

ADAPTIVE PLENOPTIC SAMPLING: THEORY AND APPLICATIONS

by
CHRISTOPHER GILLIAM

A Thesis submitted in fulfilment of requirements for the degree of
Doctor of Philosophy of Imperial College London

Communications & Signal Processing Group
Department of Electrical & Electronic Engineering
Imperial College London
2012

Statement of Originality

I declare that this thesis, and the research it contains, is the product of my own work under the guidance of my thesis supervisors: Dr. Pier Luigi Dragotti and Mike Brookes. Any ideas or quotations from the work of other people, published or otherwise, are fully acknowledged in accordance with the standard referencing practices of the discipline. The material of this thesis has not been submitted for any degree at any other academic or professional institution.

Abstract

Image-Based Rendering (IBR) is an effective technique for rendering novel views of a scene from multi-view images. The plenoptic function enables IBR to be formulated in terms of sampling and reconstruction. In this thesis, we combine the theoretical results from uniform plenoptic sampling with non-uniform camera placement. The central concept is that geometry of the scene can be modelled with a sequence of slanted planes. The positions of the cameras are then derived from the plenoptic spectral analysis of a slanted plane. To this end, we present novel results for the plenoptic spectral analysis of a slanted plane and an algorithm for adaptive plenoptic sampling.

The novelty of our spectral analysis lies in the inclusion of two realistic conditions when calculating the plenoptic spectrum: finite scene width and cameras with finite field of view. Using these conditions, we derive an exact closed-form expression for the plenoptic spectrum of a slanted plane with bandlimited texture. From this spectrum, we determine an expression for the maximum spacing between adjacent cameras. Using synthetic and real scenes, we show that this expression is a more accurate gauge of the Nyquist sampling density than the current state-of-the-art.

Based on these results, we design an adaptive plenoptic sampling algorithm for a scene with a smoothly varying surface and bandlimited texture. The algorithm operates by determining the best sequence of slanted planes to model the scene given its geometry and a limited number of cameras. Once this sequence of planes is obtained, the algorithm then positions the cameras using our sampling analysis of a slanted plane. Using synthetic and real scenes, we show that this algorithm outperforms uniform sampling. Finally, we also present a novel reconstruction filter for plenoptic sampling that

outperforms the state-of-the-art for both synthetic and real scenes. The filter uses interpolators of maximum-order-minimal-support (MOMS).

Acknowledgement

There are many people I would like to thank for helping me through this PhD.

First and foremost, I would like to thank my supervisors Pier Luigi Dragotti and Mike Brookes. Their constant support and guidance throughout these 4 years has been invaluable. I am deeply grateful to them for offering me the chance to undertake a PhD and all the opportunities associated with it. I also thank them for their time and patience (especially during the writing of this thesis). It has been a pleasure to work with you both, thank you.

I would then like to thank my friends from the CSP Group. In particular, I want to mention Andriy Gelman, David Looney, James Pearson, Mark Thomas, Sira Gonzalez and Nick Gaubitch. You have given this PhD a social side.

Also, I wish to thank the 2004-08 EEE group: Edo Abraham, Besart Robo, Liang Long Ho, Can Ozelci and Anand Subramanian. From the Egg Race in our 1st year to the present day, we have had lots of good moments together and I hope for lots more to come.

Staying with Imperial College, I want to thank past and present members of the fencing club in particular Chris Namih, Nathan Harmston, Joan Chang, Hannah Bryars, Emily Bottle, Clare Perry, Ed Gilhead and Alex Bishop. You have made my time at Imperial a great experience - one that I would not change for the world. Thank you.

Finally, I would like to thank my family for their love and support throughout. They have worried on my behalf and offered encouragement at every step - for which I am eternally grateful.

Last but not least, I thank Beth - you are my inspiration.

Contents

Statement of Originality	3
Abstract	5
Acknowledgement	7
Contents	9
List of Figures	13
Abbreviations	25
List of Symbols	27
Chapter 1. Introduction	33
1.1 Motivation	33
1.2 Problem Statement	35
1.3 Original Contribution	36
1.4 Thesis Outline	39
Chapter 2. Image Based Rendering and the Plenoptic Function	41
2.1 Introduction	41
2.2 Parametrisation of the Plenoptic Function	43
2.2.1 Light Field and Surface Light Field Parametrisations	44
2.2.2 Epipolar Plane Image	46
2.3 Plenoptic Sampling	47

2.3.1	Uniform Sampling in a Fourier Framework	48
2.3.2	Non-Uniform Sampling	54
2.3.3	Reconstructing the Sampled Plenoptic Function	56
2.4	Depth Information in IBR	57
2.5	Summary	60
Chapter 3. Plenoptic Spectral Analysis		61
3.1	Introduction	61
3.2	Scene Geometry for a Slanted Plane	62
3.3	Spectral Analysis under F FoV and FSW	64
3.3.1	Analysis of a Slanted Plane	67
3.3.2	Extending to Multiple Slanted Planes	72
3.4	Behaviour of the Spectrum for a Slanted Plane	74
3.4.1	Varying the Angle of Slant	76
3.4.2	Varying the Maximum Frequency of the Texture	78
3.4.3	Varying the Minimum Depth of the Scene	78
3.4.4	Varying the Field of View	79
3.5	Rotation of the Camera Line	79
3.6	Summary	81
Chapter 4. Uniform Plenoptic Sampling		83
4.1	Introduction	83
4.2	The Essential Bandwidth for a Slanted Plane	84
4.2.1	The Essential Bandwidth Parameters	85
4.2.2	Validation of the Essential Bandwidth EB_S	93
4.3	Sampling and Reconstructing the Plenoptic Function	95
4.3.1	Analysing the Maximum Camera Spacing	99
4.3.2	Analysing the Reconstruction Filter	100
4.4	Rendering Real Images	105
4.5	Summary	112

Chapter 5. Non-Uniform Plenoptic Sampling	115
5.1 Introduction	115
5.2 Evaluating the Model of the Scene Surface	116
5.2.1 Estimating the Geometric Error	118
5.2.2 Estimating the Aliasing Error	119
5.2.3 Behaviour of the Distortion Metric	124
5.3 Adaptively Sampling a Smoothly Varying Scene	127
5.3.1 The Surface Optimisation	128
5.3.2 The Adaptive Sampling Algorithm	128
5.3.3 Reconstruction from Adaptive Sample Locations	131
5.4 Results and Simulations for Synthetic Scenes	134
5.4.1 Analysis of Camera Locations	134
5.4.2 Validating the Adaptive Sampling Algorithm	138
5.5 Results and Simulations for Real Scenes	140
5.5.1 Analysis of the Piecewise Planar Scene	141
5.5.2 Analysis of the Smoothly Varying Scene	141
5.6 Summary	145
Chapter 6. Conclusions	147
6.1 Summary of Thesis Achievements	147
6.2 Future Research	149
6.2.1 Extensions to Presented Research	149
6.2.2 Beyond Our Research - Multi-View Depth Cameras	150
Appendix A. The Plenoptic Spectrum for a Slanted Plane	153
A.1 Complex Exponential Texture	155
A.2 Bandlimited Texture	157
Appendix B. Properties of the $P_S(\omega_t, \omega_v)$	161
B.1 The Consistency Property	161
B.2 The Symmetry Property	163

Appendix C. The Plenoptic Spectrum with Rotated Camera Line	165
Appendix D. Determining the Essential Bandwidth for $\text{sinc}'(\omega)$	173
Appendix E. The Cardinal Operator	175
Appendix F. Rendered Images for the Planar Scene	177
Appendix G. Rendered Images for the Smoothly Varying Scene	181
Bibliography	185

List of Figures

- 1.1 Diagram illustrating IBR. The set of multi-view images are shown in (a) and two examples of rendered views in (b). The images in (b) were rendered using the algorithm presented in [45]. 34
- 2.1 Diagram showing the 7D plenoptic function, where (x, y, z) is the viewing position, (v, w) is the viewing direction in terms of pixel coordinates, and τ and μ are the time and wavelength, respectively. 42
- 2.2 Diagram (a) illustrates the 4D light field, $p(t, u, v, w)$, in which a light ray is defined by its location on the camera plane, (t, u) , and its pixel location on the image plane, $(v, w) = (v' - t, w' - u)$. The distance between the two planes is the focal length f . Diagram (b) illustrates the surface light field, $l(s, r, \theta_s, \theta_r)$, which is the intensity of a light ray emitted from a point (s, r) on the scene surface \mathcal{S} at a viewing direction (θ_s, θ_r) 44
- 2.3 Diagram (a) illustrates the 2D light field, $p(t, v)$, in which a light ray is defined by its intersection with the camera line at a location t and the corresponding pixel location, v , on the image line. Diagram (b) illustrates the 2D surface light field, $l(s, \theta_s)$, in which a light ray is defined by its point of origin, s , on the object surface, \mathcal{S} , at a viewing angle θ_s . 46
- 2.4 Diagram showing the 2D parametrisation of the light field (a), and its EPI representation, (b). A point in (a) translates to a line in the EPI with a slope inversely proportional to the depth of the point. 47

- 2.5 Diagrams of the plenoptic spectrum: (a) the plenoptic spectrum bounded between $\omega_v = \omega_t z_{\min}/f$ and $\omega_v = \omega_t z_{\max}/f$. (b) the ‘Bow-tie’ shaped plenoptic spectrum caused by the pixel resolution, Δv , inducing low-pass filtering in ω_v . (c) the optimal packing for the sampled plenoptic spectrum, where Δt is the camera spacing. (d) the reconstruction filter, $\Psi(\omega_t, \omega_v)$, required for perfect reconstruction. 49
- 2.6 General scene model showing the intersection of a light ray (t, v) with the scene surface at $(x, z(x))$, where $z(x) \in [z_{\min}, z_{\max}]$ is the depth of the surface. Note that f is the focal length of the cameras and θ is the viewing angle. 52
- 2.7 Diagrams illustrating how new images are introduced when using AIC to non-uniformly sample the scene. In (a) a single image is introduced in a concentric mosaics representation [67] and in (b) multiple images are introduced in a 4D light field representation [70]. A new image is inserted between an image pair based on the reconstruction error. 55
- 2.8 Diagram showing the number of images, N_c , required in IBR as a function of the depth layers, N_d , available. The minimum sampling curve, proposed in [10], is determined from (2.20). Point A in the diagram marks the number of images required if $N_d = 1$, it is determined using Equation (2.17). 58
- 3.1 Diagram illustrating the scene geometry for a slanted plane with a texture signal, $g(s)$, pasted to the surface. The plane is dictated by three elements; the slant of the plane, ϕ , the width of the plane, T , and the starting position of the surface, (x_1, z_1) . The texture signal is defined in terms of a curvilinear coordinate $s \in [0, T]$. Notice that if $\phi > 0$ then $z_2 > z_1$, however, if $\phi < 0$ then $z_2 < z_1$ 63

- 3.2 Diagram illustrating the windowing effects of FSW and FFoV in the EPI domain for a synthetic EPI with bandlimited texture. The constraints are as follows; (a) unconstrained EPI, (b) only FFoV, (c) only FSW, and (d) both FFoV and FSW. 68
- 3.3 Diagram illustrating the windowing effects of FSW and FFoV on the plenoptic spectrum for a synthetic EPI with bandlimited texture. The constraints are as follows; (a) unconstrained EPI, (b) only FFoV, (c) only FSW, and (d) both FFoV and FSW. The corresponding EPI's are shown in Figure 3.2. 69
- 3.4 Diagram (a) shows the plenoptic spectrum for a slanted plane with a texture signal $g(s) = \cos(\omega_s s)$. The same spectrum is shown in (b), (c) and (d) each with different characteristics of the spectrum superimposed. These characteristics are: (b) the diagonal lines relating to the maximum and minimum depths of the plane, (c) the four vertical lines relating to the modulation of the texture signal when projected onto the image plane and (d) the two quadrilateral regions enclosed, \mathcal{T}_{R1} and \mathcal{T}_{R2} . Note that $\Omega_{\max} = \omega_s f / (f \cos(\phi) - \sin(\phi)v_m)$ and $\Omega_{\min} = \omega_s f / (f \cos(\phi) + \sin(\phi)v_m)$. 74
- 3.5 Diagrams showing the behaviour of the magnitude of the plenoptic spectrum for a slanted plane, $|P_S(\omega_t, \omega_v)|$, as four scene parameters are varied. Note that, for clarity, the magnitude of each spectrum is normalised. 77
- 3.6 Diagram illustrating the rotation of the camera line around a point x_r at an angle α . Note that d_1 and d_2 are the distances between the camera line and plane at x_1 and x_2 , respectively. 80

