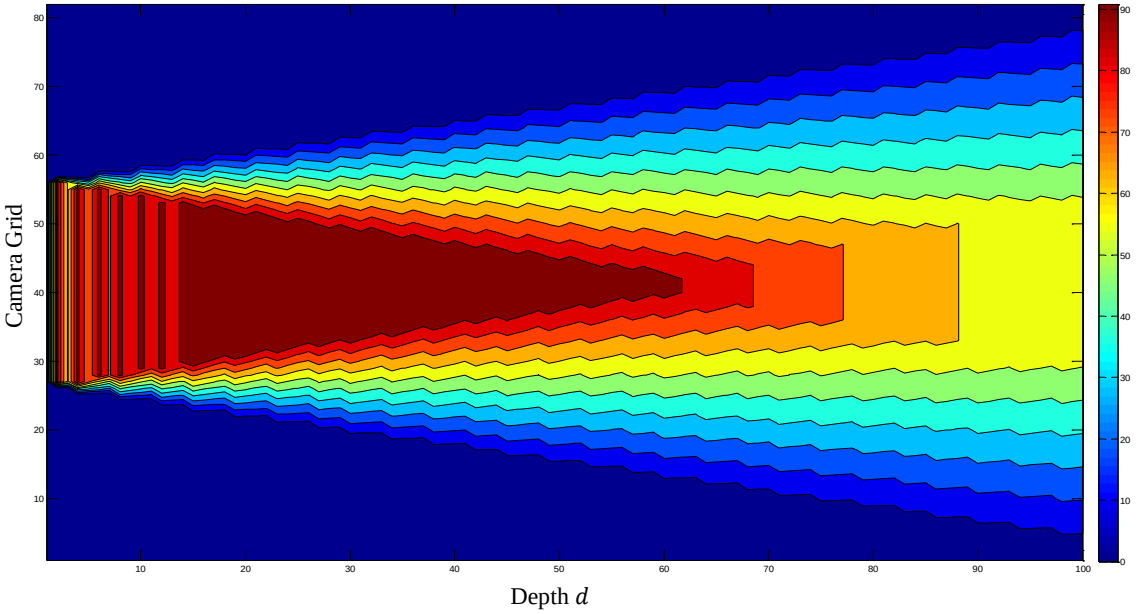


Figure 9.7. Sampling Density (SD) contour for a sample 2D light field

Notice from Figure 9.7 that SD of a point in the scene decreases when it moves further from the camera grid and away from the centre of the scene. A semi-triangle in the centre of the scene with one edge on s, has the highest SD in 2D LF.

For 3D light field with a regular camera grid acquisition the same results can be generalized as the configuration is symmetric. Hence,

$$SD\ for\ a\ discrete\ rectangular\ segment\ in\ 3D\ light\ field =$$

$$\left(\frac{number\ of\ cameras\ seeing\ the\ line\ segment}{ld}\right)^2 \tag{9.10}$$

Figure 9.8 demonstrates a 3D light field with a regular camera grid acquisition with the same system parameters as Figure 9.7

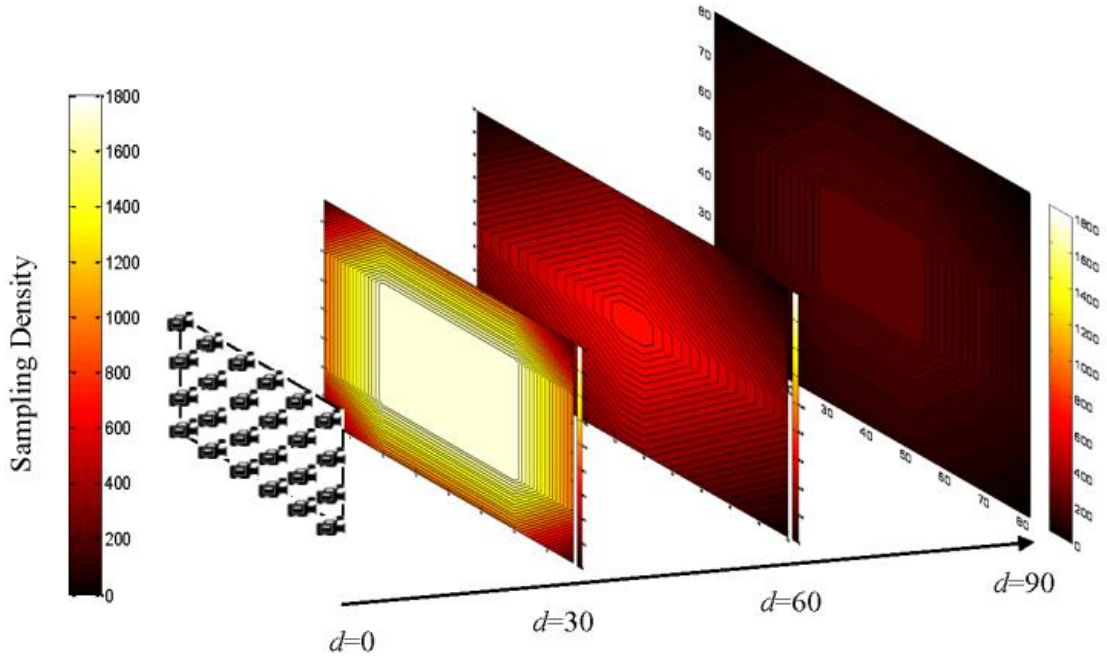


Figure 9.8. Sampling Density (SD) contour for a sample 3D light field with regular camera grid acquisition

9.3 Appendix III: Details of ESD (Effective Sampling Density) Calculation for a Regular Camera Grid LF Acquisition

This appendix describes the mathematical methodology to calculate ESD for a regular camera grid acquisition. A simple scenario for a 2D light field is demonstrated here which can be easily extended to 3D light field and arbitrary ray selection as described before in chapter 3. Figure 9.9 illustrates a 2D light field, an unknown ray r and a subset of acquired rays ω employed by rendering process to estimate r. Two rays B_1 and B_2 are boundary rays and all rays in ω geometrically bounded by them as discussed before in Appendix I. In this simple scenario, *selection mechanism M* is selected a subset of rays in ω bounded by n cameras and m pixels. A is the area of interpolation generated by intersections of rays in ω with the scene. The length of A in a simplified 2D LF can be computed from the distance between B_1 and B_2 at depth d. Mathematically A can be expressed as a function of k, l, n, m, and d :

$$A(k,l,n,m,d) = |B_1(d) - B_2(d)| \tag{9.11}$$

The line equations of B_1 and B_2 rays could be easily calculated from their intersections with u : $(U_i(0,u), U_{i+n}(0,u+n.k))$ and s: $\left(S_j(1,s), S_{j+m}(1,s+m.l)\right)$ where the distance between u and s is assumed to be 1, the distance between adjacent cameras is k and the pixel length is l. In addition to *k*, *l*, *n*, *m*, and *d* A can be also a function of G the geometric information of the scene generated by *depth estimation mechanism* (refer to chapter 3). For the corresponding 3D light field, the area of interpolation can be computed as A^2 by assuming a symmetrical ray selection.

The effective sampling density ESD can then be calculated from ESD $= \frac{|\omega|}{A}$ as discussed in chapter 3.

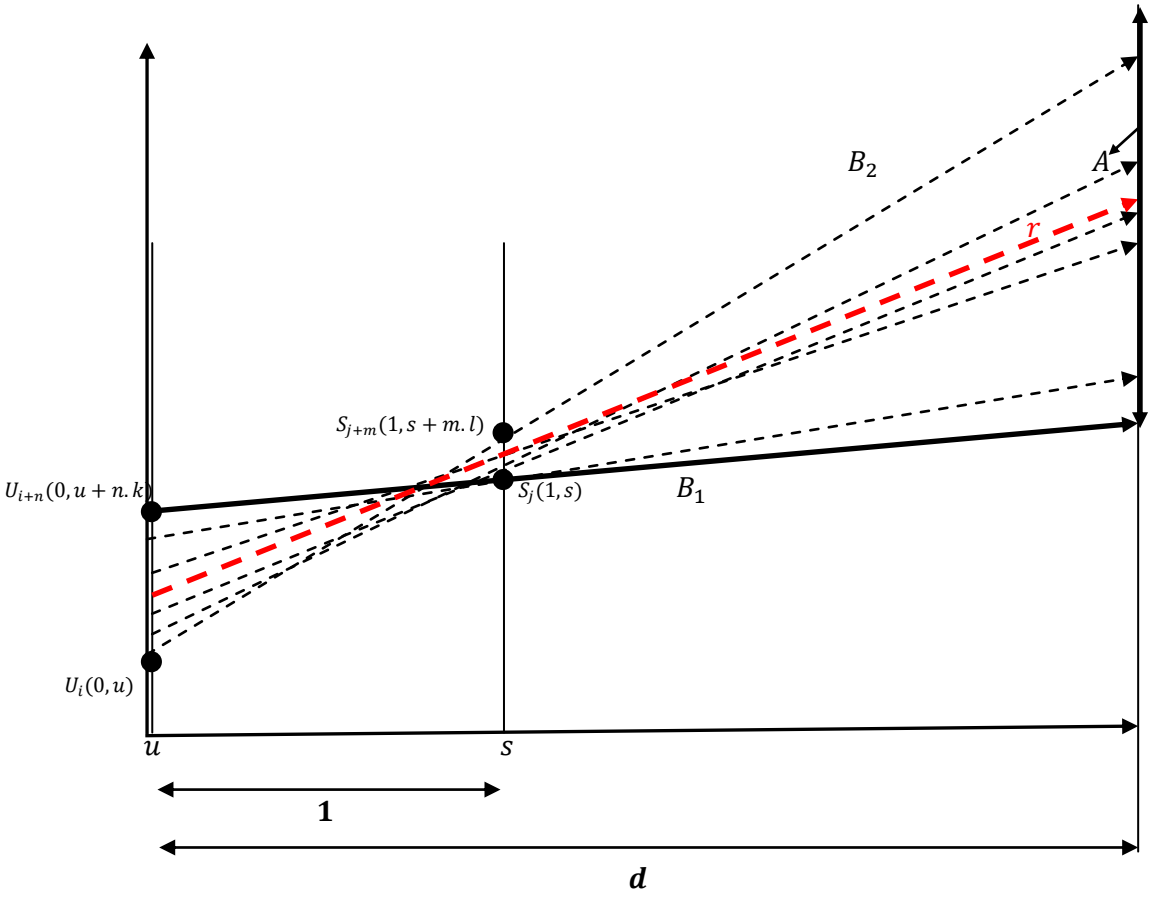


Figure 9.9. Effective Sampling calculation for a simplified 2D LF system

9.3.1 ESD for Blind LF Rendering Methods

9.3.1.1 UV Interpolation

To estimate the unknown ray *r* in UV interpolation, the nearest known neighbouring ray to the intersection point of r with the st plane is chosen. Then four known rays from four immediate neighbour cameras in the uv plane to that point are interpolated by employing a 2D interpolation method such as bilinear interpolation. Figure 9.10 shows the simplified 2D light field with two lines u and s instead of two slabs uv and st. Necessary rays for interpolation are demonstrated in this Figure. For any rays intersecting u between U_1 and U_2 and s between S_1 and S_m (rays bounded by $B_1 = Y_5$ and $B_2 = Y_3$), rays Y_1 and Y_3 and between S_m and S_2 (rays bounded by $B_1 = Y_2$ and $B_2 = Y_9$), rays Y_2 and Y_4 are chosen to be interpolated, where S_m is the

middle point between S_1 and S_2. Let's assume A_{UV} is the length of interpolation area between bounded rays in depth d. All the rays flowing through A_{UV} are interpolated from two rays. Due to symmetric configuration $A_{UV} = Y_5 - Y_3 = Y_2 - Y_9$.