- 4.1 Diagram illustrating the parametric essential bandwidth for the plenoptic spectrum corresponding to a slanted plane with bandlimited texture. The essential bandwidth, EB_S , is superimposed on an example spectrum in which the texture signal is $g(s) = \cos(\omega_s s)$. The model of the essential bandwidth comprises four parameters: Ω_t and Ω_v , the maximum values in ω_t and ω_v respectively, z_G/f the parameter that controls skew of the model relative to the ω_v -axis, and A the width of the essential bandwidth in ω_t 85
- 4.2 Graph validating the essential bandwidth for the signal expressed in (4.20). It shows the percentage of the signal's energy that is contained within the region defined in (4.26) as the angle ϕ varies. The angle increases from zero to just under the no-occlusion constraint. 92
- 4.3 Graphs illustrating the bandlimited signal used as texture for the synthetic scenes. The signal in graph (a) has a maximum frequency of $\omega_s = 40\pi$ rad/m whereas the signal in graph (b) has a maximum frequency of $\omega_s = 50$ rad/m. In both cases the signals are plotted against a curvilinear coordinate $s \in [0, 3.5\text{m}]$ 94
- 4.4 Graph showing the variation in the percentage of energy within the essential bandwidth for the plenoptic spectrum generated by a slanted plane with bandlimited texture. Changes in the angle of slant of the plane, ϕ , are illustrated along the y -axis and changes in the maximum frequency of the texture signal, ω_s , along the x -axis. The other parameters are as follows; $x_1 = 0$ m, $z_{\min} = 1.5$ m and $T = 3.5$ m. The cameras have a focal length of 32 mm, corresponding to a FoV $\approx 40^\circ$ 95
- 4.5 Diagrams showing the synthetic EPIs and plenoptic spectra for three different slanted planes. The EPIs, $p(t, v)$, are shown in (a), (c) and (e), and their plenoptic spectra, $|P_S(\omega_t, \omega_v)|$, in (b), (d) and (f), respectively. The three planes have the following parameters in common; $x_1 = 0$ m, $z_{\min} = 2.1$ m, $T = 3.5$ m and $\omega_s = 50$ rads/m. In (a) and (b) $\phi = 3^\circ$, (c) and (d) $\phi = 30^\circ$, and (e) and (f) $\phi = 60^\circ$ 97

- 4.6 Diagrams showing the synthetic EPIs and plenoptic spectra for three different slanted planes. The EPIs, $p(t, v)$, are shown in (a), (c) and (e), and their plenoptic spectra, $|P_S(\omega_t, \omega_v)|$, in (b), (d) and (f), respectively. The three planes have the following parameters in common; $x_1 = 0$ m, $z_{\min} = 2.1$ m, $T = 3.5$ m and $\omega_s = 125$ rads/m. In (a) and (b) $\phi = 3^\circ$, (c) and (d) $\phi = 30^\circ$, and (e) and (f) $\phi = 60^\circ$ 98
- 4.7 Graphs comparing the PSNR curves for the reconstruction of three EPIs using two different reconstruction filters. Each EPI relates to a slanted plane with the following parameters in common; $x_1 = 0$ m, $z_{\min} = 2.1$ m, $T = 3.5$ m and $\omega_s = 50$ rads/m. In (a) $\phi = 3^\circ$, (b) $\phi = 30^\circ$ and (c) $\phi = 60^\circ$. The filters differ in their skew; the first has a skew of z_G/f and the second z_C/f . N_G and N_C are the number of cameras required to achieve Nyquist sampling for each reconstruction method. 101
- 4.8 Graphs comparing the PSNR curves for the reconstruction of three EPIs using two different reconstruction filters. Each EPI relates to a slanted plane with the following parameters in common; $x_1 = 0$ m, $z_{\min} = 2.1$ m, $T = 3.5$ m and $\omega_s = 125$ rads/m. In (a) $\phi = 3^\circ$, (b) $\phi = 30^\circ$ and (c) $\phi = 60^\circ$. The filters differ in their skew; the first has a skew of z_G/f and the second z_C/f . N_G and N_C the number of cameras required to achieve Nyquist sampling for each reconstruction method. 102
- 4.9 Graphs comparing the reconstruction of three EPIs as the skew of the reconstruction filter varies from z_{\min}/f to z_{\max}/f . The intensity in each image represents the difference in PSNR relative to the reconstruction when the skew is z_G/f . Each EPI relates to a slanted plane with the following parameters in common; $x_1 = 0$ m, $z_{\min} = 2.1$ m, $T = 3.5$ m and $\omega_s = 125$ rads/m. 103

- 4.10 The set-up and data acquisition for the EPI-volume corresponding to a scene with a single slanted plane. Diagram (a) illustrates the bird's eye view of the scene geometry and (b) shows the data acquisition. The resulting EPI-volume consists of 133 images spaced 1 cm apart. Each image is 3008 by 1888 pixels. 104
- 4.11 Graph (a) shows the PSNR curves for the reconstruction of an EPI-volume, relating to a slanted plane, using a filter skew of z_G/f and z_C/f . Graph (b) compares the reconstruction of the same EPI volume as the skew of the reconstruction filter varies from 25.7 to 37.1. The intensity in each image represents the difference in PSNR relative to the reconstruction when the skew is z_G/f . The parameters for this slanted plane are; $x_1 = 1.15$ m, $z_{\min} = 1.65$ m, $T = 1.246$ m and $\phi = 57.5^\circ$. The cameras have a focal length of 70 mm, which corresponds to a FoV $\approx 20^\circ$. The Nyquist number of cameras required for Δt_G and Δt_C are indicated with N_G and N_C , respectively. 106
- 4.12 Graph (a) compares a 2nd order cardinal O-MOMS to the sinc function. Graph (b) compares a 3rd order I-MOMS to the sinc function. Notice that the O-MOMS has infinite support and discontinuities, whereas the I-MOMS is of compact support. 107
- 4.13 Graph (a) compares the PSNR curves for the reconstruction of the EPI-volume obtained using a reconstruction filter with a 3rd order I-MOMS, a 2nd order cardinal O-MOMS and a sinc function. The skew of the filter is z_G/f in all three cases. The graph also includes the PSNR curve obtained using the baseline algorithm with a constant depth of z_G . Graph (b) shows the PSNR curves for the reconstruction of the same EPI-volume using the I-MOMS but with differing filter skews, z_G/f and z_C/f . The graph also shows the PSNR curve for the baseline algorithm. The parameters for this slanted plane are; $x_1 = 1.15$ m, $z_{\min} = 1.65$ m, $T = 1.246$ m and $\phi = 57.5^\circ$. The cameras have a focal length of 70 mm, which corresponds to a FoV $\approx 20^\circ$ 109

4.14	An example of a rendered image from the EPI-volume using four different reconstruction methods. The original image is shown in (a), it's rendering using a sinc function in (b), using the baseline algorithm in (c), using a 2 nd order O-MOMS in (d) and lastly, using a 3 rd order I-MOMS in (e). The filter skew in each case is z_G/f . Each image is rendered using 34 cameras.	111
4.15	An example of a rendered image from the EPI-volume using four different reconstruction methods. The original image is shown in (a), it's rendering using a sinc function in (b), using the baseline algorithm in (c), using a 2 nd order O-MOMS in (d) and lastly, using a 3 rd order I-MOMS in (e). The filter skew in each case is z_C/f . Each image is rendered using 34 cameras.	112
5.1	Diagram illustrating the regions \mathcal{O}_a , \mathcal{O}_b and \mathcal{O}_c used to approximate the energy of the plenoptic spectrum that resides outside the support of the reconstruction filter $\mathcal{R}_{\psi,i}$	120
5.2	Diagram showing a piecewise quadratic scene, which consists of three quadratic pieces.	123
5.3	Graph (a) plots the estimated distortion for the plenoptic function as the number of planes used to model the surface increases from 5 to 150. The plenoptic function relates to the surface shown in Figure 5.2. The graph compares the distortion for 20, 40 and 80 cameras. Graphs (b) and (c) show respectively the amount of aliasing and geometric error that comprises the distortion in (a). Note that the cameras have a FoV = 20°, in which $f = 32$ mm, and $\omega_s = 125$ rads/m.	126
5.4	Diagram showing how the model of the scene surface, using a sequence of slanted planes, is optimised within a binary tree framework. Note that each leaf node in the tree represent a plane in the sequence, hence pruning the tree means merging two planes into one.	129

- 5.5 Diagrams illustrating the non-uniform placement of 20 cameras using a piecewise constant sample density in t . Graph (a) shows the piecewise constant sample density for the scene as a function of t . Graph (b) shows the warping function generated from the sample density. The function is constructed by integrating the sampling density, hence it is piecewise linear. The graph also shows how uniform sampling in the warped domain is converted to non-uniform sampling in t 130
- 5.6 Graphs illustrating four different reconstruction filters for a sampled synthetic EPI. The EPI corresponds to the quadratic scene in Figure 5.2. It is sampled uniformly in (a) and (b), and adaptively in (c) and (d) using the algorithm in Section 5.3.2. Accordingly, in (a) the filter is uniform with a constant filter skew of f/z_G . In (b) the filter corresponds to uniform sampling but it has a varying filter skew of $f/z(t, v)$. In (c) the filter is adaptive but with a fixed filter skew of f/z_G . Finally, in (d) the filter is adaptive with a varying skew of $f/z(t, v)$. Note that in all of the graphs the filter is superimposed on the synthetic EPI and the actual sample are highlighted. 133
- 5.7 Diagrams showing the position of N_c cameras generated by the adaptive sampling algorithm for a piecewise quadratic surface, comprising 5 pieces, with texture bandlimited to 60 rads/m. The positions of the cameras are indicated along the x -axis and, below each surface, is the corresponding sampling density as a function of x . The sampling density is relative to uniform sampling. Each graph also shows the planar model of the surface generated by the algorithm. 135

- 5.8 Diagrams showing the position of N_c cameras generated by the adaptive sampling algorithm for a piecewise quadratic surface, comprising 5 pieces, with texture bandlimited to 120 rads/m. The positions of the cameras are indicated along the x -axis and, below each surface, is the corresponding sampling density as a function of x . The sampling density is relative to uniform sampling. Each graph also shows the planar model of the surface generated by the algorithm. 136
- 5.9 Diagrams showing the position of N_c cameras generated by the adaptive sampling algorithm for a piecewise quadratic surface, comprising 5 pieces, with texture bandlimited to 300 rads/m. The positions of the cameras are indicated along the x -axis and, below each surface, is the corresponding sampling density as a function of x . The sampling density is relative to uniform sampling. Each graph also shows the planar model of the surface generated by the algorithm. 137
- 5.10 Graphs comparing the PSNR curves for the reconstruction of three synthetic EPIs when sampled uniformly and adaptively. The adaptive sample positions are determined using our sampling algorithm. The EPIs correspond to a piecewise quadratic surface, comprising 5 pieces, with texture bandlimited to (a) 60 rads/m, (b) 120 rads/m and (c) 300 rads/m. Note that the cameras have a focal length equal to 70 mm and a FoV $\approx 20^\circ$ 139
- 5.11 The set-up and data acquisition for the EPI-volume corresponding to the planar scene. Diagram (a) illustrates the bird's eye view of the scene geometry and (b) shows the data acquisition. The resulting EPI-volume consists of 253 images spaced 1 cm apart. Each image is 3008 by 1888 pixels. 140