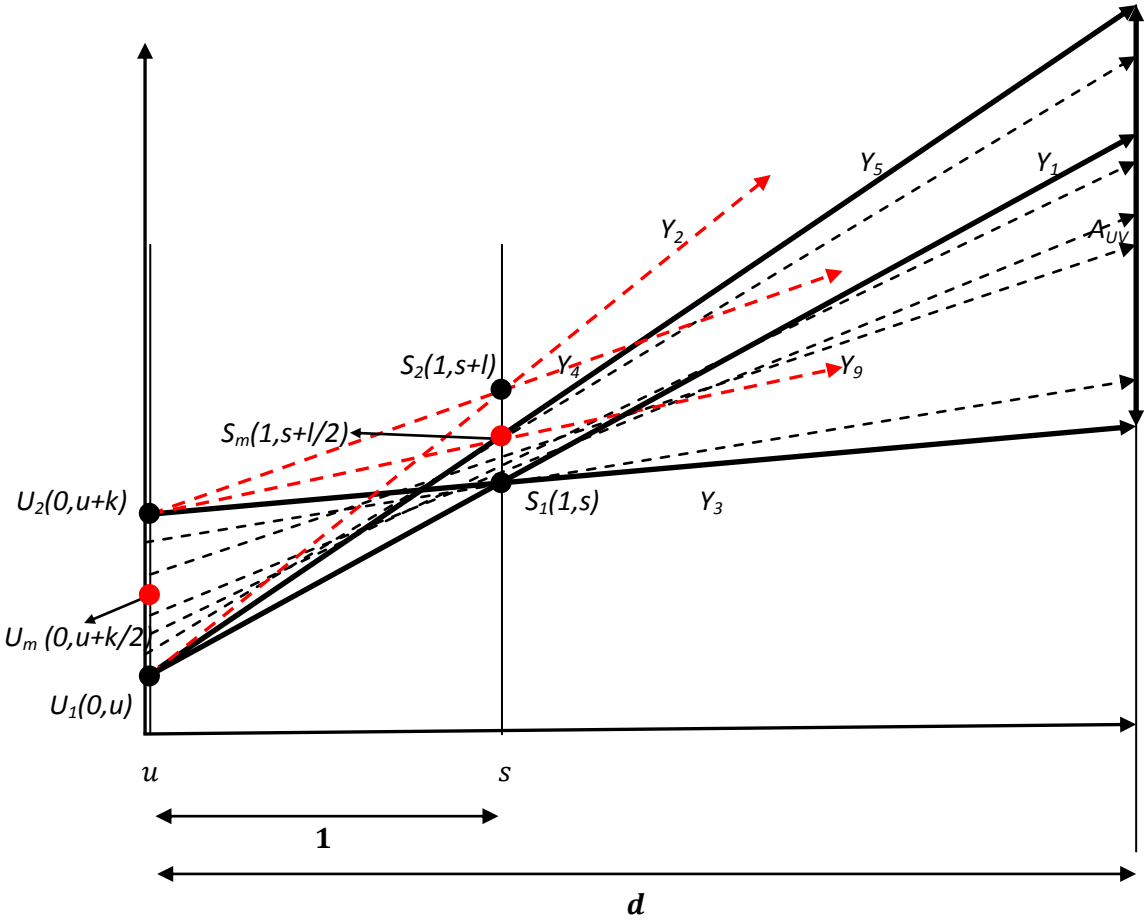


Figure 9.10. A simplified 2D LF system with required rays for UV interpolation ESD calculation

9.3.1.2 ST, UVST and NN Interpolations

The same procedure can be employed for ST, UVST and NN methods. In ST interpolation a bilinear interpolation in st plane is applied. In the nearest neighbourhood estimation (NN), the unknown ray is estimated with a ray with nearest known (s,t) in st and (u,v) in uv. In UVST method, a quadrilinear interpolation is applied to all 16 known rays, surrounded the given unknown ray. In the simplified 2D LF representation, the bilinear

interpolation in UV and ST methods are simplified to linear interpolation of two rays and quadrilinear interpolation in UVST to a bilinear interpolation. Figure 9.11 demonstrates all the rays required for blind LF rendering methods ESD calculation.

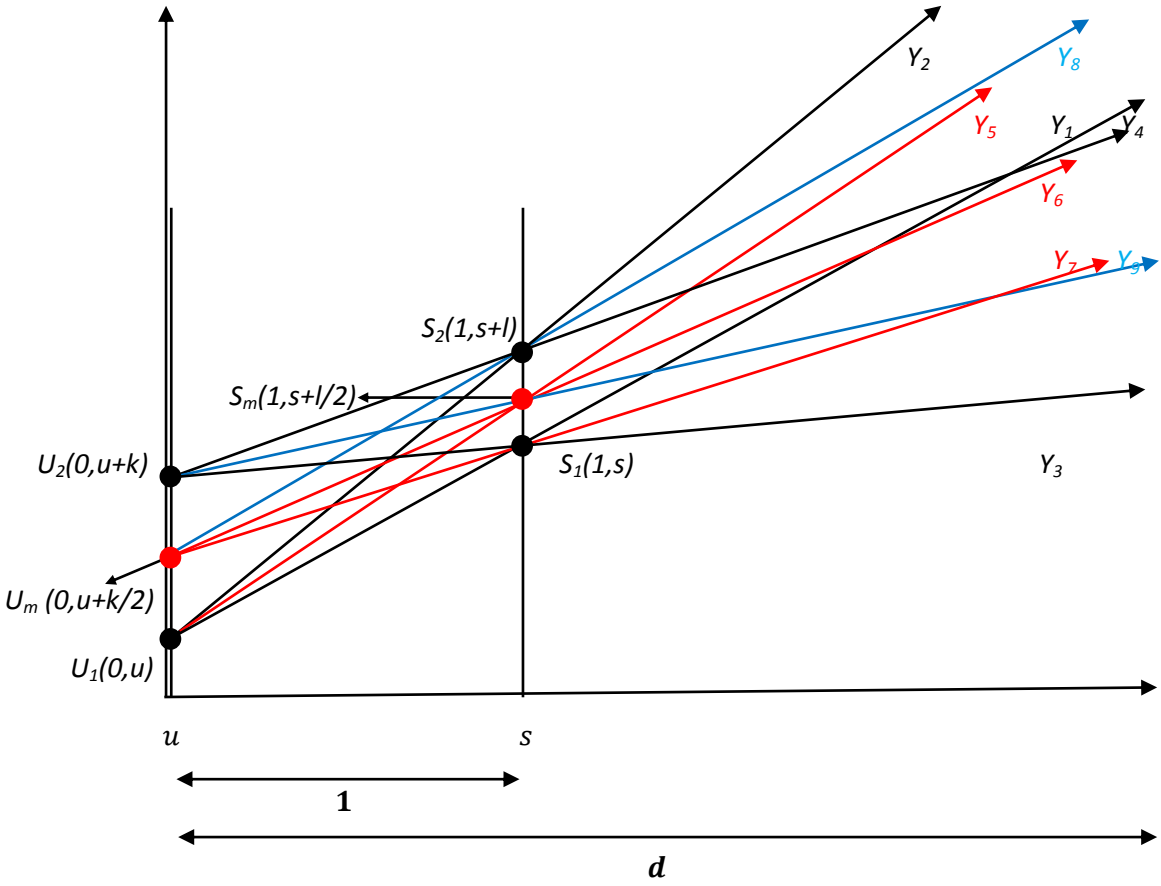


Figure 9.11. Required rays for LF blind rendering methods ESD calculation

Table 9.1 summarizes the line equations for all the rays demonstrated in Figure 9.11. Table 9.2 summarizes the interpolation area *A*, number of rays employed in interpolation, and ESD calculation for the blind LF rendering methods in both 2D and 3D LF representations.

Table 9.1

Line equations for all the rays demonstrated in Figure 9.11

Ray symbol	***Line equation***
Y_1	$Y_1 = (s-u)X + u$
Y_2	$Y_2 = (s+l-u)X + u$
Y_3	$Y_3 = (s-u-k)X + (u+k)$
Y_4	$Y_4 = (s+l-u-k)X + (u+k)$
Y_5	$Y_5 = \left(s + \frac{l}{2} - u\right)X + u$
Y_6	$Y_6 = \left(s - u + \frac{l-k}{2}\right)X + (u + \frac{k}{2})$
Y_7	$Y_7 = \left(s - u - \frac{k}{2}\right)X + (u + \frac{k}{2})$
Y_8	$Y_8 = \left(s + l - u - \frac{k}{2}\right)X + (u + \frac{k}{2})$
Y_9	$Y_9 = \left(s + \frac{l}{2} - u - k\right)X + (u + k)$

9.3.1.3 Discussions

Please note that the above calculations in particular, Figure 9.11 and Table 9.2, demonstrate the ESD calculation for the simplified blind LF rendering method when the $selection\ process$ of rendering component only selects the immediate neighbours of the unknown ray r. The details of general case, has been completely discussed before in chapter 3. The main aim of this section was to show how the ESD can be calculated with applying geometric concepts to LF system. The same approach can be employed to calculate the ESD for any given LF system. The general ray equation will be discussed later in section 9.3.3.

Table 9.2

Summary of ESD calculation for the simplified blind LF rendering methods

Rendering Method	***Area of interpolation A for 2D LF***	***A in 3D LF***	***Number of rays employed in interpolation |ω| for 2D LF***	***|ω| in 3D LF***	**ESD** $= \frac{\|\omega\|}{A}$
NN	$A_{NN} = Y_5(d) - Y_7(d) = Y_8(d) - Y_9(d) = Y_2(d) - Y_6(d) = Y_6(d) - Y_3(d) = (\frac{l+k}{2})d - \frac{k}{2}$	$A_{NN}{}^2$	1 Y_1 / Y_4 / Y_2 / Y_3	1	$ESD_{NN} = \frac{1}{A_{NN}{}^2}$
ST	$A_{ST} = Y_2(d) - Y_7(d) = Y\ (d) - Y_3(d) = (l + \frac{k}{2})d - \frac{k}{2}$	$A_{ST}{}^2$	2 (Y_1,Y_2) / (Y_3, Y_4)	4	$ESD_{ST} = \frac{4}{A_{ST}{}^2}$
UV	$A_{UV} = Y_5(d) - Y_3(d) = Y\ (d) - Y_9(d) = (k + \frac{l}{2})d - k$	$A_{UV}{}^2$	2 (Y_1,Y_3) / (Y_2, Y_4)	4	$ESD_{UV} = \frac{4}{A_{UV}{}^2}$
UVST	$A_{UVST} = Y_2(d) - Y_3(d) = (l + k)d - k$	$A_{UVST}{}^2$	4 (Y_1,Y_2, Y_3, Y_4)	16	$ESD_{UVST} = \frac{16}{A_{UVST}{}^2}$

9.3.2 ESD for LF Rendering Methods with Depth Information

In the previous section, it was assumed that no depth information is available. As discussed before in chapter 3, by employing an estimated depth, the ESD and hence the rendering quality can be improved. Figure 9.12 demonstrates a simple rendering method using minimal depth information such as object focusing depth, which we refer to as UV-D. In this simple version of UV-D, just two immediate neighbours are interpolated to estimate the unknown ray r. In more advanced versions of UV-D, all the rays that intersect the object surface on *interpolation area A* at depth d could be employed in the rendering process.

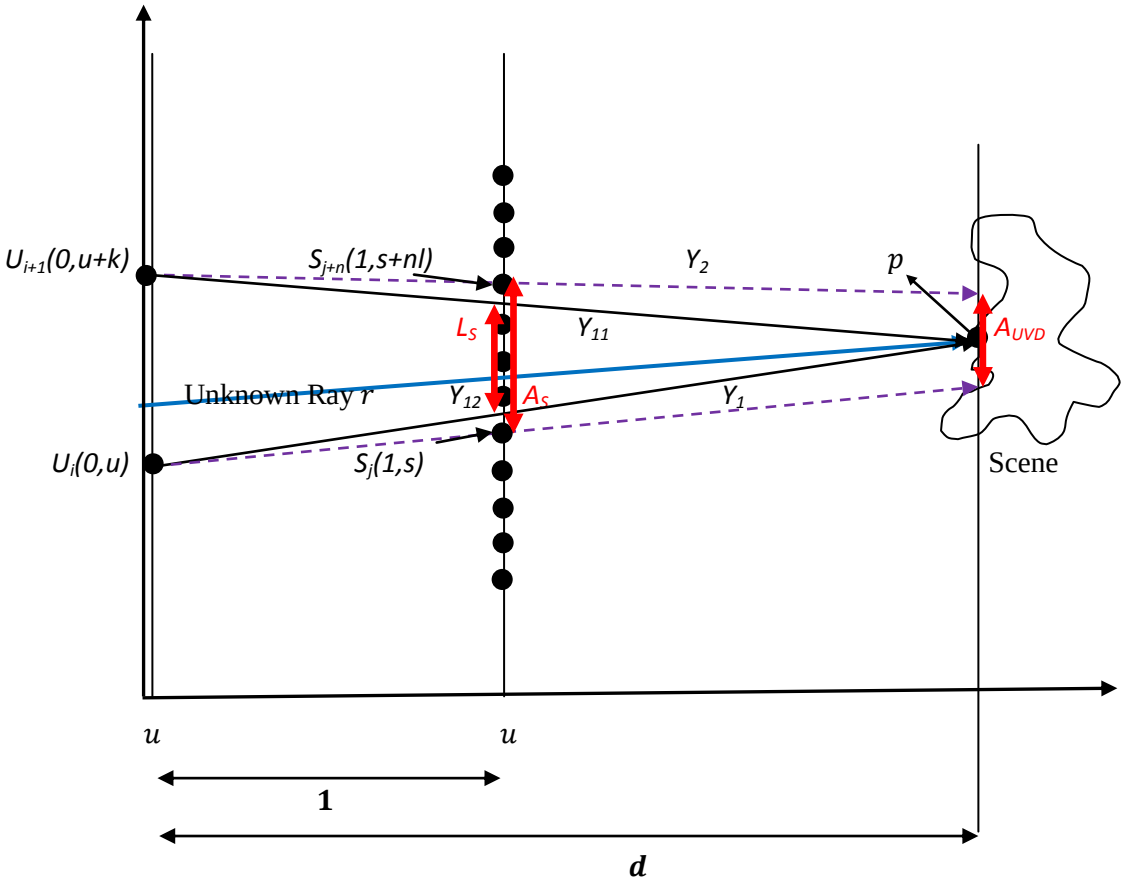


Figure 9.12. Simplified 2D LF representation with employing focused depth d for a better ray selection in UV-D rendering method

Note from Fgiure 9.12 that it is assumed that exact intersection point of unknown ray r with the object surface is known as p at depth d. A_{UVD} refers to the interpolation area for UV-D method, Y_{11} and Y_{12} are two immediate neighbour rays, intersecting with r on point p at depth d on the object surface. If these two rays pass through the known s values, $A_{UVD} = 0$, if not, neighbourhood or bilinear interpolation could be used to estimate them. In both cases Y_1 and Y_2 are boundary rays used for interpolation. A_{UVD} is equal to the distance between Y_1 and Y_2 at depth d. Approximated Y_{11} and Y_{22} are interpolated in a bilinear way on uv to estimate r.

With triangular similarity it is obvious that:

$$\frac{d-1}{d} = \frac{L_s}{k} \Rightarrow L_s = \frac{k(d-1)}{d} \tag{9.12}$$

The worst case scenario happens when the unknown ray (Y_{11} or Y_{12}) is in the middle of the known s values.

$$A_S = L_S + \frac{l}{2} + \frac{l}{2} = \frac{k(d-1) + ld}{d} \tag{9.13}$$

And it can be easily shown that:

$$A_{UVD} = ld \tag{9.14}$$

In reality the exact depth d of point p is not known and the estimated depth d of p has a Δd error. p' refers to the estimated p with Δd error. Again the same procedure is taken by the *selection process* for the estimated p' as demonstrated in Figure 9.13.

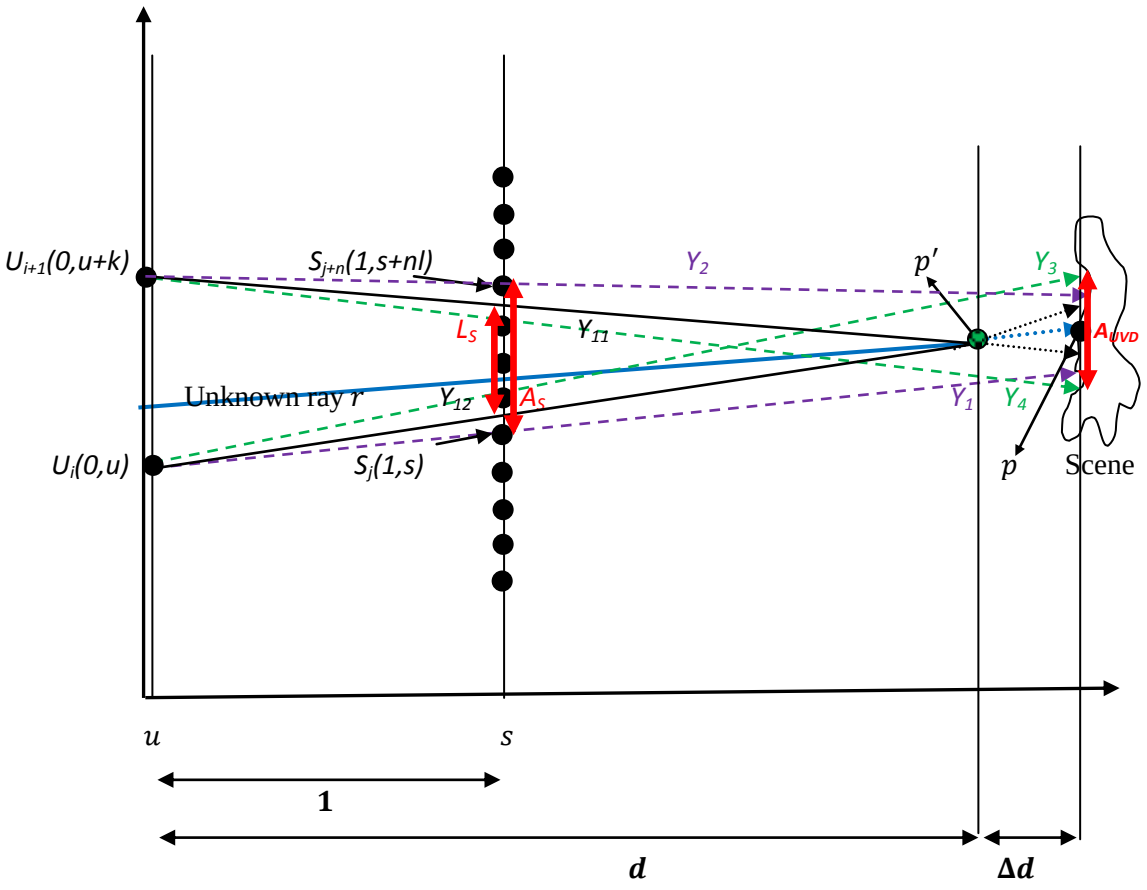


Figure 9.13. General UV-D rendering with approximated point p with Δd error in depth estimation

In the general UV-D rendering presented in Figure 9.13, again the rays intersecting in approximated point p' at depth d are found, however the real point p on the object surface is somewhere in the range of $d \pm \Delta d$. Clearly, this error in depth estimation would increase the area of interpolation A_{UVD} as:

$$A_{UVD} = \max[\,|Y_2(d+\Delta\mathrm{d}) - Y_1(d+\Delta\mathrm{d})|, |Y_3(d+\Delta\mathrm{d}) - Y_4(d+\Delta\mathrm{d})|\,] \tag{9.15}$$

and,

$$|Y_2(d+\Delta\mathrm{d}) - Y_1(d+\Delta\mathrm{d})|$$

$$= l(d+\Delta\mathrm{d}) - |Y_{11}(d+\Delta\mathrm{d}) - Y_{12}(d+\Delta\mathrm{d})| = \left|l(d+\Delta\mathrm{d}) - \frac{\Delta\mathrm{d.k}}{\mathrm{d}}\right| \tag{9.16}$$

and,

$$|Y_3(d+\Delta\mathrm{d}) - Y_4(d+\Delta\mathrm{d})| = |Y_{11}(d+\Delta\mathrm{d}) - Y_{12}(d+\Delta\mathrm{d})| + l(d+\Delta\mathrm{d}) =$$

$$\left|l(d+\Delta\mathrm{d}) + \frac{\Delta\mathrm{d.k}}{\mathrm{d}}\right| \tag{9.17}$$

From equations (9.15), (9.16) and (9.17), it can be concluded that:

$$A_{\mathrm{UVD}} = l(d+\Delta\mathrm{d}) + \frac{\Delta\mathrm{d.k}}{\mathrm{d}} \tag{9.18}$$

Subsequently, the $\mathrm{ESD_{UVD}}$ for 3D light field can be computed as:

$$\mathrm{ESD_{UVD}} = \frac{4}{{A_{\mathrm{UVD}}}^2} \tag{9.19}$$

In above discussions and (9.19) it is assumed that only immediate neighbours of r are selected as ω, i.e., $|\omega| = 4$. To model a more complex surface reflection behaviour, more rays from Ω can be selected as ω by selection mechanism M to be employed in interpolation F. Here Ω is a set of all the rays intersecting the interpolation area A in vicinity of point p (or more realistic p' when the error estimation has an error $\Delta\mathrm{d}$). Figure 9.14 illustrates a very simple scenario of the rays in Ω for the case when the exact point p is known. The extension of this to the case with estimation error is straight forward.

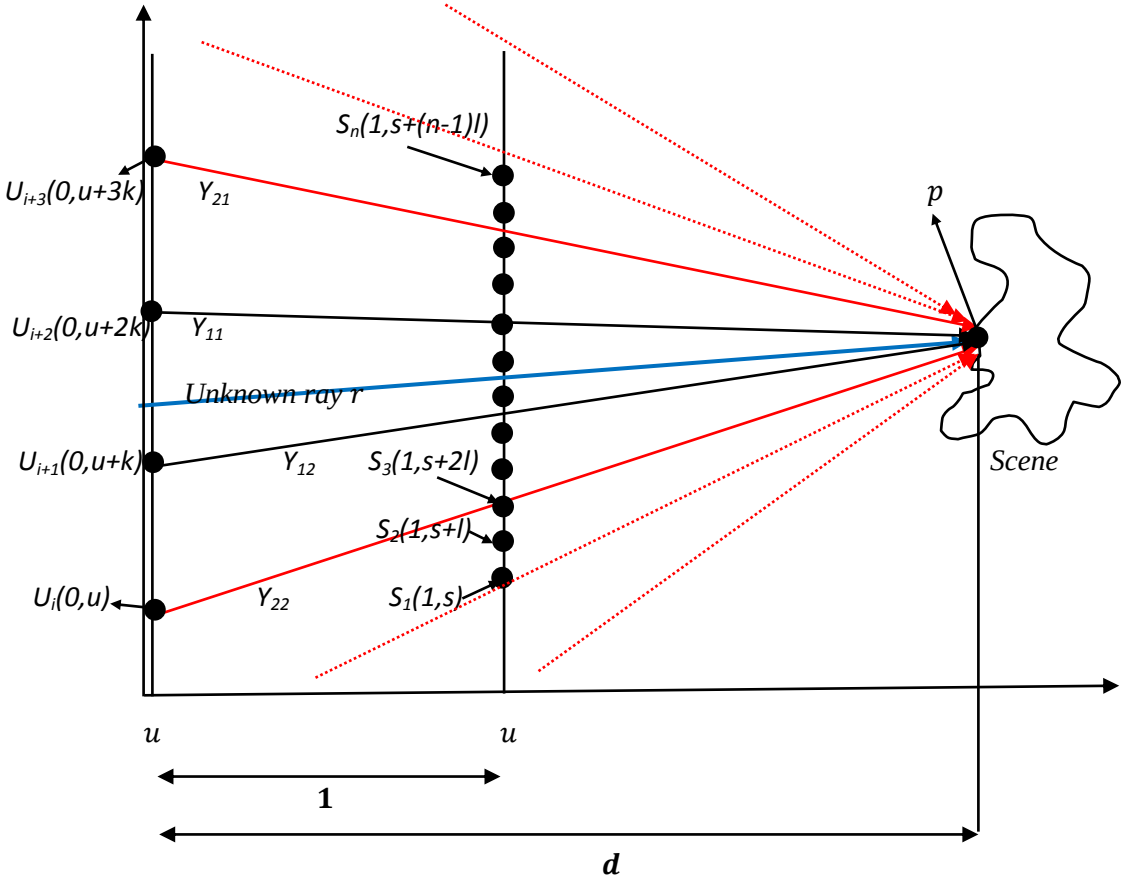


Figure 9.14. UV-D rendering with exact depth information with more rays selected as ω from available rays in Ω

9.3.2.1 Discussions

This section provides more details of the geometry and ESD calculation for LF rendering methods with depth information as a supplementary to chapter 3. UV-D rendering method with exact and estimated depth information has been chosen as a candidate to demonstrate the geometry. However, the full extension of the analysis of UV-D to UVST, UV-DM, and UVST-DM as well as arbitrary ray selections have been completely discussed in chapter 3 and has not been repeated here.