- 5.12 Graph comparing the PSNR curves for the reconstruction of a EPI-volume. The curves correspond to adaptive sampling, uniform sampling and uniform sampling with the baseline algorithm. The EPI-volume relates to the three plane scene shown in Figure 5.11. The adaptive samples are determined using our sampling algorithm. The cameras have a focal length of 70 mm, which corresponds to a FoV $\approx 20^\circ$ 142
- 5.13 An example of a rendered image from a EPI-volume using three different sampling and reconstruction methods. The EPI-volume relates to the three plane scene shown in Figure 5.11. The original image is shown in (a), it's rendering using uniform sampling in (b), it's rendering using the baseline algorithm uniform in (c), and it's rendering using adaptive sampling in (d). Each image is rendered using 37 cameras. 142
- 5.14 The set-up and data acquisition for the EPI-volume corresponding to the smoothly varying scene. Diagram (a) illustrates the bird's eye view of the scene geometry. Diagram (b) shows the twin data acquisition of both colour and depth images using a standard camera and Microsoft's Xbox Kinect. The resulting EPI-volume consists of 73 images spaced 1 cm apart. Each image is 3008 by 2000 pixels. We also obtain 73 depth images spaced 1 cm apart, but at a resolution of 374 by 248 pixels. . . . 143

5.15	Graph comparing the PSNR curves for the reconstruction of a EPI-volume. The curves correspond to adaptive sampling, uniform sampling and uniform sampling with the baseline algorithm. The EPI-volume relates to the smoothly varying scene shown in Figure 5.14. The adaptive samples are determined using our sampling algorithm. The cameras have a focal length of 34 mm, which corresponds to a FoV $\approx 39^\circ$	144
5.16	An example of a rendered image from a EPI-volume using three different sampling and reconstruction methods. The EPI-volume relates to the smoothly varying scene shown in Figure 5.14. The original image is shown in (a), it's rendering using uniform sampling in (b), it's rendering using the baseline algorithm in (c), and it's rendering using adaptive sampling in (d). Each image is rendered using 10 cameras.	144
C.1	Diagram illustrating a light ray travelling from the scene to a camera at position t . The camera line is rotated around a point x_r at an angle α . Note that θ is the viewing angle and v is the relating pixel position. . .	166
D.1	Graph (a) illustrates the Energy Spectral Density (ESD) for $\text{sinc}'(\omega)$. The percentage of the overall energy within a bandwidth W as a function of W is shown in (b).	174
F.1	Rendered image 1.	178
F.2	Rendered image 2.	178
F.3	Rendered image 3.	179
F.4	Rendered image 4.	179
F.5	Rendered image 5.	180
F.6	Rendered image 6.	180
G.1	Rendered image 1.	182
G.2	Rendered image 2.	182
G.3	Rendered image 3.	183
G.4	Rendered image 4.	183

G.5	Rendered image 5.	184
G.6	Rendered image 6.	184

Abbreviations

3DTV:	Three Dimensional Television
AIC:	Active Incremental Capturing
ARC:	Active Rearranged Capturing
DIBR:	Depth Image-Based Rendering
EB:	Essential Bandwidth
EPI:	Epipolar Plane Image
ESD:	Energy Spectral Density
FoV:	Field of View
FFoV:	Finite Field of View
FSW:	Finite Scene Width
FTV:	Free-viewpoint Television
IBR:	Image-Based Rendering
LDI:	Layered Depth Image
MBR:	Model-Based Rendering
MOMS:	Maximal-Order-Minimal-Support
I-MOMS:	Interpolating Maximal-Order-Minimal-Support
O-MOMS:	Optimal Maximal-Order-Minimal-Support
PSNR:	Peak Signal-to-Noise Ratio
SR:	Sample Reduction

List of Symbols

Coordinate Variables

r	Surface curvilinear coordinate (y-axis)
s	Surface curvilinear coordinate (x-axis)
t	Camera spatial coordinate (x-axis)
u	Camera spatial coordinate (y-axis)
v	Pixel spatial coordinate (x-axis)
\bar{v}	Pixel coordinate scale by focal length, v/f
w	Pixel spatial coordinate (y-axis)
x	World spatial coordinate (x-axis)
y	World spatial coordinate (y-axis)
z	World spatial coordinate (z-axis)
μ	Wavelength
θ_s	Angular direction a light ray leaves a surface point s relative to the z -axis
θ_r	Angular direction a light ray leaves a surface point r relative to the z -axis
ϑ_x	Angular direction relative to x -axis
ϑ_y	Angular direction relative to y -axis
τ	Time
ω_t	Frequency coordinate equivalent to t
ω_v	Frequency coordinate equivalent to v

Functions

$d_{\text{MOMS}}^{\{M\}}(i)$	Discrete MOMS of order M
$e_i(t, v)$	Error function in the EPI-domain for the i th slanted plane in a sequence
$g(s)$	Texture signal pasted to the scene surface
$h(s)$	Windowed texture signal
$l(s, r, \theta_s, \theta_r)$	4D Surface light field
$l(s, \theta_s)$	2D Surface light field
$l_x(x, \bar{v})$	2D Intensity of a light ray emitted from a spatial position $(x, z(x))$ at a viewing direction \bar{v}
$n(\phi, \bar{v}_m)$	Weight function for the essential bandwidth of a slanted plane
$p(x, y, z, v, w, \mu, \tau)$	7D Plenoptic function
$p(t, u, v, w)$	4D Light field
$p(t, v)$	2D Light field
$\tilde{p}_i(t, v)$	2D Light field for the i th slanted plane in a sequence
$x(s)$	Function mapping the curvilinear coordinate s to x
$z(x)$	Depth of the scene surface as a function of x
$D_{\text{MOMS}}^{\{M\}}(z)$	Z-transform of the discrete MOMS of order M
$E_1(x)$	Exponential integral of x
$E_1^*(X)$	Complex conjugate of the Exponential integral of x
$G(\omega)$	Fourier transform of the texture signal $g(s)$
$H(\Omega)$	Fourier transform of the signal $h(s)$
$L_x(x, \omega_v)$	1D Fourier transform of $l_x(x, \bar{v})$ with respect to \bar{v}
$P(\omega_t, \omega_v)$	2D Plenoptic spectrum (2D Fourier transform of $p(t, v)$)
$P_{FPP}(\omega_t, \omega_v)$	2D Plenoptic spectrum for a fronto-parallel plane
$P_{MP}(\omega_t, \omega_v)$	2D Plenoptic spectrum for a scene comprising multiple slanted planes
$P_S(\omega_t, \omega_v)$	2D Plenoptic spectrum for a slanted plane
$\beta_M(x)$	B-spline of order M

$\delta(x)$	Dirac delta function
$\varphi_{\text{MOMS}}^{\{M\}}(x)$	MOMS of order M
$\eta(x)$	Interpolating function
$\eta_{\text{MOMS}}^{\{M\}}(x)$	Cardinal MOMS function of order M
$\eta_I^{\{M\}}(x)$	Cardinal I-MOMS function of order M
$\eta_O^{\{M\}}(x)$	Cardinal O-MOMS function of order M
$\xi_{\text{MOMS}}^{\{M\}}(i)$	Inverse Z-transform of $1/D_{\text{MOMS}}^{\{M\}}(z)$
$\psi(t, v)$	Uniform reconstruction filter in the EPI-domain
$\phi_{i,k}(t, v)$	Warped reconstruction filter for the i th sample in t and k th sample in v
$\zeta(x)$	Function comprising Exponential integrals
$\Phi_{\text{MOMS}}^{\{M\}}(\omega)$	Fourier transform of $\varphi_{\text{MOMS}}^{\{M\}}(x)$
$\Psi(\omega_t, \omega_v)$	Fourier transform of the uniform reconstruction filter
$\mathcal{W}(t)$	Warping function for the camera line

Variables

a, b, c	Variables used to define $P_S(\omega_t, \omega_v)$
f	Focal length of the cameras
\hat{t}	Warped camera line
t_{range}	Overall support of the plenoptic function in t
v_m	Maximum value in v due to FFoV
\bar{v}_m	Maximum value in \bar{v} due to FFoV, v_m/f
x_1	Spatial starting position of a plane in x
x_2	Spatial ending position of a plane in x
z_1	Spatial starting position of a plane in z
z_2	Spatial ending position of a plane in z
z_{max}	Maximum depth of the scene
z_{min}	Minimum depth of the scene
z_C	Depth proposed in [10] for the skew of the reconstruction filter

z_G	Depth proposed in this thesis for the skew of the reconstruction filter
A	Width in ω_t of the essential bandwidth for a slanted plane
E_a, E_b, E_c	Energy contributions used to determine an estimate of the aliasing error
$E_{\text{sinc}}, E_{\text{dsinc}}$	Energy contributions relating to a sinc function and its derivative
E_T	Total energy of the plenoptic function for a slanted plane
EB_S	Essential bandwidth for the plenoptic function of a slanted plane
L_s	Number of planes
N_c	Number of cameras
N_d	Number of depth layers
N_p	Number of pixels
N_v	Number of views
N_G	Number of samples required to achieve Δt_G
N_C	Number of samples required to achieve Δt_C
T	Width of the plane
α	The angle the camera line is rotated relative to x -axis
ϕ	Angle of slant of a plane relative to x -axis
γ_i	Geometric error for the i th plane in the sequence
$\tilde{\gamma}_i$	Estimated geometric error for the i th plane in the sequence
λ	Lagrangian multiplier
ω_s	Maximum frequency of the texture signal $g(s)$
Δt	Spacing between adjacent cameras assuming uniform distribution
Δt_G	Maximum camera spacing proposed in this thesis
Δt_C	Maximum camera spacing proposed in [10]
Δv	Finite pixel resolution of the camera
Δz	Depth range of the scene, $z_{\text{max}} - z_{\text{min}}$
Γ	Width variable for a Lorentzian function

Λ_i	Aliasing error for the i th plane in the sequence
$\tilde{\Lambda}_i$	Estimated aliasing error for the i th plane in the sequence
Ω_s	Maximum frequency of $H(\Omega)$ assuming its essential bandwidth
Ω_t	Maximum value in ω_t used to determine EB_S
Ω_v	Maximum value in ω_v used to determine EB_S
\mathcal{G}_{FPF}	Scene geometry equations describing a fronto-parallel plane
\mathcal{G}_{MP}	Scene geometry equations describing a scene comprising multiple planes
\mathcal{G}_S	Scene geometry equations describing a slanted plane
$\mathcal{O}_a, \mathcal{O}_b, \mathcal{O}_c$	Regions in frequency used to estimate the aliasing error
\mathcal{P}	Support of the plenoptic function in the EPI-domain
\mathcal{R}_ψ	Support of the uniform reconstruction filter in the frequency domain
\mathcal{S}	Scene surface
$\mathcal{T}_{R1}, \mathcal{T}_{R2}$	Quadrilateral regions that characterise the plenoptic spectrum of a slanted plane

Operators

\mathcal{C}	Cardinal operator (generates cardinal MOMS)
\mathcal{F}	Fourier transform operator
\mathcal{Z}	Z-transform operator

Chapter 1

Introduction

1.1 Motivation

Visual media is currently undergoing a transition from a two dimensional based display system to one that offers users a three dimensional (3D) experience. Examples of this transition can be seen in the recent increase in 3D films, like Avatar, the deployment of consumer 3D displays [33], such as stereoscopic TV sets, and the emergence of dedicated broadcast channels to supply 3D content. One way of creating this 3D experience is 3DTV, which offers the user a depth impression of the observed scene [34]. However this experience is limited when compared to the real world as the user's viewing position within the scene is either fixed or severally restricted. A solution is free-viewpoint television (FTV) [60], which offers the user an interactive 3D experience by allowing them free control over the viewpoint within the scene. This type of user interactivity has many potential applications from televised sporting events to virtual guided tours of museums. The free-viewpoint experience is achieved by creating new views of the scene from a collection of images, each taken from a different direction. The technique is known as virtual view synthesis and is a topic within the growing area of multi-view imaging.