9.3.3 General Ray Equation

As it is shown in previous chapters, the main method to calculate ESD is based on deriving the ray equations for boundary lines. In addition, to select the rays during LF rendering, in particular form the ω set for methods with depth information, deriving the ray equation and computing its intersection with uv and st planes is a very critical problem in LF rendering. This section presents a vector representation of line equations for all the rays passing through a given point $p\ (p_x, p_y, p_z)$ in the scene space. This analysis is mainly based on the representation proposed in [11].

With assuming the image plane at $z = -1$ and camera plane at $z = 0$, it is obvious from camera pinhole model that:

$$r = \left(0,0,-\frac{p_x}{p_z},-\frac{p_y}{p_z}\right) + u.\left(1,0,1+\frac{1}{p_z},0\right) + v.(0,1,0,1+\frac{1}{p_z}) \tag{9.20}$$

By rewriting (9.20) based on the ray parameters in LF representation, we have:

$$r(u,v,s,t) = \left(0,0,-\frac{p_x}{p_z},-\frac{p_y}{p_z}\right) + u.\left(1,0,1+\frac{1}{p_z},0\right) + v.(0,1,0,1+\frac{1}{p_z}) \tag{9.21}$$

Equation (9.21) indicates all the rays passing through the given point $p\ (p_x, p_y, p_z)$ and intersecting uv camera plane.

Equation (9.21) can be rewritten based on the scalar disparities along epipolar lines. Note that cameras are regularly located in rows and columns of a grid, hence the images are rectified and their epipolar lines are horizontal or vertical. In other word, the disparities can be described as horizontal and/or vertical shifts between the corresponding pixels of image pairs. Each correspondence requires two rays to be represented. Let's demonstrate them with $r_1(\ u_1, v_1, s_1, t_1)$ and $r_2(\ u_2, v_2, s_2, t_2)$. It is easy to show:

$$\frac{s_2 - s_1}{u_2 - u_1} = \frac{t_2 - t_1}{v_2 - v_1} = \frac{1}{p_z} = disp \tag{9.22}$$

Subsequently, (9.21) could be rewritten in term of its disparity by employing (9.22) as :

$$r(u,v,s,t) = (0,0,s_0,t_0) + u.(1,0,disp,0) + v.(0,1,0,disp) \quad (9.23)$$

s_0 and t_0 could be directly calculated from r_1 and r_1. Again (9.23) can be used to compute all the rays passing through a given 3D point in space. In addition, for a given camera or a given intersection point (u,v) in uv plane the corresponding (s,t) can be directly calculated from (9.23). The solutions of (9.23) with integer values of u and v is correspondence to a back projection procedure from the point p to the real cameras at camera plane uv. This back projection procedure can compute all the rays in Ω.

9.4 Appendix IV: Analytical Solution to Lagrangean Optimization of ESD

Chapter 6 introduced a joint optimization on ESD by applying Lagrangean optimization. Though the optimization method has been completely discussed in chapter 6, but the details of analytical solutions derivation has been postponed to this appendix.

Partial derivation of $ESD_{UVDM(d,\Delta d,k,l,|\omega|)}$ based on $|\omega|$ can be calculated as follow:

$$\frac{\partial ESD_{UVDM(d,\Delta d,k,l,|\omega|)}}{\partial|\omega|} = \frac{\left(l(d+\Delta d)+\frac{\Delta d.k}{d}\left(\sqrt{|\omega|}-1\right)\right)^2 - 2|\omega|\left(l(d+\Delta d)+\frac{\Delta d.k}{d}\left(\sqrt{|\omega|}-1\right)\right)\frac{\Delta d.k}{2d\sqrt{|\omega|}}}{\left(l(d+\Delta d)+\frac{\Delta d.k}{d}\left(\sqrt{|\omega|}-1\right)\right)^4} =$$

$$\frac{\left(1-\frac{\Delta d.k\sqrt{|\omega|}}{d\left(l(d+\Delta d)+\frac{\Delta d.k}{d}\left(\sqrt{|\omega|}-1\right)\right)}\right)}{\left(l(d+\Delta d)+\frac{\Delta d.k}{d}\left(\sqrt{|\omega|}-1\right)\right)^2} \tag{9.24}$$

It is obvious that $\left(l(d+\Delta d)+\frac{\Delta d.k}{d}\left(\sqrt{|\omega|}-1\right)\right) > 0$ for $l > 0$, hence,

$$\frac{\partial ESD_{UVDM(d,\Delta d,k,l,|\omega|)}}{\partial|\omega|} = 0 \Rightarrow 1 - \frac{\Delta d.k\sqrt{|\omega|}}{d\left(l(d+\Delta d)+\frac{\Delta d.k}{d}\left(\sqrt{|\omega|}-1\right)\right)} = 0 \Rightarrow \Delta d.k\sqrt{|\omega|} = d\left(l(d+\Delta d)+\frac{\Delta d.k}{d}\left(\sqrt{|\omega|}-1\right)\right) \Rightarrow \Delta d.k\sqrt{|\omega|} = ld^2 + ld\Delta d + \Delta d.k\sqrt{|\omega|} - \Delta d.k \Rightarrow ld^2 + ld\Delta d - \Delta d.k = 0 \Rightarrow k = \frac{ld^2+ld\Delta d}{\Delta d}$$

And partial derivation based on k is:

$$\frac{\partial ESD_{UVDM(d,\Delta d,k,l,|\omega|)}}{\partial k} = \frac{-2\frac{\Delta d}{d}|\omega|\left(\sqrt{|\omega|}-1\right)\left(l(d+\Delta d)+\frac{\Delta d.k}{d}\left(\sqrt{|\omega|}-1\right)\right)}{\left(l(d+\Delta d)+\frac{\Delta d.k}{d}\left(\sqrt{|\omega|}-1\right)\right)^4} =$$

$$\frac{-2\Delta d|\omega|\left(\sqrt{|\omega|}-1\right)}{d\left(l(d+\Delta d)+\frac{\Delta d.k}{d}\left(\sqrt{|\omega|}-1\right)\right)^3} \tag{9.25}$$

It is obvious that for any given $d > 0$, $\Delta d \geq 0$, $l > 0$, $k > 0$ and $|\omega| \geq 1$, $d\left(l(d + \Delta d) + \frac{\Delta d.k}{d}\left(\sqrt{|\omega|} - 1\right)\right)^3 > 0$, and $-2\Delta d|\omega|\left(\sqrt{|\omega|} - 1\right) \leq 0$ hence, $\frac{\partial ESD_{UVDM(d,\Delta d,k,l,|\omega|)}}{\partial k} \leq 0$, where equality occurs for $|\omega| = 1$.

From (9.24) and (9.25) we can calculate (9.26):

$$\nabla\left(\mathrm{ESD}_{\mathrm{UVDM}(d,\Delta d,k,l,|\omega|)}\right) = \left(\frac{\partial \mathrm{ESD}_{\mathrm{UVDM}(d,\Delta d,k,l,|\omega|)}}{\partial|\omega|}, \frac{\partial \mathrm{ESD}_{\mathrm{UVDM}(d,\Delta d,k,l,|\omega|)}}{\partial k}\right) =$$

$$\left(\frac{\left(1 - \frac{\Delta d.k\sqrt{|\omega|}}{d\left(l(d+\Delta d) + \frac{\Delta d.k}{d}\left(\sqrt{|\omega|}-1\right)\right)}\right)}{\left(l(d+\Delta d) + \frac{\Delta d.k}{d}\left(\sqrt{|\omega|}-1\right)\right)^2}, \frac{-2\Delta d|\omega|\left(\sqrt{|\omega|}-1\right)}{d\left(l(d+\Delta d) + \frac{\Delta d.k}{d}\left(\sqrt{|\omega|}-1\right)\right)^3}\right) \tag{9.26}$$

The ESD Lagrangean equation can be written as:

$$\Lambda\,(k, |\omega|, \lambda) = C(k, |\omega|) + \lambda\left(ESD_{UVDM(d,\Delta d,\mathrm{k,l},|\omega|)} - T\right) =$$

$$\frac{C_k}{k^2} + C_\omega|\omega| + \lambda\left(\frac{|\omega|}{\left(l(d+\Delta\mathrm{d}) + \frac{\Delta\mathrm{d.k}}{\mathrm{d}}\left(\sqrt{|\omega|}-1\right)\right)^2} - T\right) \tag{9.27}$$

The optimum k and $|\omega|$ should satisfy (9.28),

$$\nabla_{k,|\omega|,\lambda}\Lambda\,(k, |\omega|, \lambda) = 0 \Rightarrow \begin{cases} \frac{\partial \Lambda\,(k,|\omega|,\lambda)}{\partial k} = 0 \\ \frac{\partial \Lambda\,(k,|\omega|,\lambda)}{\partial|\omega|} = 0 \\ \frac{\partial \Lambda\,(k,|\omega|,\lambda)}{\partial\lambda} = 0 \end{cases} \tag{9.28}$$

By expanding the equations in (9.28) we have:

$$\frac{\partial \Lambda\,(k,|\omega|,\lambda)}{\partial k} = \frac{-2C_k}{k^3} + \frac{-2\lambda\Delta d|\omega|\left(\sqrt{|\omega|}-1\right)}{d\left(l(d+\Delta d) + \frac{\Delta d.k}{d}\left(\sqrt{|\omega|}-1\right)\right)^3} = 0 \tag{9.29}$$

$$\frac{\partial \Lambda\,(k,|\omega|,\lambda)}{\partial|\omega|} = C_\omega + \frac{\lambda\left(1 - \frac{\Delta d.k\sqrt{|\omega|}}{d\left(l(d+\Delta d) + \frac{\Delta d.k}{d}\left(\sqrt{|\omega|}-1\right)\right)}\right)}{\left(l(d+\Delta d) + \frac{\Delta d.k}{d}\left(\sqrt{|\omega|}-1\right)\right)^2} = 0 \tag{9.30}$$

$$\frac{\partial \Lambda\,(k,|\omega|,\lambda)}{\partial \lambda} = \frac{|\omega|}{\left(l(d+\Delta d)+\frac{\Delta d.k}{d}\left(\sqrt{|\omega|}-1\right)\right)^2} - T = 0 \tag{9.31}$$

By rewriting (9.31) based on k:

$$\frac{|\omega|}{\left(l(d+\Delta d)+\frac{\Delta d.k}{d}\left(\sqrt{|\omega|}-1\right)\right)^2} = T \Rightarrow k = \frac{\sqrt{|\omega|}d - ld\sqrt{T}(d+\Delta d)}{\Delta d\sqrt{T}\left(\sqrt{|\omega|}-1\right)} \tag{9.32}$$

From (9.29) and (9.30) we have:

$$\lambda = \frac{-C_k d\left(l(d+\Delta d)+\frac{\Delta d.k}{d}\left(\sqrt{|\omega|}-1\right)\right)^3}{k^3 \Delta d|\omega|\left(\sqrt{|\omega|}-1\right)} \text{ and,}$$

$$\lambda = \frac{-C_\omega\left(l(d+\Delta d)+\frac{\Delta d.k}{d}\left(\sqrt{|\omega|}-1\right)\right)^2}{\left(1-\frac{\Delta d.k\sqrt{|\omega|}}{d\left(l(d+\Delta d)+\frac{\Delta d.k}{d}\left(\sqrt{|\omega|}-1\right)\right)}\right)}$$

$$\xrightarrow{by\ removing\ \lambda} \frac{-C_k d\left(l(d+\Delta d)+\frac{\Delta d.k}{d}\left(\sqrt{|\omega|}-1\right)\right)}{k^3 \Delta d|\omega|\left(\sqrt{|\omega|}-1\right)} = \frac{-C_\omega}{\left(1-\frac{\Delta d.k\sqrt{|\omega|}}{d\left(l(d+\Delta d)+\frac{\Delta d.k}{d}\left(\sqrt{|\omega|}-1\right)\right)}\right)} \Rightarrow$$

$$C_k d\left(l(d+\Delta d)+\frac{\Delta d.k}{d}\left(\sqrt{|\omega|}-1\right)\right)\left(1-\frac{\Delta d.k\sqrt{|\omega|}}{d\left(l(d+\Delta d)+\frac{\Delta d.k}{d}\left(\sqrt{|\omega|}-1\right)\right)}\right) = C_\omega k^3 \Delta d|\omega|\left(\sqrt{|\omega|}-1\right) \Rightarrow$$

$$C_k d\left(l(d+\Delta d)+\frac{\Delta d.k}{d}\left(\sqrt{|\omega|}-1\right)\right) - C_k \Delta d.k\sqrt{|\omega|} - C_\omega k^3 \Delta d|\omega|\left(\sqrt{|\omega|}-1\right) = 0$$

$$\xrightarrow{\text{Rewriting based on } k} C_k dl(d+\Delta d) + C_k \Delta d.k\sqrt{|\omega|} - C_k \Delta d.k - C_k \Delta d.k\sqrt{|\omega|} -$$

$$C_\omega k^3 \Delta d|\omega|\sqrt{|\omega|} + C_\omega k^3 \Delta d|\omega| = 0 \Rightarrow$$

$$\left(C_\omega \Delta d|\omega|\left(1-\sqrt{|\omega|}\right)\right)k^3 - (C_k \Delta d)k + C_k dl(d+\Delta d) = 0 \tag{9.33}$$

Now by substituting k from (9.32) in (9.33) we have:

$$\left(C_\omega \Delta d|\omega|\left(1-\sqrt{|\omega|}\right)\right)\left(\frac{\sqrt{|\omega|}d-ld\sqrt{T}(d+\Delta\mathrm{d})}{\Delta\mathrm{d}\sqrt{T}\left(\sqrt{|\omega|}-1\right)}\right)^3-(C_k\Delta d)\left(\frac{\sqrt{|\omega|}d-ld\sqrt{T}(d+\Delta\mathrm{d})}{\Delta\mathrm{d}\sqrt{T}\left(\sqrt{|\omega|}-1\right)}\right)+C_k dl(d+\Delta d)=0\Rightarrow$$

$$\left(C_\omega \Delta d|\omega|-C_\omega \Delta d|\omega|\sqrt{|\omega|}\right)\left(\frac{\left(|\omega|\sqrt{|\omega|}d^3-l^3d^3T\sqrt{T}(d+\Delta\mathrm{d})^3-3|\omega|d^3l\sqrt{T}(d+\Delta\mathrm{d})+3l^2d^3T(d+\Delta\mathrm{d})^2\sqrt{|\omega|}\right)}{\left(\Delta\mathrm{d}\sqrt{T}\left(\sqrt{|\omega|}-1\right)\right)^3}\right)-$$

$$(C_k\Delta d)\left(\frac{\sqrt{|\omega|}d-ld\sqrt{T}(d+\Delta\mathrm{d})}{\Delta\mathrm{d}\sqrt{T}\left(\sqrt{|\omega|}-1\right)}\right)+C_k dl(d+\Delta d)=0\Rightarrow$$

$$\left(\frac{\left(C_\omega\Delta d|\omega|^2\sqrt{|\omega|}d^3-C_\omega\Delta d|\omega|l^3d^3T\sqrt{T}(d+\Delta\mathrm{d})^3-3C_\omega\Delta d|\omega|^2d^3l\sqrt{T}(d+\Delta\mathrm{d})+3C_\omega\Delta dl^2d^3T(d+\Delta\mathrm{d})^2|\omega|\sqrt{|\omega|}\right)}{\left(\Delta\mathrm{d}\sqrt{T}\left(\sqrt{|\omega|}-1\right)\right)^3}\right)+$$

$$\left(\frac{\left(-C_\omega\Delta d|\omega|^3d^3+C_\omega\Delta d|\omega|\sqrt{|\omega|}l^3d^3T\sqrt{T}(d+\Delta\mathrm{d})^3+3C_\omega\Delta d|\omega|^2\sqrt{|\omega|}d^3l\sqrt{T}(d+\Delta\mathrm{d})-3C_\omega\Delta d|\omega|^2l^2d^3T(d+\Delta\mathrm{d})^2\right)}{\left(\Delta\mathrm{d}\sqrt{T}\left(\sqrt{|\omega|}-1\right)\right)^3}\right)+$$

$$\left(\frac{-C_k\Delta d\sqrt{|\omega|}d+C_k\Delta dld\sqrt{T}(d+\Delta\mathrm{d})}{\Delta\mathrm{d}\sqrt{T}\left(\sqrt{|\omega|}-1\right)}\right)+C_k dl(d+\Delta d)=0\Rightarrow$$

$$\left(\frac{\left(-C_\omega d^3\Delta d|\omega|^3+C_\omega d^3\Delta d|\omega|^2\sqrt{|\omega|}+3C_\omega d^3\Delta dl\sqrt{T}(d+\Delta\mathrm{d})|\omega|^2\sqrt{|\omega|}-3C_\omega d^3\Delta dl\sqrt{T}(d+\Delta\mathrm{d})|\omega|^2-3C_\omega l^2d^3T(d+\Delta\mathrm{d})^2\Delta d|\omega|^2+3C_\omega\Delta dl^2d^3T(d+\Delta\mathrm{d})^2|\omega|\sqrt{|\omega|}+C_\omega\Delta dl^3d^3T\sqrt{T}(d+\Delta\mathrm{d})^3|\omega|\sqrt{|\omega|}-C_\omega\Delta dl^3d^3T\sqrt{T}(d+\Delta\mathrm{d})^3|\omega|\right)}{\left(\Delta\mathrm{d}\sqrt{T}\left(\sqrt{|\omega|}-1\right)\right)^3}\right)+$$

$$\left(\frac{-C_k\Delta d\sqrt{|\omega|}d+C_k\Delta dld\sqrt{T}(d+\Delta\mathrm{d})}{\Delta\mathrm{d}\sqrt{T}\left(\sqrt{|\omega|}-1\right)}\right)+C_k dl(d+\Delta d)=0$$

$$\xRightarrow{\Delta\mathrm{d}\sqrt{T}\left(\sqrt{|\omega|}-1\right)\neq 0}\left(\left(-C_\omega d^3\Delta d|\omega|^3+C_\omega d^3\Delta d|\omega|^2\sqrt{|\omega|}+3C_\omega d^3\Delta dl\sqrt{T}(d+\Delta\mathrm{d})|\omega|^2\sqrt{|\omega|}-3C_\omega d^3\Delta dl\sqrt{T}(d+\Delta\mathrm{d})|\omega|^2-3C_\omega l^2d^3T(d+\Delta\mathrm{d})^2\Delta d|\omega|^2+3C_\omega\Delta dl^2d^3T(d+\Delta\mathrm{d})^2|\omega|\sqrt{|\omega|}+C_\omega\Delta dl^3d^3T\sqrt{T}(d+\Delta\mathrm{d})^3|\omega|\sqrt{|\omega|}-C_\omega\Delta dl^3d^3T\sqrt{T}(d+\Delta\mathrm{d})^3|\omega|\right)\right)+\left(\Delta\mathrm{d}\sqrt{T}\left(\sqrt{|\omega|}-1\right)\right)^2\left(-C_k\Delta d\sqrt{|\omega|}d+C_k\Delta dld\sqrt{T}(d+\Delta\mathrm{d})\right)+\left(\Delta\mathrm{d}\sqrt{T}\left(\sqrt{|\omega|}-1\right)\right)^3\left(C_k dl(d+\Delta d)\right)=0\Rightarrow$$

$$\left(\left(-C_\omega d^3\Delta d|\omega|^3+C_\omega d^3\Delta d|\omega|^2\sqrt{|\omega|}+3C_\omega d^3\Delta dl\sqrt{T}(d+\Delta\mathrm{d})|\omega|^2\sqrt{|\omega|}-3C_\omega d^3\Delta dl\sqrt{T}(d+\Delta\mathrm{d})|\omega|^2-3C_\omega l^2d^3T(d+\Delta\mathrm{d})^2\Delta d|\omega|^2+\right.\right.$$

$$3C_\omega \Delta d l^2 d^3 T(d+\Delta\mathrm{d})^2|\omega|\sqrt{|\omega|} + C_\omega \Delta d l^3 d^3 T\sqrt{T}(d+\Delta\mathrm{d})^3|\omega|\sqrt{|\omega|} - C_\omega \Delta d l^3 d^3 T\sqrt{T}(d+\Delta\mathrm{d})^3|\omega|\Big)\Big) + \Big(\Delta\mathrm{d}^2 T\left(|\omega| + 1 - 2\sqrt{|\omega|}\right)\Big)\Big(-C_k \Delta d\sqrt{|\omega|}d + C_k \Delta d l d\sqrt{T}(d+\Delta\mathrm{d})\Big) + \Big(\Delta\mathrm{d}^3\mathrm{T}\sqrt{T}\left(|\omega|\sqrt{|\omega|} - 1 - 3|\omega| + 3\sqrt{|\omega|}\right)\Big)\left(C_k dl(d+\Delta d)\right) = 0 \Rightarrow$$

$$\Big(\Big(-C_\omega d^3 \Delta d|\omega|^3 + C_\omega d^3 \Delta d|\omega|^2\sqrt{|\omega|} + 3C_\omega d^3 \Delta d l\sqrt{T}(d+\Delta\mathrm{d})|\omega|^2\sqrt{|\omega|} - 3C_\omega d^3 \Delta d l\sqrt{T}(d+\Delta\mathrm{d})|\omega|^2 - 3C_\omega l^2 d^3 T(d+\Delta\mathrm{d})^2 \Delta d|\omega|^2 + 3C_\omega \Delta d l^2 d^3 T(d+\Delta\mathrm{d})^2|\omega|\sqrt{|\omega|} + C_\omega \Delta d l^3 d^3 T\sqrt{T}(d+\Delta\mathrm{d})^3|\omega|\sqrt{|\omega|} - C_\omega \Delta d l^3 d^3 T\sqrt{T}(d+\Delta\mathrm{d})^3|\omega|\Big)\Big) + \Big(\Delta\mathrm{d}^2 T|\omega| + \Delta\mathrm{d}^2 T - 2\Delta\mathrm{d}^2 T\sqrt{|\omega|}\Big)\Big(-C_k \Delta d\sqrt{|\omega|}d + C_k \Delta d l d\sqrt{T}(d+\Delta\mathrm{d})\Big) + \Big(\Big(\Delta\mathrm{d}^3\mathrm{T}\sqrt{T}|\omega|\sqrt{|\omega|} - \Delta\mathrm{d}^3\mathrm{T}\sqrt{T} - 3\Delta\mathrm{d}^3\mathrm{T}\sqrt{T}|\omega| + 3\Delta\mathrm{d}^3\mathrm{T}\sqrt{T}\sqrt{|\omega|}\Big)\Big)\left(C_k dl(d+\Delta d)\right) = 0 \Rightarrow$$