From a multi-view imaging perspective, there are two approaches to performing virtual view synthesis. The first approach, from 3D computer graphics, involves exact modelling of the scene using 3D meshes and light source descriptions. A virtual view is

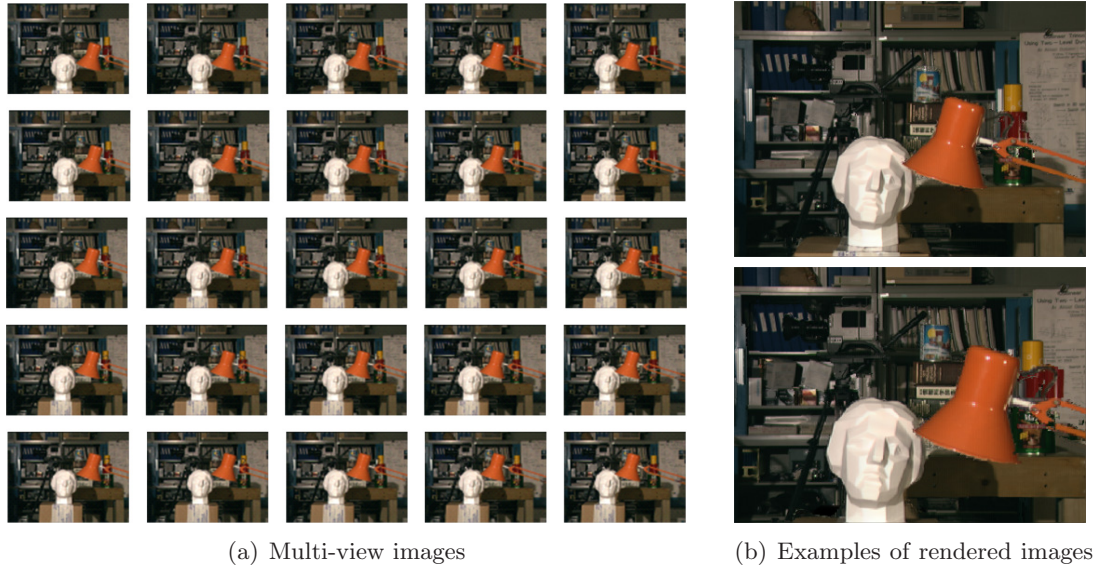


Figure 1.1: Diagram illustrating IBR. The set of multi-view images are shown in (a) and two examples of rendered views in (b). The images in (b) were rendered using the algorithm presented in [45].

then generated by warping the existing images of the scene onto the 3D geometry and then projecting the result into the image domain. This approach is known as model-based rendering (MBR). The 3D model of the scene is generated using multi-view stereo vision algorithms [48]. These algorithms operate by matching consistent points, lines or patches across the set of multi-view images. The matches relate to objects and surfaces within the scene, and are used to generate a 3D model. However, the disadvantage of this approach is that this matching process is typically expensive and error prone, and hence it requires human assistance for high quality rendering [34].

The second approach, and the focus of this thesis, is known as image-based rendering (IBR). The underlying principle is that since each image captures a set of light rays travelling from the scene to the camera, it is possible to synthesise a new view by interpolating nearby light rays. Therefore, instead of using an explicit 3D model, the multi-view images of the scene are used to create new views directly [11]. An example of IBR is shown in Figure 1.1. The initial set of multi-view images are shown in Figure 1.1(a) and two rendered images from the set are shown in Figure 1.1(b). The images are rendered using the algorithm presented in [45]. The figure also illustrates

the main benefit of interpolating from real images: photorealistic rendering quality. Therefore IBR offers high quality rendering from a set of images without the difficulty of determining 3D models. Such flexibility has allowed IBR to find application in many areas aside from FTV. Examples of these applications include surveillance [56], immersive communications [3, 18] and virtual or augmented reality [64].

There are, however, some important challenges involved in IBR. In particular, although it is easy to capture and store a single image, high quality rendering in IBR requires a dense sampling of the real world [34]. For instance the Stanford multi-camera array [62] contains 100 video cameras, each 640 by 480 pixels with a frame rate of 30 frame/sec. Therefore, even with the recent advancements in computing technology, the requirement for so many images imposes significant costs. The penalty for not having enough cameras is the appearance of artefacts in the rendering process. In view of this, an important topic of research in IBR is to determine the minimum number of cameras required for artefact-free rendering. It is this particular problem that we examine in this thesis. We also examine the associated problem of determining the best position of these cameras.

1.2 Problem Statement

A natural framework for studying IBR is the seven dimensional plenoptic function [2]. This function specifies the intensity of a light ray passing through a 3D spatial location, in a certain direction, with a certain wavelength and at a certain time [50]. Using this function, IBR can be regarded in terms of sampling and reconstruction. Assuming a pin-hole camera model, the set of multi-view images represents samples of the continuous plenoptic function and the rendering of a new view entails the function's reconstruction from the samples [69].

Within this framework, artefacts in the rendering process, such as blurring and ghosting effects, are manifestations of aliasing caused by undersampling the plenoptic function [10]. Therefore, we can determine the minimum number of images required for artefact-free rendering by analysing the sampling of the plenoptic function. In

particular, we want to determine the minimum sampling of the plenoptic function required for alias-free reconstruction. Assuming uniform sampling, the authors of [10] addressed this problem in a Fourier framework. They performed a spectral analysis of the plenoptic function to determine the maximum distance between adjacent cameras for alias-free reconstruction. In contrast non-uniform sampling of the plenoptic function, such as [47, 67, 71], is based on minimising an estimate of the rendering error between two adjacent cameras.

In this thesis, we aim to combine theoretical results from the spectral analysis of the plenoptic function with non-uniform camera placement. To achieve this, we start by analysing the uniform sampling of the plenoptic function of a slanted plane. The purpose of this analysis is to determine the maximum spacing between adjacent cameras for a slanted plane. Although this scene is very simple, our interest comes from using it as an elementary element with which to construct more complicated geometries. To this end, we propose approximating the scene of interest using a set of slanted planes. The position of the cameras are then determined by the plenoptic sampling analysis for each slanted plane we use to approximate the scene. Accordingly, unless the same sampling is required for all the planes, the placement of the cameras will be non-uniform. Note that the focus of this thesis is on the sampling of the plenoptic function rather than the effect of depth information in IBR. As a result, to simplify this analysis, we concentrate on IBR rather than depth IBR (DIBR).

1.3 Original Contribution

The following aspects of this thesis are believed to be original contributions:

Plenoptic Spectral Analysis under Realistic Conditions

For the first time, we present a spectral analysis of the 2D plenoptic function that incorporates two realistic conditions: finite scene width and cameras with finite field of view. The novelty of our spectral analysis is that these realistic conditions are applied directly when deriving the 2D plenoptic spectrum, thus allowing a greater class of scene

geometry to be analysed. In particular, we are able to derive an exact expression for the 2D plenoptic spectrum of a 1D slanted plane with complex exponential texture in Section 3.3.1 and extend this expression to bandlimited texture in Section 4.2.1. Using these exact expressions, we are then able to provide greater insight into the behaviour of the plenoptic spectrum of a slanted plane, which in turn leads to greater understanding in plenoptic sampling.

Uniform Plenoptic Sampling using the Essential Bandwidth

In [10] uniform sampling of the plenoptic function is analysed in a Fourier framework and the maximum spacing between adjacent cameras is determined from the spectral support of the 2D plenoptic function. However, in Chapter 3, we show that the 2D plenoptic spectrum under certain realistic conditions is band-unlimited. In view of this, we present a novel approach to uniform plenoptic sampling using its essential bandwidth - a region in frequency containing approximately 90% of the signal's energy. This approach involves sampling the 2D plenoptic function assuming it is bandlimited to its essential bandwidth. Therefore, in Chapter 4, we determine a non-separable 2D essential bandwidth for the plenoptic function of a slanted plane and use it to sample the scene. From this analysis, in Section 4.3, we present a new expression for the maximum spacing between adjacent cameras for a slanted plane and validate it using both synthetic and real scenes. This validation shows that our expression for the maximum camera spacing is a more accurate gauge of the optimal plenoptic sampling for a slanted plane than the current state-of-the-art.

Non-Uniform Plenoptic Sampling based on Plenoptic Spectral Analysis

In Chapter 5, we present a framework for non-uniformly sampling the 2D plenoptic function for a scene with a smoothly varying surface and bandlimited texture. In this framework, we combine our results from plenoptic spectral analysis with adaptive camera placement. The main concept is that the essential geometry of the scene is captured using a sequence of slanted planes. The cameras are then positioned based on the plenoptic spectral analysis presented in Chapters 3 and 4. Using this framework,

we present a novel algorithm that adaptively positions a finite number of cameras by determining the best model of the scene geometry using slanted planes. This model adapts depending on the geometry of the scene, hence the camera placement adapts as well. In Sections 5.4 and 5.5, we validate the algorithm using synthetic and real scenes and show that the reconstruction from the adaptive samples outperforms uniform sampling.

The Reconstruction Filter for the Plenoptic Function

Assuming a uniform camera distribution, a reconstruction filter for the plenoptic function was presented in [10] based on its spectral support. If we sample the plenoptic function using the essential bandwidth, however, a different filter is required. Following from this, in Section 4.3, we present a new parametrisation of the reconstruction filter and show it to be an improvement when sampling and reconstructing synthetic and real scenes. Furthermore, we improve reconstruction again by using functions of maximum-order-minimal-support (MOMS) to interpolate the plenoptic function as described in Section 4.4. Therefore, taking both contributions into account, we present a novel reconstruction filter for uniform plenoptic sampling of a slanted plane that outperforms existing ones. An adaptive version of this filter is presented in Section 5.3.3. It allows reconstruction from non-uniform plenoptic samples and the incorporation of varying depth information.

Publications

The work presented in this thesis has lead to the following publications:

- C. Gilliam, P. L. Dragotti and M. Brookes, “On Uniform Plenoptic Sampling”, to be submitted, *IEEE Transactions on Image Processing*.
- C. Gilliam, M. Brookes and P. L. Dragotti, “Image Based Rendering and the Sampling of the Plenoptic Function”, in *Emerging Technologies for 3D Video: Creation, Coding, Transmission and Rendering*, F. Dufaux, B. Pesquet-Popescu and M. Cagnazzo Eds., Wiley, 2013.

- C. Gilliam, J. Pearson, M. Brookes and P. L. Dragotti, “Image Based Rendering with Depth Cameras: How Many are Needed?”, in *Proc. IEEE International Conference on Acoustic, Speech and Signal Processing*, Kyoto (Japan), March 2012, pp. 5437-5440.
- C. Gilliam, P. L. Dragotti and M. Brookes, “Adaptive Plenoptic Sampling”, in *Proc. IEEE International Conference on Image Processing*, Brussels, September 2011, pp. 2581-2584.
- C. Gilliam, P. L. Dragotti and M. Brookes, “A Closed-form Expression for the Bandwidth of the Plenoptic Function under Finite Field of View Constraints”, in *Proc. IEEE International Conference on Image Processing*, Hong Kong, September 2010, pp. 3965-3968.

1.4 Thesis Outline

This thesis is organised as follows.