$$\Big(\Big(-C_\omega d^3 \Delta d|\omega|^3 + C_\omega d^3 \Delta d|\omega|^2\sqrt{|\omega|} + 3C_\omega d^3 \Delta d l\sqrt{T}(d+\Delta\mathrm{d})|\omega|^2\sqrt{|\omega|} - 3C_\omega d^3 \Delta d l\sqrt{T}(d+\Delta\mathrm{d})|\omega|^2 - 3C_\omega l^2 d^3 T(d+\Delta\mathrm{d})^2 \Delta d|\omega|^2 + 3C_\omega \Delta d l^2 d^3 T(d+\Delta\mathrm{d})^2|\omega|\sqrt{|\omega|} + C_\omega \Delta d l^3 d^3 T\sqrt{T}(d+\Delta\mathrm{d})^3|\omega|\sqrt{|\omega|} - C_\omega \Delta d l^3 d^3 T\sqrt{T}(d+\Delta\mathrm{d})^3|\omega|\Big)\Big) + \Big(-C_k d\Delta\mathrm{d}^3 T|\omega|\sqrt{|\omega|} + C_k ld(d+\Delta\mathrm{d})\Delta\mathrm{d}^3 T\sqrt{T}|\omega|\Big) + \Big(-d\Delta\mathrm{d}^3 T C_k\sqrt{|\omega|} + \Delta\mathrm{d}^3 T\sqrt{T}C_k ld(d+\Delta\mathrm{d})\Big) + \Big(2d\Delta\mathrm{d}^3 T C_k|\omega| - 2\Delta\mathrm{d}^3 T\sqrt{T}C_k ld(d+\Delta\mathrm{d})\sqrt{|\omega|}\Big) + \Big(\Big(\Delta\mathrm{d}^3\mathrm{T}\sqrt{T}C_k dl(d+\Delta d)|\omega|\sqrt{|\omega|} - \Delta\mathrm{d}^3\mathrm{T}\sqrt{T}C_k dl(d+\Delta d) - 3\Delta\mathrm{d}^3\mathrm{T}\sqrt{T}C_k dl(d+\Delta d)|\omega| + 3\Delta\mathrm{d}^3\mathrm{T}\sqrt{T}C_k dl(d+\Delta d)\sqrt{|\omega|}\Big)\Big) = 0 \Rightarrow$$

$$-C_\omega d^3 \Delta d|\omega|^3 + C_\omega d^3 \Delta d|\omega|^2\sqrt{|\omega|} + 3C_\omega d^3 \Delta d l\sqrt{T}(d+\Delta\mathrm{d})|\omega|^2\sqrt{|\omega|} - 3C_\omega d^3 \Delta d l\sqrt{T}(d+\Delta\mathrm{d})|\omega|^2 - 3C_\omega l^2 d^3 T(d+\Delta\mathrm{d})^2 \Delta d|\omega|^2 + 3C_\omega \Delta d l^2 d^3 T(d+\Delta\mathrm{d})^2|\omega|\sqrt{|\omega|} - C_k d\Delta\mathrm{d}^3 T|\omega|\sqrt{|\omega|} + C_\omega \Delta d l^3 d^3 T\sqrt{T}(d+\Delta\mathrm{d})^3|\omega|\sqrt{|\omega|} + \Delta\mathrm{d}^3\mathrm{T}\sqrt{T}C_k ld(d+\Delta d)|\omega|\sqrt{|\omega|} - C_\omega \Delta d l^3 d^3 T\sqrt{T}(d+\Delta\mathrm{d})^3|\omega| + C_k ld(d+\Delta\mathrm{d})\Delta\mathrm{d}^3 T\sqrt{T}|\omega| + 2d\Delta\mathrm{d}^3 T C_k|\omega| - 3\Delta\mathrm{d}^3\mathrm{T}\sqrt{T}C_k ld(d+\Delta d)|\omega| - d\Delta\mathrm{d}^3 T C_k\sqrt{|\omega|} - 2\Delta\mathrm{d}^3 T\sqrt{T}C_k ld(d+\Delta\mathrm{d})\sqrt{|\omega|} + 3\Delta\mathrm{d}^3\mathrm{T}\sqrt{T}C_k dl(d+\Delta d)\sqrt{|\omega|} + \Delta\mathrm{d}^3 T\sqrt{T}C_k ld(d+\Delta\mathrm{d}) - \Delta\mathrm{d}^3\mathrm{T}\sqrt{T}C_k dl(d+\Delta d) = 0 \Rightarrow$$

$$-C_\omega d^3 \Delta d|\omega|^3 + C_\omega d^3 \Delta d|\omega|^2\sqrt{|\omega|} + 3C_\omega d^3 \Delta d l\sqrt{T}(d+\Delta d)|\omega|^2\sqrt{|\omega|} - 3C_\omega d^3 \Delta d l\sqrt{T}(d+\Delta d)|\omega|^2 - 3C_\omega l^2 d^3 T(d+\Delta d)^2 \Delta d|\omega|^2 + 3C_\omega \Delta d l^2 d^3 T(d+\Delta d)^2|\omega|\sqrt{|\omega|} - C_k d\Delta d^3 T|\omega|\sqrt{|\omega|} + C_\omega \Delta d l^3 d^3 T\sqrt{T}(d+\Delta d)^3|\omega|\sqrt{|\omega|} + \Delta d^3 T\sqrt{T}C_k ld(d+\Delta d)|\omega|\sqrt{|\omega|} - C_\omega \Delta d l^3 d^3 T\sqrt{T}(d+\Delta d)^3|\omega| - 2C_k ld(d+\Delta d)\Delta d^3 T\sqrt{T}|\omega| + 2d\Delta d^3 T C_k|\omega| - d\Delta d^3 T C_k\sqrt{|\omega|} + \Delta d^3 T\sqrt{T}C_k ld(d+\Delta d)\sqrt{|\omega|} = 0 \xrightarrow{\sqrt{|\omega|}\neq 0}$$

$$(-C_\omega d^3 \Delta d)|\omega|^2\sqrt{|\omega|} + \left(3C_\omega d^3 \Delta d l\sqrt{T}(d+\Delta d) + C_\omega d^3 \Delta d\right)|\omega|^2 + \left(-3C_\omega d^3 \Delta d l\sqrt{T}(d+\Delta d) - 3C_\omega l^2 d^3 T(d+\Delta d)^2 \Delta d\right)|\omega|\sqrt{|\omega|} + \left(3C_\omega \Delta d l^2 d^3 T(d+\Delta d)^2 - C_k d\Delta d^3 T + C_\omega \Delta d l^3 d^3 T\sqrt{T}(d+\Delta d)^3 + \Delta d^3 T\sqrt{T}C_k ld(d+\Delta d)\right)|\omega| + \left(-C_\omega \Delta d l^3 d^3 T\sqrt{T}(d+\Delta d)^3 - 2C_k ld(d+\Delta d)\Delta d^3 T\sqrt{T} + 2d\Delta d^3 T C_k\right)\sqrt{|\omega|} + \left(-d\Delta d^3 T C_k + \Delta d^3 T\sqrt{T}C_k ld(d+\Delta d)\right) = 0 \quad (9.34)$$

Let's introduce an auxiliary variable $X = \sqrt{|\omega|}$, $|\omega|$ can be computed by rewriting (9.34) based on X as shown as (9.35) which is a quintic equation (polynomial equation of order 5). Analytical methods for solving a quintic equation by using radicals are available under specific criteria, such as Arthur Cayley method [105]. Alternatively well-known Newton-Raphson numerical method [106] can be used. Note that only real roots $|\omega| > 1$ are acceptable. After calculating $|\omega|$, k can be calculated from (9.32).

$$\alpha_5 X^5 + \alpha_4 X^4 + \alpha_3 X^3 + \alpha_2 X^2 + \alpha_1 X + \alpha_0 = 0 \quad (9.35)$$

where,

$$\begin{cases} \alpha_5 = (-C_\omega d^3 \Delta d) \\ \alpha_4 = \left(3C_\omega l d^3 \Delta d\sqrt{T}(d+\Delta d) + C_\omega d^3 \Delta d\right) \\ \alpha_3 = \left(-3C_\omega l d^3 \Delta d\sqrt{T}(d+\Delta d) - 3C_\omega l^2 d^3 \Delta d T(d+\Delta d)^2\right) \\ \alpha_2 = \left(3C_\omega l^2 d^3 \Delta d T(d+\Delta d)^2 - C_k d\Delta d^3 T + C_\omega l^3 d^3 \Delta d T\sqrt{T}(d+\Delta d)^3 + C_k ld\Delta d^3 T\sqrt{T}(d+\Delta d)\right) \\ \alpha_1 = \left(-C_\omega l^3 d^3 \Delta d T\sqrt{T}(d+\Delta d)^3 - 2C_k ld\Delta d^3 T\sqrt{T}(d+\Delta d) + 2C_k d\Delta d^3 T\right) \\ \alpha_0 = \left(-C_k d\Delta d^3 T + C_k ld\Delta d^3 T\sqrt{T}(d+\Delta d)\right) \end{cases}$$

9.5 Appendix V: The Quantitative Analysis for Comparison and Evaluation of LF based FVV Systems

9.5.1 Summary

Although several acquisition models and rendering algorithms have been suggested for LF based FVV systems but the lack of appropriate datasets with known ground truth has prevented a comparison and evaluation study of these models and algorithms. In most of the reported studies, the proposed method is applied to several test cases for validation and as a result, just a subjective visualized output is given. To overcome this problem and to validate the theory of ESD, this thesis has widely used a new quantitative approach for LF-based FVV systems to validate different theoretical expectations from ESD theory reported in previous chapters. This short chapter briefly presents this novel quantitative approach for comparison and evaluation of LF-based FVV systems. The core of the proposed methodology is a simulation model and a 3D engine. The platform produces the reference images and ground truth data for a given 3D model. Subsequently, data are injected to a comparison engine to compare synthesized images from light field engine with original images from simulation, generating objective experimental results for evaluation. The methodology is highly flexible and efficient to automatically generate different datasets and objectively compare and analyze any subset of rendering methods or acquisition models for any given experiment design scheme. Overall, it is shown that the proposed quantitative methodology could be used for objective evaluation and comparison of LF-based FVV systems and to validate the ESD theory.

9.5.2 Methodology and Simulation Model

The main idea behind the proposed methodology is to develop a computer simulation model employing a 3D engine. The simulator roles could be categorized as:

1) Any LF-based FVV system has its own acquisition architecture and configurations employing different number of cameras, different topologies/architectures, different camera configurations and different parameterizations such as basic 2 slabs/planes (2PP), two-sphere parameterization (2SP) and sphere-plane parameterization (SPP). This information is given to the simulator and the simulator generates all the reference images/videos for the given LF acquisition component.

2) The experimental scheme gives a random validation or test set of virtual/observation cameras distributed in the user navigability space. These observation images for different viewpoints are generated by the simulator. These images are considered as ground truth data for objective evaluation.