In Chapter 2, we examine the state-of-the-art in sampling of the plenoptic function and its application to IBR. In more detail, we start by examining the structure of the plenoptic function and present two key parametrisations: the light field and the surface light field representations. Once these key parametrisations are established, we examine their sampling and reconstruction in both uniform and non-uniform frameworks. Finally, we discuss the role of depth information when sampling the plenoptic function.

In Chapter 3, we re-examine the spectral analysis of the plenoptic function assuming two realistic conditions; finite scene width and cameras with finite field of view. We show that these conditions lead to spectral spreading in frequency, which results in a band-unlimited plenoptic spectrum. Using these conditions, we then examine the plenoptic spectrum for a simple scene: a slanted plane with bandlimited texture. We derive an exact expression for the plenoptic spectrum of this scene. From this expression, we analyse the behaviour of the plenoptic spectrum and present a model to characterise this behaviour. We end the chapter by extending this analysis to multiple

planes.

In Chapter 4, we study the uniform sampling and reconstruction of the plenoptic function of a slanted plane in a Fourier framework. We explore this sampling and reconstruction of the plenoptic function using its essential bandwidth. In particular, we present a new expression for the maximum spacing between adjacent cameras for a slanted plane and a new parametrisation for the reconstruction filter. We show the validity of these results using synthetic and real plenoptic functions that correspond to slanted planes. The chapter also explores the reconstruction of the plenoptic function using different interpolating functions.

Chapter 5 extends the sampling analysis to a non-uniform framework and generalises the scene geometry. It examines the non-uniform sampling and reconstruction of the plenoptic function relating to a scene with a smoothly varying surface. To sample this scene, we propose a novel framework that combines the results from uniform plenoptic sampling with adaptive camera placement. Based on this framework, we present an adaptive sampling algorithm that adaptively positions cameras to sample the plenoptic function. Using synthetic and real data, we show the validity of the algorithm when compared to uniform sampling. We also present an adaptive filter to reconstruct the non-uniformly sampled plenoptic function.

Finally, Chapter 6 concludes this thesis with a summary of its achievements and presents possible directions for future research.

Chapter 2

Image Based Rendering and the Plenoptic Function

2.1 Introduction

The central concept in IBR is that a scene can be represented as a collection of light rays emanating from its surface. A set of multi-view images, therefore, records the scene as each image captures a collection of light rays travelling from the scene to a camera. Using these images, a virtual view is rendered by interpolating a sub-set of the light rays. This interpolation is easily performed if the number of images available is very large. However, the requirement for so many images imposes significant costs. The number of images can be reduced through knowledge of the scene, such as a 3D model of its geometry. In this case, a virtual view is rendered by projecting each light ray to the required viewpoint via the model. The resulting projections are then combined using interpolation. A disadvantage of this approach is that high quality rendering requires detailed 3D models and the generation of such models can be time consuming. Another approach, and the focus of this thesis, is to model the light rays using the concept of the *plenoptic function* proposed by Adelson and Bergen [2].

The plenoptic function models the intensity of all the light rays passing through each point in space, thus providing a natural framework in which to study IBR [17].

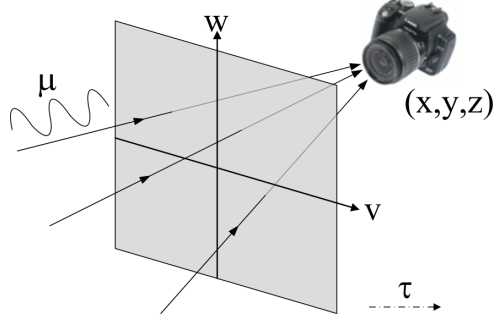


Figure 2.1: Diagram showing the 7D plenoptic function, where (x, y, z) is the viewing position, (v, w) is the viewing direction in terms of pixel coordinates, and τ and μ are the time and wavelength, respectively.

Its full parametrisation comprises seven variables, three for the spatial viewing position, (x, y, z) , two for the angular viewing direction, $(\vartheta_x, \vartheta_y)$, and the last two for the time, τ , and wavelength, μ , dimensions. In many situations, however, it is convenient to parametrise the viewing direction in pixel coordinates, (v, w) . Therefore the full plenoptic function is

$$p_7 = p(x, y, z, v, w, \mu, \tau), \quad (2.1)$$

as illustrated in Figure 2.1.

Using this function, IBR can be seen in terms of sampling and reconstruction. The multi-view images represent the samples of the plenoptic function and the rendering of a new view its reconstruction [69]. If there are too few samples or they are incorrectly positioned, then the plenoptic function will be incorrectly sampled leading to a degradation in the rendering quality. As a result, studying the sampling of the plenoptic function can yield answers to the following: how many images are required in IBR and where should these images be placed.

In view of this, the following chapter presents the state-of-the-art in sampling of the plenoptic function and its application to IBR. It is organised as follows. Section 2.2 describes the range of plenoptic parametrisations that arise by restricting aspects of the scene and the viewing position. In particular the section focuses on the structure of two common plenoptic representations: the light field and the surface light field. We then examine the sampling and reconstruction of the plenoptic function, known as

plenoptic sampling, in Section 2.3. The section covers both uniform and non-uniform camera placement, and the corresponding synthesis in each case. In Section 2.4 we examine the use of depth information in IBR, in particular how rendering quality can be improved with additional geometric knowledge. Finally, in Section 2.5, we end the chapter with a summary of key points relevant to the remainder of the thesis. For other detailed surveys on the plenoptic function and IBR see [52,69] and more recently [34,50].

2.2 Parametrisation of the Plenoptic Function

The high dimensionality of the plenoptic function makes theoretical analysis a challenging problem. A common solution is to reduce this dimensionality by restricting certain aspects of the scene and sensing set-up [69]. With these restrictions, we can re-parametrise the 7D plenoptic function into more tractable representations. In [69], Zhang and Chen categorised plenoptic representations based on a combination of six assumptions required to produce them. The proposed six assumptions can be divided into those that restrict the scene and those that progressively restrict the viewing position, i.e. from 3D to a 2D surface, then a 1D path and finally a fixed position.

For example the 6D surface plenoptic function [68] removes one dimension by assuming the radiance of a light ray is constant along its path through empty space. Likewise, McMillan and Bishop [40] remove both the time and wavelength parameters to construct their 5D plenoptic model. The time is removed by assuming a static scene and then the wavelength is removed by splitting it into three bands (red, green and blue). By using all of these assumptions, coupled with restricting the viewing position to a surface, the authors in [51] construct the concentric mosaic representation. In this representation the scene is captured by a single camera mounted to the end of a rotating beam. As a result the intensity of a light ray is described using just three parameters; the 2D pixel location and the beam’s angle of rotation.

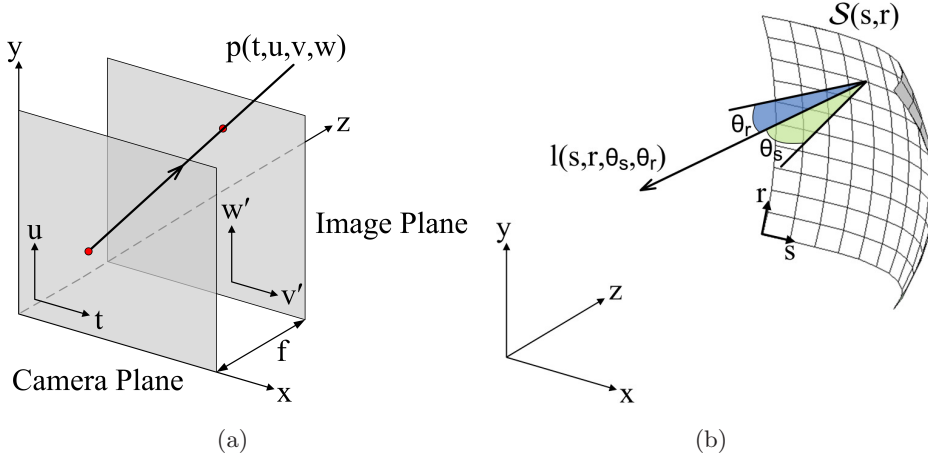


Figure 2.2: Diagram (a) illustrates the 4D light field, $p(t, u, v, w)$, in which a light ray is defined by its location on the camera plane, (t, u) , and its pixel location on the image plane, $(v, w) = (v' - t, w' - u)$. The distance between the two planes is the focal length f . Diagram (b) illustrates the surface light field, $l(s, r, \theta_s, \theta_r)$, which is the intensity of a light ray emitted from a point (s, r) on the scene surface S at a viewing direction (θ_s, θ_r) .

2.2.1 Light Field and Surface Light Field Parametrisations

In this thesis we will focus on two popular plenoptic representations: the light field [37] (similar to the lumigraph [27] or ray-space [24] representations) and the surface light field [41,65]. Both parametrisations assume that the scene is static and the radiance of a light ray is constant along its path through space. Therefore the spatial location of the cameras can be simplified to a 2D surface. As a result both the light field and the surface light field are four dimensional plenoptic representations. They differ, however, in their approach to characterising a light ray using these four dimensions.

In the light field parametrisation the scene is bounded within a box and a pin-hole camera model is assumed. Therefore each light ray is defined by its intersection with two parallel planes, the camera plane, (t, u) , and the image plane, (v, w) . The separation between the two planes is equal to the focal length, f . Therefore the intensity of the light ray at camera location (t, u) and pixel location (v, w) is

$$p_4 = p(t, u, v, w), \quad (2.2)$$

see Figure 2.2(a) for a diagram. By re-introducing the time parameter, [63] proposed

a time dependent version of the light field that can handle dynamic scenes. This new representation was termed light field video. Another variant, is the spherical light field proposed in [28]. In this case the scene is bounded within a sphere and each light ray is defined by its intersection with two concentric spheres. Notice, however, that in this framework, and its variations, the light ray is defined with respect to the receiving camera position.

In contrast, the light ray in the surface light field is defined relative to its point of origin on the scene surface, \mathcal{S} . This surface is parametrised using two curvilinear surface coordinates, (s, r) , such that a point on the surface is defined as $\mathcal{S}(s, r) = [x(s, r), y(s, r), z(s, r)]^T$, where $[x, y, z]^T$ is the real world coordinates [44]. The direction the light ray leaves the surface is defined by the viewing angle (θ_s, θ_r) , where θ_s and θ_r are defined relative to the z -axis. Therefore the intensity of a light ray emitted from a point (s, r) on the scene surface at a viewing direction (θ_s, θ_r) is

$$l_4 = l(s, r, \theta_s, \theta_r). \quad (2.3)$$

Similar to the light field, a diagram of this framework is shown in Figure 2.2(b).