3) These are different LF rendering algorithms. Given the rendering algorithm, the LF rendering engine generates the synthesized/rendered images for the same viewpoints of test set, from the given reference images.

4) Image comparison engine compares each pair of ground truth and rendered images based on a desired metric such as PSNR to quantitatively evaluates the quality of rendering. The average of PSNR (or other metrics) for all the images in test set is an objective assessment of that LF-based FVV system for that test set.

The simulation model is able to generate variety of datasets and ground truth data for any given experimental configuration and FVV system, automatically and efficiently. These datasets are statistically reliable. The data analysis also could be automatically done by comparison engine. The output of the system is well-designed charts including any statistical analysis required by experiment scheme. Figure 9.15 illustrates the architecture of the system.

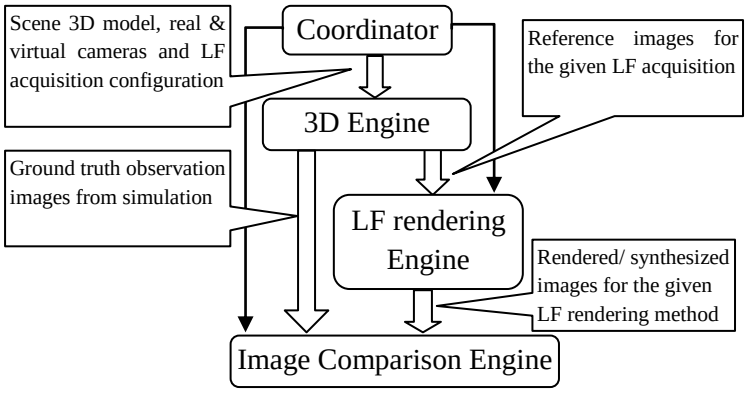


Figure 9.15. System architecture and simulation data flow

Coordinator interacts with user, reads input files, calculates the positions and orientations of reference cameras for a given LF acquisition, calculates the positions and orientations of virtual/observation cameras (test/validation set) and provides the scene 3D model, reference cameras and observation cameras configurations to 3D engine and supply LF rendering engine by rendering algorithm, rendering parameters and reference cameras configurations.

3D engine renders the scene 3D model, automatically generates reference and virtual/observation cameras in 3D model and generates image files corresponding to the cameras. In the developed system, open source Blender was utilized for 3D engine and a Python script for simulation procedure was implemented. All other subsystems are implemented in C to improve efficiency.

LF rendering engine creates a ray space for the given acquisition and then synthesize/render output images for each given observation point by employing the given rendering algorithm and parameters.

Image comparison engine receives two sets of images, a ground truth observation image set from the 3D engine and a rendered observation image set from the LF rendering engine. It employs an image comparison method to calculate the images dissimilarity related to

LF rendering error for that given observation point and scene geometry. The Mean of this error indicates the LF error for the given scene. In this thesis the Peak Signal to Noise Ratio (PSNR) is used in to calculate the error in *db*. This module provides coordinator with all error data and subsequently, coordinator provide the user with all kind of data analysis and charts required for that given experiment scheme.

Interestingly, since 3D models are used to represent the scene, a full depth map is available to use in the rendering process. Controlled amount of depth map error can then be introduced to study how the rendering would be impacted when the depth map is noisy or inaccurate. Figure 9.16 demonstrates a snapshot of the camera grid for acquisition and some random virtual cameras for rendering and generation of ground truth.

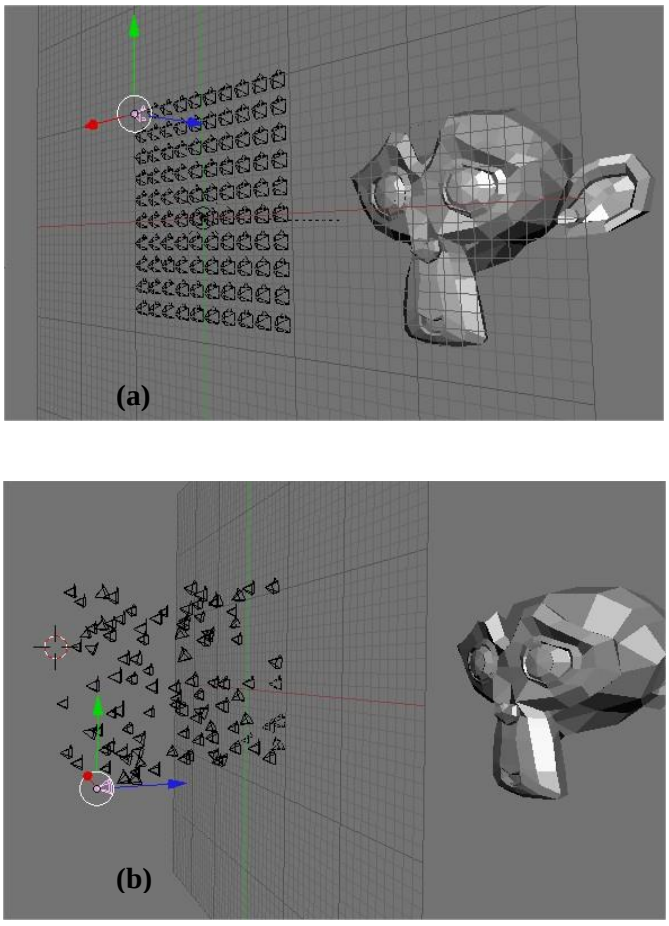


Figure 9.16. a) A simulated regular camera grid; b) Virtual viewpoints.

Figure 9.17 demonstrates the rendering outputs for three LF rendering methods for a given 3D object and the same viewpoint.

UVST UV-D UV-DM

Figure 9.17. Sample rendering output for three LF rendering methods

9.5.3 Discussion and Conclusion

A quantitative methodology for LF-based FVV system evaluation and comparison is proposed in this appendix. This methodology has been widely used throughout the thesis to validate the ESD theory and different evaluation and optimization models proposed in different chapters.

References

[1] M. Tanimoto, *et al.*, "Free-Viewpoint TV," *IEEE Signal Processing Magazine*, vol. 28, pp. 67-76, 2011.

[2] M. Tanimoto, "FTV: Free-viewpoint Television," *Signal Processing: Image Communication*, vol. 27, pp. 555-570, 2012.

[3] T. Fujii and M. Tanimoto, "Free viewpoint TV system based on ray-space representation," in *ITCom 2002: The Convergence of Information Technologies and Communications*, 2002, pp. 175-189.

[4] M. Hussain, "Free Viewpoint TV," Linköping, 2010.

[5] M. Tanimoto, "Overview of free viewpoint television," *Signal Processing: Image Communication*, vol. 21, pp. 454-461, 2006.

[6] E. H. Adelson and J. Bergen, "The plenoptic function and the elements of early vision," *Computational Models of Visual Processing*, pp. 3-20, 1991.

[7] M. Levoy and P. Hanrahan, "Light field rendering," *Proc. SIGGRAPH (ACM Trans. Graphics)*, pp. 31-42, Aug. 1996.

[8] S. J. Gortler, *et al.*, "The lumigraph," *Proc. SIGGRAPH (ACM Trans. Graphics)*, pp. 43-54, Aug. 1996.

[9] K. Takahashi and T. Naemura, "Layered light-field rendering with focus measurement," *Signal Processing: Image Communication*, vol. 21, pp. 519-530, 2006.

[10] N. W. Daniel, *et al.*, "Surface light fields for 3D photography," presented at the 27th annual conference on Computer graphics and interactive techniques, 2000.

[11] Y. Jingyi, *et al.*, "Scam light field rendering," in *10th Pacific Conference on Computer Graphics and Applications, 2002*, pp. 137-144.

[12] H. Y. Shum, *et al.*, "Pop-up light field: An interactive image-based modeling and rendering system," *ACM Trans. Graphics*, vol. 23, pp. 143-162, Apr. 2004.

[13] W. Wen, *et al.*, "An efficient method for all-in-focused light field rendering," in *3rd IEEE International Conference on Computer Science and Information Technology (ICCSIT), 2010*, pp. 399-404.

[14] I. Aaron, *et al.*, "Dynamically reparameterized light fields," presented at 27th annual conference on Computer graphics and interactive techniques, 2000.

[15] L. Wenfeng, *et al.*, "Virtual View Specification and Synthesis for Free Viewpoint Television," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 19, pp. 533-546, 2009.

[16] H. Shidanshidi, *et al.*, "A Method for Calculating the Minimum Number of Cameras in a Light Field Based Free Viewpoint Video System," in *ICME*, 2013, pp.1-6.

[17] K. Takahashi, "Theoretical Analysis of View Interpolation With Inaccurate Depth Information," *IEEE Transactions on Image Processing*, vol. 21, pp. 718-732, 2012.

[18] J.X. Chai, *et al.*, "Plenoptic sampling," *Proc. SIGGRAPH (ACM Trans. Graphics)*, pp. 307-318, Jul. 2000.

[19] C. Zhang and T. Chen, "Spectral analysis for sampling image-based rendering data," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, pp. 1038-1050, 2003.

[20] C. Zhang and T. Chen, "Light field sampling," *Synthesis Lectures on Image, Video, and Multimedia Processing*, vol. 2, pp. 1-102, 2006.

[21] L. Zhouchen and S. Heung-Yeung, "A Geometric Analysis of Light Field Rendering," *International Journal of Computer Vision*, vol. 58, pp. 121-138, 2004.

[22] N. King-To, *et al.*, "A Multi-Camera Approach to Image-Based Rendering and 3-D/Multiview Display of Ancient Chinese Artifacts," *IEEE Transactions on Multimedia*, vol. 14, pp. 1631-1641, 2012.

[23] M. N. Do, *et al.*, "On the bandwidth of the plenoptic function," *IEEE Transactions on Image Processing*, vol. 21, pp. 708-717, 2012.
[24] K. Hansung, *et al.*, "Outdoor Dynamic 3-D Scene Reconstruction," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, pp. 1611-1622, 2012.
[25] S. X. Liu, *et al.*, "High quality virtual view synthesis based on corrected surface mapping and image fusion," *Electronics Letters*, vol. 45, pp. 30-32, 2009.
[26] E. Ekmekcioglu, *et al.*, "Content Adaptive Enhancement of Multi-View Depth Maps for Free Viewpoint Video," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, pp. 352-361, 2011.
[27] T. Scandarolli, *et al.*, "Attention-Weighted Rate Allocation in Free-Viewpoint Television," IEEE *Signal Processing Letters*, vol. 20, pp. 359-362, 2013.
[28] W. Qifei, *et al.*, "Free Viewpoint Video Coding With Rate-Distortion Analysis," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, pp. 875-889, 2012.
[29] H. Zhun and D. Qionghai, "A New Scalable Free Viewpoint Video Streaming System Over IP Network," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2007, pp. II-773-II-776.
[30] H. Shidanshidi, *et al.*, "Objective evaluation of light field rendering methods using effective sampling density," in *MMSP*, 2011, pp. 1-6.
[31] H. Shidanshidi, *et al.*, "Estimation of Signal Distortion using Effective Sampling Density for Light Field based Free Viewpoint Video," *IEEE Transactions on Multimedia*, received a qualified accept and is now under second revision.
[32] Z. Lin and H.-Y. Shum, "A Geometric Analysis of Light Field Rendering," *International Journal of Computer Vision*, vol. 58, pp. 121-138, 2004.
[33] H. Y. Shum, *et al.*, "Survey of image-based representations and compression techniques," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, pp. 1020-1037, Nov. 2003.
[34] C. Zhang and T. Chen, "A survey on image-based rendering—Representation, sampling and compression," *EURASIP Signal Processing: Image Commun.*, vol. 19, pp. 1-28, Jan. 2004.
[35] Y. Liu, *et al.*, "A point-cloud-based multiview stereo algorithm for free-viewpoint video," *IEEE Transactions on Visualization and Computer Graphics*, vol. 16, pp. 407-418, 2010.
[36] H. Y. Shum, *et al.*, "Image-Based Rendering," *Springer-Verlag*, 2006.
[37] L. McMillan and G. Bishop, "Plenoptic modeling: An image-based rendering system," *Proc. SIGGRAPH (ACM Trans. Graphics)*, pp. 39-46, 1995.
[38] R. Szeliski and H. Y. Shum, "Creating full view panoramic image mosaics and environment maps," *Proc. SIGGRAPH (ACM Trans. Graphics)*, pp. 251-258, 1997.
[39] H. Y. Shum and L. W. He, "Rendering with concentric mosaics," *Proc. SIGGRAPH (ACM Trans. Graphics)*, pp. 299-306, 1999.
[40] C. L. Zitnick, *et al.*, "High-quality video view interpolation using a layered representation," *Proc. SIGGRAPH (ACM Trans. Graphics)*, pp. 600-609, Aug 2004.
[41] S. C. Chan, *et al.*, "The plenoptic videos," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, pp. 1650-1659, Dec. 2005.
[42] T. Fujii, "A basic study on integrated 3-D visual communication," PhD thesis in Japanese, The University of Tokyo, 1994.
[43] T. Fujii, *et al.*, "Ray space coding for 3D visual communication," *Proc. Picture Coding Symp. '96*, pp. 447-451, Mar. 1996.
[44] W. Matusik, *et al.*, "Image-based visual hulls," in *27th annual conference on Computer graphics and interactive techniques*, 2000, pp. 369-374.
[45] S. C. Chan, *et al.*, "Image-Based Rendering and Synthesis [Technological advances and challenges]," *IEEE SIGNAL PROCESSING MAGAZINE*, pp. 22-33, Nov. 2007.
[46] J. Shade, *et al.*, "Layered depth images," *Proc. SIGGRAPH (ACM Trans. Graphics)*, pp. 231-242, Jul. 1998.