On a final note, several authors [10, 17, 68] further reduce the dimensionality of both parametrisations by considering only a horizontal slice of the scene. In the case of the light field, u and w are fixed; this corresponds to the situation where the camera positions are constrained to a line parallel to the x -axis and only one scan-line is considered in each image. Therefore the intensity of the light ray at camera location t and pixel location v is

$$p_2 = p(t, v). \quad (2.4)$$

For the surface light field, r and θ_r are fixed corresponding to a one dimensional surface, $\mathcal{S}(s)$. Therefore the intensity of the light ray emitted from a surface point $[x(s), z(s)]^T$, at viewing angle θ_s is

$$l_2 = l(s, \theta_s). \quad (2.5)$$

Diagrams of the 2D light field and 2D surface light field are shown in Figure 2.3. For

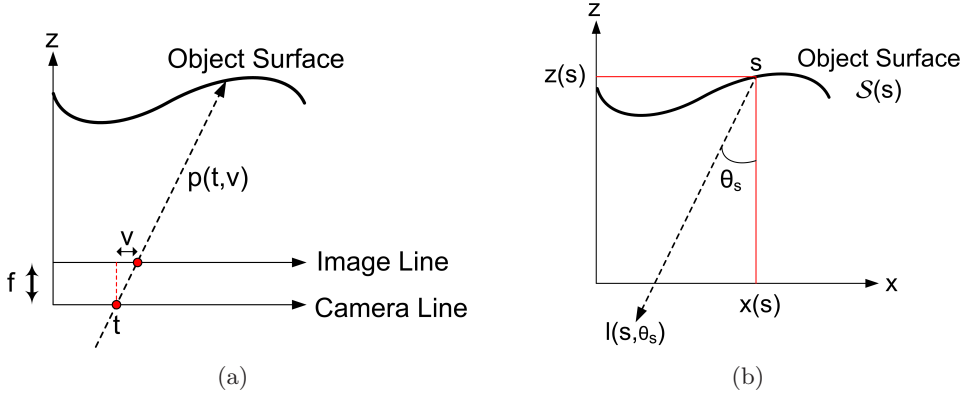


Figure 2.3: Diagram (a) illustrates the 2D light field, $p(t, v)$, in which a light ray is defined by its intersection with the camera line at a location t and the corresponding pixel location, v , on the image line. Diagram (b) illustrates the 2D surface light field, $l(s, \theta_s)$, in which a light ray is defined by its point of origin, s , on the object surface, S , at a viewing angle θ_s .

the remainder of this thesis we shall use (2.4) and (2.5) when analysing the light field and surface light field, respectively. Accordingly, in the 2D surface light field, we drop the s subscript from the viewing angle, referring to it only as θ .

2.2.2 Epipolar Plane Image

As covered in the last section, the 2D light field representation explicitly defines the intensity of a light ray captured at a coordinate (t, v) . Now, if we consider all possible (t, v) coordinates, a visual representation of the 2D light field can be constructed. This visual representation is known as the Epipolar Plane Image (EPI) [7] or EPI-volume if we are considering 2D images. It highlights how the inherent structure in the linear camera path leads to structure in the light field. For example a point at a depth z_0 , see Figure 2.4(a), is mapped to a line in the EPI with a slope that is inversely proportional to z_0 , see Figure 2.4(b). This structure leads to the following important characteristic: lines with steeper slopes will always occlude lines with shallower slopes in the EPI domain, in other words a point close to the camera will occlude those that are more distant. Feldmann et al [21] introduced the notion of Image Cube Trajectories in order to extend this visual tool to non-linear camera paths.

This concept of mapping a point in space to a line in the EPI can be generalised to

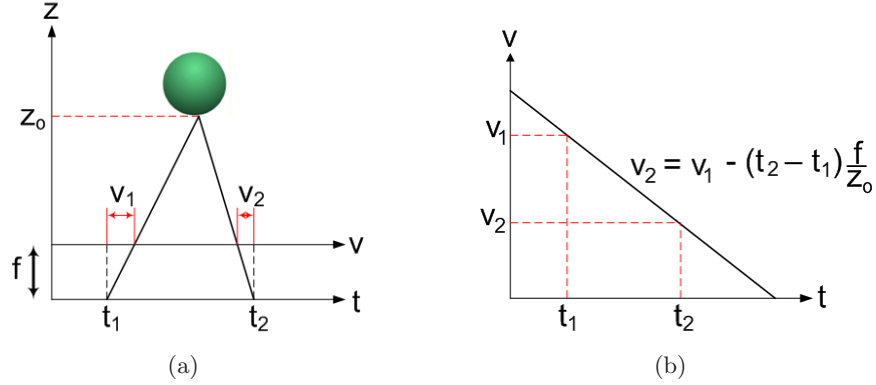


Figure 2.4: Diagram showing the 2D parametrisation of the light field (a), and its EPI representation, (b). A point in (a) translates to a line in the EPI with a slope inversely proportional to the depth of the point.

higher dimensional structures. For instance all the points at a certain depth will have the same gradient or trajectory in the EPI. Thus a region in space, made from neighbouring points, will result in a collection of trajectories in the EPI-volume [4]. Using this observation, [15] decomposes the scene into layers and groups the resulting trajectories, from each layer, into volumes called EPI-tubes. More generally, for higher dimensional plenoptic representations, these volumes or hypervolumes are termed plenoptic manifolds [5].

2.3 Plenoptic Sampling

The discussion so far has focused on the continuous plenoptic function (or light field) neglecting the constraints of the acquisition system (i.e. the cameras and the camera network). The most basic constraint is that we have a finite number of cameras N_c with a finite pixel resolution Δv (hence a finite number of pixels N_p). Consequently we only have access to samples of the plenoptic function with which to render new views. The quality of these rendered views depends on the how we reconstruct the continuous plenoptic function from its samples. Plenoptic sampling, therefore, focuses on maximising this reconstruction and in turn maximising the rendering quality. It comprises the following problems: determining the minimum number of cameras required to reconstruct the plenoptic function; deciding how to position the cameras in

order to maximise the reconstruction; and deciding how to combine the images such that the best rendering quality is achieved.

The following discussion examines the state-of-the-art in plenoptic sampling. In particular it focuses on uniform camera distribution in plenoptic sampling using a Fourier framework, the approaches to non-uniform camera distributions in plenoptic sampling and the reconstruction of the plenoptic function assuming either uniform or non-uniform sampling. Note that we assume the sampling within each camera is an intrinsic characteristic, which we cannot alter.

2.3.1 Uniform Sampling in a Fourier Framework

When the cameras are uniformly spaced, at a spacing Δt , it is natural to analyse plenoptic sampling within a classical Fourier framework. In such a framework, uniform sampling leads to spectral replication in frequency and the minimum sampling requirement - the Nyquist density - is such that the replicas do not overlap. If these replicas overlap then aliasing will occur. In plenoptic sampling, aliasing manifests itself as artefacts in the rendering process [10]. As a result the plenoptic sampling problem is reduced to determining the maximum camera spacing Δt , such that the replicated spectra do not overlap the original, and designing a reconstruction filter, $\Psi(\omega_t, \omega_v)$, that removes the replicas. With this in mind, we are interested in the properties of the plenoptic spectrum, the Fourier transform of the plenoptic function. In particular its spectral support which determines the Nyquist spatial sampling density.

The first spectral analysis of the plenoptic function was performed in [10]. Their analysis involved using the structure of the EPI to map each image to a reference position. Assuming a Lambertian scene¹ with no-occlusion, the intensity of a point in the scene, at a depth z_0 , is mapped to the reference image at $t = 0$ by $p(t, v) =$

¹The Lambertian assumption means that the intensity of a light ray leaving a point on the scene surface is independent of the angle the light ray leaves the surface. In other words the point looks the same from any viewing angle.

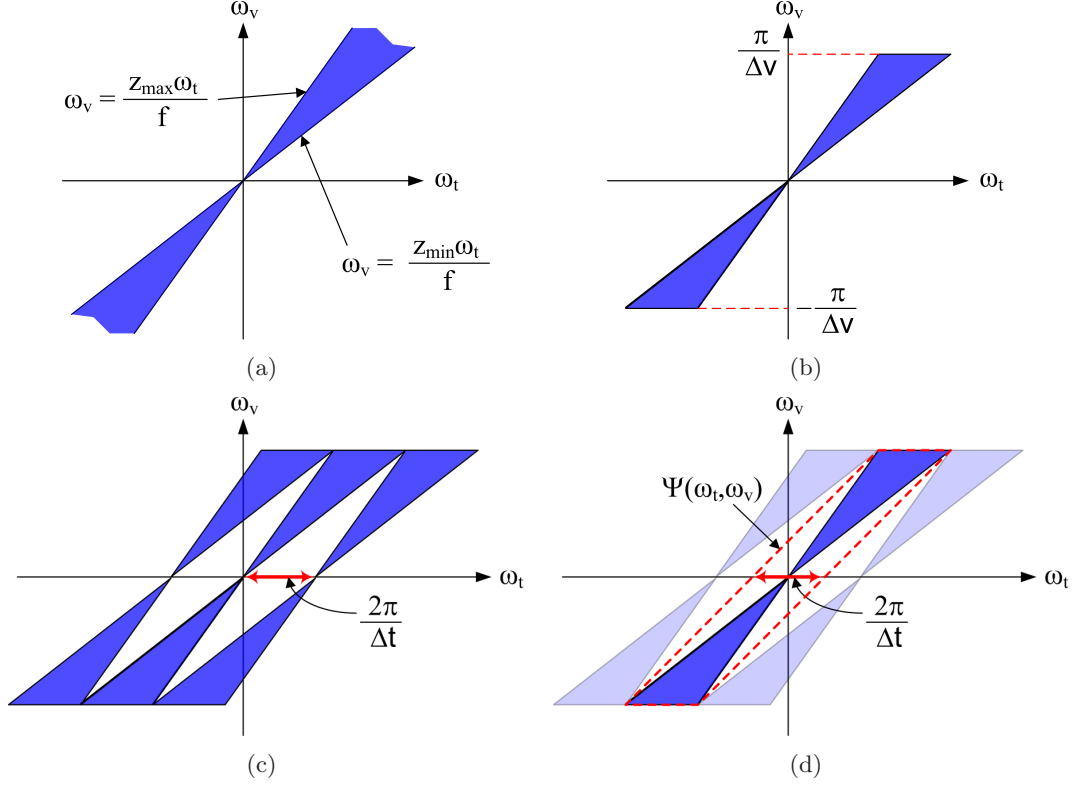


Figure 2.5: Diagrams of the plenoptic spectrum: (a) the plenoptic spectrum bounded between $\omega_v = \omega_t z_{\min}/f$ and $\omega_v = \omega_t z_{\max}/f$. (b) the ‘Bow-tie’ shaped plenoptic spectrum caused by the pixel resolution, Δv , inducing lowpass filtering in ω_v . (c) the optimal packing for the sampled plenoptic spectrum, where Δt is the camera spacing. (d) the reconstruction filter, $\Psi(\omega_t, \omega_v)$, required for perfect reconstruction.

$p(0, v - ft/z_0)$. Using this mapping, the plenoptic spectrum for the point at z_0 is

$$\begin{aligned}
 P(\omega_t, \omega_v) &= \mathcal{F}_{t,v} \{p(t, v)\} \\
 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(t, v) e^{-j(\omega_t t + \omega_v v)} dt dv \\
 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p\left(0, v - \frac{ft}{z_0}\right) e^{-j\omega_v v} e^{-j\left(\omega_t - \omega_v \frac{f}{z_0}\right)t} dt dv \\
 &= P(\omega_v) \delta\left(\omega_t - \omega_v \frac{f}{z_0}\right), \tag{2.6}
 \end{aligned}$$

where $P(\omega_v)$ is the Fourier transform of the reference image. Therefore the plenoptic spectrum is defined along the line $\omega_v = \omega_t z_0/f$.² Now, as pointed out by Chai et al [10], if the scene has a maximum depth, z_{\max} , and a minimum depth, z_{\min} , then

²Note that in [10] the line is perpendicular to this due to a difference in the direction of v relative to the camera line.

the plenoptic spectrum is approximately bounded between the lines $\omega_v = \omega_t z_{\min}/f$ and $\omega_v = \omega_t z_{\max}/f$, see Figure 2.5(a). This analysis, however, implicitly assumes the scene depth is approximately piecewise constant. This assumption is relaxed in [17, 68], allowing the authors to examine more complicated scenes.

Specifically, [17, 68] derived spectral properties for a broader range of scenes by exploiting the equivalence between the plenoptic function and the surface light field. This equivalence is formalised by modelling the scene with a functional surface. In this framework, the depth of the scene surface, relative to the real world coordinate x , is defined by the function $z(x)$ and its texture is modelled as a bandlimited signal, $g(s)$, where s is the curvilinear coordinate on the surface. Now, assuming the camera line t coincides with the x coordinate system, the authors link a light ray arriving at (t, v) to its point of origin on the surface at $(x, z(x))$ using the following geometric relationship

$$t = x - z(x) \tan(\theta) = x - z(x) \frac{v}{f}, \quad (2.7)$$

where f is the focal length and θ is the viewing angle. An illustration of this relationship is shown in Figure 2.6. Provided this geometric relationship is a one-to-one mapping, then the spatial position $(x, z(x))$ specifies a single curvilinear position s , which allows the plenoptic function to be mapped to the surface light field and vice versa. The provision of a one-to-one mapping is enforced in [17] by excluding scenes with occlusions. Therefore $z(x)$ is constrained such that

$$|z'(x)| < \frac{f}{v_m}, \quad (2.8)$$

where $z'(x)$ is the first derivative of z with respect to x , and v_m is the maximum value of v for a camera with a finite field of view, hence $v \in [-v_m, v_m]$. Although this constraint is not directly enforced in [68], a one-to-one relationship is achieved by selecting the closest point to the scene that satisfies (2.7).

Formally, the mapping between the plenoptic function and the surface light field is as follows. Equation (2.7) allows the mapping of the plenoptic function, $p(t, v)$,

to $l_x(x, \bar{v})$ the intensity of a light ray emitted from the spatial position $(x, z(x))$ at a viewing direction defined by $\bar{v} = v/f$, i.e.

$$l_x(x, \bar{v}) = p(x - z(x)\bar{v}, f\bar{v}). \quad (2.9)$$

The surface light field, $l(s, \theta)$, is then obtained by mapping the spatial position x to the curvilinear coordinate s and the viewing direction \bar{v} to the viewing angle θ , hence

$$l(s, \theta) = l_x(x(s), \bar{v}(\theta)). \quad (2.10)$$

The importance of this mapping is that spectral properties of the plenoptic function can be derived by assuming properties of the surface light field without explicitly defining the scene's geometry. Therefore [17] derives the plenoptic spectrum in terms of $l_x(x, \bar{v})$ and determines spectral properties based on its behaviour.

The plenoptic spectrum in question is obtained as follows: starting from its definition,

$$\begin{aligned} P(\omega_t, \omega_v) &= \mathcal{F}_{t,v} \{p(t, v)\} \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(t, v) e^{-j(\omega_t t + \omega_v v)} dt dv, \end{aligned} \quad (2.11)$$

both integration variables are changed using (2.7) and $\bar{v} = \tan(\theta) = v/f$, which results in a Jacobian of $(1 - z'(x)\bar{v})f$. Consequently the following is obtained

$$\begin{aligned} P(\omega_t, \omega_v) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(x - z(x)\bar{v}, f\bar{v}) e^{-j(\omega_t(x - z(x)\bar{v}) + \omega_v f\bar{v})} (1 - z'(x)\bar{v}) f dx d\bar{v} \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} l_x(x, \bar{v}) e^{-j(\omega_t(x - z(x)\bar{v}) + \omega_v f\bar{v})} (1 - z'(x)\bar{v}) f dx d\bar{v}. \end{aligned} \quad (2.12)$$

At this point two identities are introduced, the first is $h(x, \bar{v}) = l_x(x, \bar{v})(1 - z'(x)\bar{v})$ and the second is $L_x(x, \omega_v) = \mathcal{F}_{\bar{v}} \{l_x(x, \bar{v})\}$. Using these identities, the integral in \bar{v}

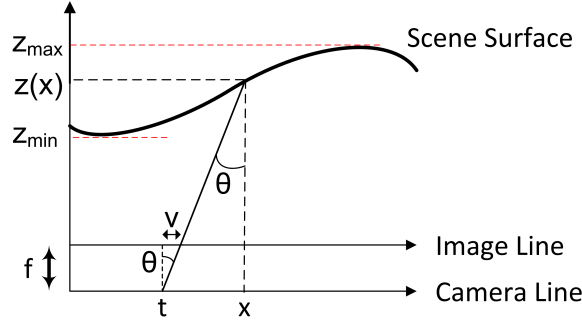


Figure 2.6: General scene model showing the intersection of a light ray (t, v) with the scene surface at $(x, z(x))$, where $z(x) \in [z_{\min}, z_{\max}]$ is the depth of the surface. Note that f is the focal length of the cameras and θ is the viewing angle.

becomes

$$\begin{aligned} H(x, \omega_v) &= \mathcal{F}_{\bar{v}} \{h(x, \bar{v})\} = \int_{-\infty}^{\infty} h(x, \bar{v}) e^{-j\omega_v \bar{v}} d\bar{v} \\ &= L_x(x, \omega_v) - jz'(x) \frac{\partial L_x(x, \omega_v)}{\partial \omega_v}. \end{aligned} \quad (2.13)$$

Finally, inserting the above into (2.12), we obtain a general equation for the plenoptic spectrum that is independent of the scene's geometry:

$$P(\omega_t, \omega_v) = f \int_{-\infty}^{\infty} H(x, \omega_v f - z(x)\omega_t) e^{-j\omega_t x} dx. \quad (2.14)$$

The first point to take from this equation is the dependency of the plenoptic spectrum on the slope of the surface $z'(x)$. A dependency that was not apparent in (2.6). Secondly, by using the fact that $l_x(x, \bar{v}) = l_x(x)$ for a Lambertian scene, [17] showed that in frequency the following is true

$$L_x(x, \omega_v) = 0, \quad \text{if } \omega_v \neq 0, \quad (2.15)$$

which leads to

$$P(\omega_t, \omega_v) = 0, \quad \text{if } \omega_v f - z(x)\omega_t \neq 0. \quad (2.16)$$

Therefore, as $z(x) \in [z_{\min}, z_{\max}]$, the plenoptic spectrum is precisely bounded by lines relating to the maximum and minimum depths of the scene, see Figure 2.5(a). More-

over, they formalised bounds for non-Lambertian scenes by assuming $l_x(x, \bar{v})$ is bandlimited in \bar{v} to B_L . Consequently relaxing the Lambertian assumption on the scene results in an extended region around the plenoptic spectrum, a fact also highlighted in [68, 70].

When considering the continuous plenoptic function, without regard to the acquisition network, the bounded spectrum defined above is band-unlimited unless the scene surface is flat [17]. However the finite resolution of the acquisition devices induce lowpass filtering in ω_v at $\pi/\Delta v$, which in turn induces filtering in ω_t . Therefore the plenoptic spectrum is bandlimited in both ω_v and ω_t resulting in a ‘bow-tie’ shape, as shown in Figure 2.5(b). Based on this shape, the optimal packing of the replicated spectra at critical sampling is shown in Figure 2.5(c). To achieve this packing, without overlap occurring, [10] derive the following maximum camera spacing

$$\Delta t_C = \frac{2\pi}{\Omega_v f \left(\frac{1}{z_{\min}} - \frac{1}{z_{\max}} \right)}, \quad (2.17)$$

where Ω_v is the maximum frequency in ω_v (in the worse case it is equal to $\pi/\Delta v$). In terms of the reconstruction filter $\Psi(\omega_t, \omega_v)$, a slanted rectangular window is required to remove the replicated spectra, see Figure 2.5(d). The support of this reconstruction filter in the frequency domain is

$$\mathcal{R}_\psi = \left\{ \omega_t, \omega_v : \omega_v \in [-\Omega_v, \Omega_v], \omega_t \in \left[\frac{\omega_v f}{z_C} - \frac{\pi}{\Delta t}, \frac{\omega_v f}{z_C} + \frac{\pi}{\Delta t} \right] \right\}, \quad (2.18)$$

where the slant of the filter is determined by f and z_C , which is defined as

$$\frac{1}{z_C} = \frac{1}{2} \left(\frac{1}{z_{\max}} + \frac{1}{z_{\min}} \right). \quad (2.19)$$

It is worth noting that (2.17) and (2.19) are solutions to the uniform plenoptic sampling problem but only for the case of a Lambertian scene with no occlusion. Bearing this in mind, several researchers have extended the spectral analysis of the plenoptic function to more general situations. As noted earlier, [17, 68, 70] relaxed the Lambertian

assumption on the scene surface and [68, 70] examined scenes with occlusions. In both situations the spectral support of the plenoptic function increased, hence decreasing the camera spacing. However the structure of the reconstruction filter is not altered [70]. Finally, using the mapping introduced in [10], [13] examined the effect on the plenoptic spectrum when the camera path is varied.

2.3.2 Non-Uniform Sampling

In general uniform plenoptic sampling is most efficient when the scene in question is relatively constant (in either depth or texture), however, in reality, this cannot be guaranteed. In order to avoid undersampling, the camera spacing becomes conservative and is determined by the largest depth and the largest texture variation in the scene. The problem of such an approach is that it may limit the sampling efficiency in certain scenes [70]. For example if the scene is relatively constant, with only a small section containing large variation, then uniform sampling will result in the majority of the scene being oversampled.

The solution is to allow irregular or non-uniform camera placement that depends on the scene (termed free-form sampling in [70]). The difficulty with sampling the plenoptic function in such a way is that there are many possible camera configurations to choose from. Therefore more constraints are required to determine the camera positions. Based on the heuristics used to position the cameras, non-uniform plenoptic sampling can be split into three categories; sample reduction (SR), active incremental capturing (AIC) and active rearranged capturing (ARC) [70].

In SR an initial image set is generated by uniformly oversampling the scene. This image set is then minimised based on some quantitative criteria. For example, using a mesh model of the scene, [22] ranks each image based on the proportion of scene elements covered. The reduced image set is then generated by selecting the images with the highest rank (i.e. those that cover the highest proportion of scene elements). In a similar fashion, [43] filters the initial image set based on an analysis of the scene geometry. Another, slightly different approach, involves mapping the plenoptic function to a different sampling matrix using multi-dimensional lattice theory [70]. Thus redun-

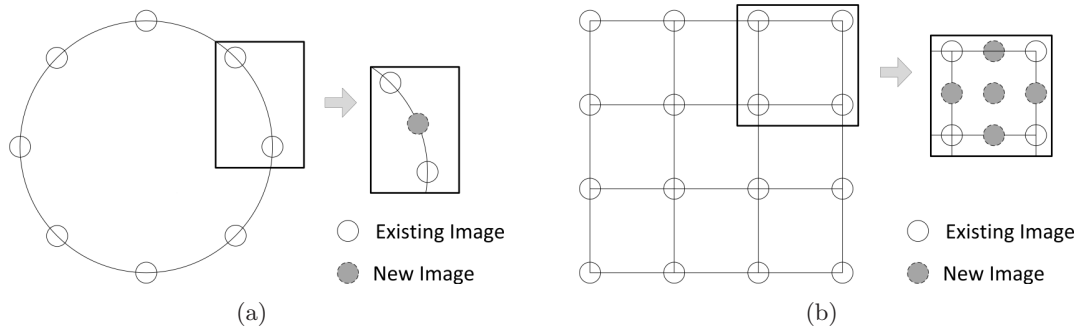


Figure 2.7: Diagrams illustrating how new images are introduced when using AIC to non-uniformly sample the scene. In (a) a single image is introduced in a concentric mosaics representation [67] and in (b) multiple images are introduced in a 4D light field representation [70]. A new image is inserted between an image pair based on the reconstruction error.

dancy in the image set is removed by non-rectangular down-sampling of the plenoptic function. However, the main limitation of SR is that it initially requires a large number of images to ensure the scene is oversampled and then discard those that are deemed redundant.

A valid alternative is to use AIC to non-uniformly sample the scene. In contrast to SR, the scene is initially undersampled uniformly and intermediate samples are added to reduce a local reconstruction error. This is repeated until the average local reconstruction error is below a threshold. Figure 2.7 illustrates how new samples are introduced in a concentric mosaic representation [67] and in a 4D light field representation [70]. A systematic framework for AIC was presented in [67] using the position-interval error (PIE) as a measure of the average reconstruction error for any pair of samples. Thus the goal is to have a uniform PIE for each pair of samples. A good estimate of the PIE is the local colour consistency [67]. Another example of AIC is [47], in which an adaptive mesh is used to define the camera positions. New camera positions are introduced on the edges of the mesh in order to reduce the estimated reconstruction error between the nearest source images.