[47] C. Chang, *et al.*, "LDI tree: A hierarchical representation for image-based rendering," *Proc. SIGGRAPH (ACM Trans. Graphics)*, pp. 291-298, Aug. 1999.
[48] P. E. Debevec, *et al.*, "Efficient view-dependent image-based rendering with projective texture-mapping," *Proc. Eurographics Workshop on Rendering*, pp. 150-161, 1998.
[49] F. Gan, *et al.*, "An object-based approach to plenoptic videos," *Proc. IEEE Int. Symp. Circuits and Systems*, pp. 3435–3438, May 2005.
[50] C. Zhang and T. Chen, "Surface plenoptic function: a tool for the sampling analysis of image-based rendering," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2003, pp. IV-768-71 vol. 4.
[51] B. S. Wilburn, *et al.*, "Light field video camera," in *Electronic Imaging*, 2001, pp. 29-36.
[52] K. T. Ng, *et al.*, "The data compression and transmission aspects of panoramic videos," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, pp. 82-95, Jan 2005.
[53] J. Foote and D. Kimber, "FlyCam: Practical panoramic video and automatic camera control," in *EEE International Conference on Multimedia and Expo (ICME)*, 2000, pp. 1419-1422.
[54] U. Neumann, *et al.*, "Immersive panoramic video," in *eighth ACM international conference on Multimedia*, 2000, pp. 493-494.
[55] A. Lippman, "Movie-maps: An application of the optical videodisc to computer graphics," *ACM SIGGRAPH Computer Graphics*, vol. 14, pp. 32-42, 1980.
[56] G. Miller, *et al.*, "The virtual museum: Interactive 3d navigation of a multimedia database," *The Journal of Visualization and Computer Animation*, vol. 3, pp. 183-197, 1992.
[57] E. Camahort, *et al.*, "Uniformly sampled light fields," *Rendering Techniques*, vol. 98, pp. 117-130, 1998.
[58] E. Camahort and D. Fussell, "A geometric study of light field representations," *Department of Computer Science Univ. of Texas Technical Report TR9935*, 1999.
[59] R. Raskar and A. K. Agrawal, "4D light field cameras," ed: Google Patents, 2010.
[60] J. C. Yang, *et al.*, "A real-time distributed light field camera," in *13th Eurographics workshop on Rendering*, 2002, pp. 77-86.
[61] B. Wilburn, *et al.*, "The light field video camera," *in SPIE Electronic Imaging: Media Processors*, 2002, vol. 4674, pp. 29-36.
[62] T. Naemura, *et al.*, "Real-time video-based modeling and rendering of 3D scenes," *IEEE Comput. Graphics Applicat*, vol. 22, pp. 66-73, Mar.-Apr. 2002.
[63] J. C. Yang, *et al.*, "A real-time distributed light field camera," *Proc. Eurographics Workshop on Rendering*, pp. 77-86, 2002.
[64] C. Zhang and T. Chen, "Active rearranged capturing of image-based rendering scenes-Theory and practice," *IEEE Trans. Multimedia*, vol. 9, pp. 520-531, Apr. 2007.
[65] *The Stanford Multi-Camera Array*. Available: http://graphics.stanford.edu/projects/array/
[66] R. Arora and H. Parthasarathy, "Navigation using a spherical camera," in *19th International Conference on Pattern Recognition (ICPR)*, 2008, pp. 1-4.
[67] R. Koch, *et al.*, "Calibration of hand-held camera sequences for plenoptic modeling," in *the Seventh IEEE International Conference on Computer Vision (ICCV)*, 1999, pp. 585-591.
[68] A. Davis, *et al.*, "Unstructured light fields," in *Computer Graphics Forum*, 2012, pp. 305-314.
[69] C. Buehler, *et al.*, "Unstructured lumigraph rendering," in *28th annual conference on Computer graphics and interactive techniques*, 2001, pp. 425-432.
[70] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Transactions on Image Processing*, vol. 15, pp. 430-444, 2006.
[71] A. Pons, *et al.*, "Image quality metric based on multidimensional contrast perception models," *Displays*, vol. 20, pp. 93-110, 1999.
[72] S. Winkler, "A perceptual distortion metric for digital color images," in *ICIP*, 1998, pp. 399-403 vol. 3.

[73] A. M. Eskicioglu and P. S. Fisher, "Image quality measures and their performance," *IEEE Transactions on Communications*, vol. 43, pp. 2959-2965, 1995.
[74] T. Feng and H. Y. Shum, "An optical analysis of light field rendering," in *Fifth Asian Conference on Computer Vision*, 2000, pp. 394-399.
[75] A. Lumsdaine and T. Georgiev, "Full resolution lightfield rendering," *Indiana University and Adobe Systems, Tech. Rep*, 2008.
[76] J. Stewart, *et al.*, "A new reconstruction filter for undersampled light fields," in *14th Eurographics workshop on Rendering*, Leuven, Belgium, 2003.
[77] S. M. Seitz, *et al.*, "A comparison and evaluation of multi-view stereo reconstruction algorithms," in *CVPR*, 2006, pp. 519-528.
[78] J. Kilner, *et al.*, "Objective quality assessment in free-viewpoint video production," *Image Commun.*, vol. 24, pp. 3-16, 2009.
[79] T. Brandão and P. Queluz, "Towards objective metrics for blind assessment of images quality," in *IEEE International Conference on Image Processing (ICIP)*, 2006, pp. 2933-2936.
[80] K. Seshadrinathan and A. C. Bovik, "A structural similarity metric for video based on motion models," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2007, pp. I-869-I-872.
[81] S. Winkler, "Video quality and beyond," in *European Signal Processing Conference*, 2007, pp. 3-7.
[82] Z. Wang, *et al.*, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, pp. 600-612, 2004.
[83] İ. Avcıbaş, *et al.*, "Statistical evaluation of image quality measures," *Journal of Electronic imaging*, vol. 11, p. 206, 2002.
[84] E. Bosc, *et al.*, "Towards a New Quality Metric for 3-D Synthesized View Assessment," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, pp. 1332-1343, 2011.
[85] E. Bosc, *et al.*, "Can 3D synthesized views be reliably assessed through usual subjective and objective evaluation protocols?," in *18th IEEE International Conference on Image Processing (ICIP)*, 2011, pp. 2597-2600.
[86] B. Goldlücke, *et al.*, "Hardware-accelerated dynamic light field rendering," *Proc. VMV'2002,* pp. 455-462, 2002.
[87] Y. Y. Chuang, *et al.*, "A Bayesian approach to digital matting," *Proc. IEEE Conf. CVPR*, vol. 5, pp. 264-271, 2001.
[88] S. C. Chan, *et al.*, " An Object-Based Approach to Image/Video-Based Synthesis and Processing for 3-D and Multiview Televisions," *IEEE Transactions on Circuits and Systems for Video Technology, vol.19, no.6, pp.821,831, June 2009*
[89] P. Sloan, *et al.*, "Precomputed radiance transfer for real-time rendering in dynamic, low-frequency lighting environment," *Proc. SIGGRAPH (ACM Trans. Graphics)*, pp. 527-536, 2002.
[90] K. Zhou, *et al.*, "Precomputed shadow fields for dynamic scenes," *Proc. SIGGRAPH (ACM Trans. Graphics)*, pp. 1196-1201, 2005.
[91] K. Gröchenig, "A discrete theory of irregular sampling," *Linear Algebra and its Applications*, vol. 193, pp. 129-150, 1993.
[92] H. G. Feichtinger, *et al.*, "Efficient numerical methods in non-uniform sampling theory," *Numerische Mathematik*, vol. 69, pp. 423-440, 1995/02/01 1995.
[93] H. Boche and U. J. Monich, "Non-uniform sampling - signal and system representation," in *International Symposium on Information Theory and Its Applications ISITA*, 2008, pp. 1-6.
[94] H. Shidanshidi, *et al.*, "Non-uniform Sampling of Plenoptic Signal based on the Scene Complexity Variations for a Free Viewpoint Video System," in *ICIP*, 2013, pp. 3147 - 3151
[95] H. Shidanshidi, *et al.*, "Optimization of Acquisition and Rendering Subsystems Using Effective Sampling Density for Light Field based Free Viewpoint Video," *IEEE Transactions on Multimedia*, under review.

[96] H. Shidanshidi, *et al.*, "A quantitative approach for comparison and evaluation of light field rendering techniques," in *ICME*, 2011, pp. 1-4.
[97] S. Schwarz, *et al.*, "Depth Sensing for 3DTV: A Survey," *IEEE MultiMedia*, vol. 20, pp. 10-17, 2013.
[98] K. Khoshelham and S. O. Elberink, "Accuracy and resolution of kinect depth data for indoor mapping applications," *Sensors*, vol. 12, pp. 1437-1454, 2012.
[99] T. Pattinson, "Quantification and description of distance measurement errors of a time-of-flight camera," M. Sc. Thesis, University of Stuttgart, Stuttgart, Germany, 2010.
[100] ITU-T, "Subjective video quality assessment methods for multimedia applications," Recommendation P.910, 1999.
[101] T1A1.1 Working Group on Multimedia Communications Coding and Performance, "Methodological Framework for Specifying Accuracy and Cross-Calibration of Video Quality Metrics," 2001.
[102] M. H. Brill, *et al.*, "Accuracy and cross-calibration of video quality metrics: new methods from ATIS/T1A1," *Signal Processing: Image Communication*, vol. 19, pp. 101-107, 2004.
[103] Computer Graphics Laboratory, Stanford University, "*The (New) Stanford Light Field Archive"*. Available: http://lightfield.stanford.edu/lfs.html
[104] R. K. Mantiuk, *et al.*, "Comparison of four subjective methods for image quality assessment," in *Computer Graphics Forum*, 2012, pp. 2478-2491.
[105] A. Cayley, "On a new auxiliary equation in the theory of equations of the fifth order," *Philosophical Transactions of the Royal Society of London*, vol. 151, pp. 263-276, 1861.
[106] T. J. Ypma, "Historical development of the Newton-Raphson method," *SIAM review*, vol. 37, pp. 531-551, 1995.