The final category in non-uniform plenoptic sampling is ARC. It involves repositioning a limited set of cameras in order to improve rendering quality on the fly. The main advantage of repositioning the cameras on the fly is that ARC can sample and render dynamic scenes. In contrast both SR and AIC methods require a two stage

sampling process, a uniform stage followed by a selective stage, consequently both are restricted to static scenes. An example of ARC is presented in [71], where N_c cameras are used to render N_v views, assuming $N_v > N_c$. The new camera positions are determined by minimising the sum of the squared rendering errors. This minimisation is solved iteratively using the local colour consistency as an estimate of the rendering error. The cameras are initialised on a uniform grid and then progressively moved to new positions based on the minimisation, hence allowing the system to be applied to dynamic scenes.

2.3.3 Reconstructing the Sampled Plenoptic Function

In the Section 2.3.1, we presented a filter for uniform reconstruction of the plenoptic function in the frequency domain. A drawback of such a filter however is that it has a fixed skew equal to z_C/f . This is equivalent to assuming the scene has a constant depth at z_C and rendering all images accordingly. To overcome this issue [59] propose reconstructing the plenoptic function using a set of different filter skews and then determining the optimum for each pixel. In other words they render at several depths and fuse the resulting images.

Alternatively, a common approach in both uniform and non-uniform sampling is to simplify the reconstruction filter to a local interpolation [9, 14, 27, 37, 38, 58]. Using this method, a point in the plenoptic function is calculated through the weighted interpolation of the K nearest samples captured by the cameras. If no depth information is available, the interpolation is performed using the K neighbouring samples of the plenoptic function. However, unless the multi-view image set is very dense, these neighbouring samples will correspond to different points in the scene leading to rendering artefacts occurring in the reconstruction. With access to depth information, the structure of the EPI can be used to aid the interpolation and improve the reconstruction. The depth information allows the local interpolation to be performed along the EPI lines. Therefore, if the depth information is correct, the K samples used in the interpolation will correspond to the same point in the scene. For example, if $K = 2$, then a 2D EPI (or 3D EPI-volume) is reconstructed using a linear interpolation along the EPI

line. In [27, 37], this linear interpolation is extended to a quadrilinear interpolation for a 4D light field. In view of this, a simple reconstruction method for a 2D EPI (or 3D EPI-volume) is to assume a rough depth model, such as a plane at a constant depth z_0 , and then perform linear interpolation along the EPI lines using a gradient of f/z_0 .

Another approach, used in [70, 71], is to consider the local interpolation in terms of light rays. Therefore, the interpolated light ray is computed using the K nearest light rays capture by the cameras. Assuming a rough depth model of the scene, these K light rays are chosen based on the angular difference between the captured light ray and the interpolated light ray. Having determined these K light rays, the inverse of their angular difference is used as the interpolation weights. Note that the weights are normalised to ensure they sum to 1. The weights can also be design to incorporate other aspects such as finite field of view and camera resolution [70].

Finally, in [58] the linear interpolation between two adjacent stereo images is analysed assuming inaccurate depth information. Assuming a probabilistic distribution of depth errors, the authors present an optimised interpolation using two pixel-varying filters, termed combining filters. A similar approach is presented in [9] for multiple unstructured input images.

2.4 Depth Information in IBR

An interesting result from the spectral analysis in Section 2.3.1 is that the number of images required to reconstruct the plenoptic function depends on aspects of the scene depth. This interplay between the number of images required and the amount of depth information available was examined in [10]. The authors observed that the maximum camera spacing, defined in (2.17), depends on knowledge of the maximum and minimum depths of the scene (i.e. the depth variation). Now if more depth information is available then the scene can be decomposed into multiple depth layers and sampled separately. This is equivalent to sampling multiple scenes with smaller depth variation thus reducing the overall number of images required. Therefore, using

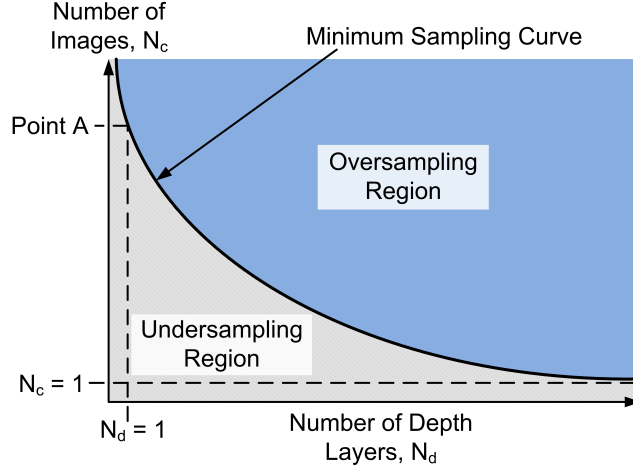


Figure 2.8: Diagram showing the number of images, N_c , required in IBR as a function of the depth layers, N_d , available. The minimum sampling curve, proposed in [10], is determined from (2.20). Point A in the diagram marks the number of images required if $N_d = 1$, it is determined using Equation (2.17).

depth layers, the maximum uniform spacing between adjacent cameras becomes

$$\Delta t_{C,d} = \frac{2\pi N_d}{\Omega_v f \left(\frac{1}{z_{\min}} - \frac{1}{z_{\max}} \right)}, \quad (2.20)$$

where $N_d \geq 1$ is the number of depth layers. From this expression, [10] constructed a minimum sampling curve in terms of the number of images required against the number of depth layers available, see Figure 2.8. Operating at a point above this line will result in redundancy (i.e. oversampling the plenoptic function), whereas undersampling occurs if you operate at a point below the line.

This trade-off between the number of images required and the amount of depth information available has been described previously in [11, 34, 50, 52, 69]. For the purpose of this thesis, however, IBR techniques involving depth information are roughly classified into three groups according to the amount of depth information used.

The first group comprises techniques that use a limited amount of depth information to render new views. An example is the uniform plenoptic sampling highlighted in Section 2.3.1, which requires only the maximum and minimum depths of the scene. Using slightly more depth information, techniques, such as the pop-up light field [53] and IBR objects [25], assign a single depth value to coherent segments of the image set.

This is equivalent to approximating the scene with a piecewise constant depth model. In a similar fashion, [5] uses a finite number of depth layers to decompose the plenoptic function into a series of plenoptic manifolds. Limited depth information is also present in the lumigraph [27], and unstructured lumigraph [9], in the form of a rough depth model.

In contrast to the first group, IBR techniques in the second classification assume that detailed depth information is available. In this context, detailed depth information means that each view in the image set has a corresponding depth map i.e. a per-pixel depth as well as a per-pixel colour. Consequently new views are rendered in two steps; in the first step the existing images are mapped to the new viewpoint and then, in the second step, the new image is generated by blending the warped images [18]. This process of rendering new views using per-pixel depth is known as depth IBR (DIBR), a term coined in [19, 20]. Recently many DIBR techniques have been proposed such as [18, 42, 54, 55, 73, 74]. In view of this, the authors in [44] analysed the rendering quality obtained from DIBR. Their analysis focused on the distortion in an arbitrary view caused by the IBR configuration, for example errors in the depth and intensity information.

Finally, the last classification covers IBR techniques that use 3D scene models. For example, the layered depth image (LDI) [49], and its extension, LDI trees [12], construct a geometric representation of the scene using a multi-valued image. The image records all the colour and depth information that exists along the line of sight of each pixel. Also, by definition, this last group includes model-based rendering (MBR) techniques, such as [66].

In general the depth information in question is computed using multi-view stereo vision algorithms [42, 48]. The disadvantage of such algorithms is that they are computationally intensive and prone to inaccuracies [18, 34]. However with the recent increase in low-cost reliable depth cameras [32], such as Microsoft's Xbox Kinect, this depth information can be captured on-line without the additional computation. This paves the way for multi-view systems containing both colour and depth cameras. For a state-of-the-art on depth camera technology and active 3D scene capture see [23, 32, 57].

2.5 Summary

IBR can be posed as the problem of sampling and reconstructing the plenoptic function. The multi-view image set, in this scenario, represents the samples of the plenoptic function and the rendering of a new viewpoint its reconstruction. A problem, however, is that a large number of images are required to sample the plenoptic function. If too few images are available, the plenoptic function is undersampled leading to artefacts when rendering new views. Accordingly, the number of images required for artefact-free rendering, and their optimum positioning, can be determined through sampling analysis of the plenoptic function. In this chapter, we have presented the state-of-the-art in plenoptic sampling analysis, in particular focusing on the optimum sampling efficiency in both a uniform and non-uniform framework. We also examined the use of depth information in IBR and how it can be used to compensate a reduction in the number of images. A summary of the key points are as follows:

- In this thesis we shall use the 2D light field and 2D surface light field representations of the plenoptic function.
- Assuming uniform camera distribution, the optimal camera spacing and reconstruction filter are determined through spectral analysis of the plenoptic function.
- The plenoptic spectrum of a Lambertian surface is precisely bounded between lines relating to the maximum and minimum depths of the scene.
- In general, maximum sampling efficiency requires non-uniform camera placement. This is divided into three approaches; sample reduction (SR), active incremental capturing (AIC) and active rearranged capturing (ARC).
- Additional depth information can either be used to improve the reconstruction of the plenoptic function or to reduce the number of images without altering the quality.
- The additional depth information is either computed off-line using multi-view stereo algorithms or captured using depth cameras.

Chapter 3

Plenoptic Spectral Analysis

3.1 Introduction

In this chapter we re-examine the spectral analysis of the plenoptic function incorporating two realistic conditions: finite scene width (FSW) and cameras with finite field of view (FFoV). The novelty of this approach is that we are able to analyse the plenoptic spectrum for a larger class of scene geometry. In contrast, the spectral analysis reviewed in Chapter 2 only uses FFoV in order to impose the no-occlusion constraint (2.8) on the scene surface. It is not applied to the actual spectral analysis. Consequently, the resulting analysis is only valid for scenes without depth variation.

Using these finite conditions, we analyse the plenoptic spectrum for a 1D slanted plane (a simplification of a 2D planar facet for the 2D light field and 2D surface light field parametrisations). The appeal of such a scene is two fold: first we are able to derive exact spectral properties of its plenoptic spectrum. Second, it can be used as a basis to construct more complicated scenes. Therefore we can apply knowledge gained from the slanted plane to sample the plenoptic functions relating to more complicated scenes.

It is worth noting that a spectral analysis of a slanted plane under FFoV has previously been analysed in [68]. This analysis, however, did not incorporate any constraint on the width of the plane. As a result the derived plenoptic spectrum was independent of ω_v . Thus, to bound the spectrum in ω_v , the authors assumed the approximate

spectral bounds presented in [10], i.e. the spectral bounds relating to the maximum and minimum depths of the scene outlined in Section 2.3.1. In contrast, our analysis results in an exact expression for the spectrum hence allowing greater understanding of its structure and behaviour as scene parameters alter.

In this chapter we start by formalising the scene geometry for a slanted plane in Section 3.2. We then, in Section 3.3, examine the effects of F FoV and FSW on the plenoptic spectrum for a slanted plane. In particular, we derive an exact closed-form expression for the plenoptic spectrum of a Lambertian slanted plane with complex exponential texture. This spectral analysis is then extended to scenes comprising multiple slanted planes. Using this closed-form expression, Section 3.4 examines in detail the behaviour of the plenoptic spectrum for a slanted plane and determines a characteristic structural model for the spectrum. In Section 3.5, as an alternative to altering the scene, we examine the effects of rotating the camera line about a point and derive the resulting plenoptic spectrum. We end the chapter with a summary of key points in Section 3.6.

3.2 Scene Geometry for a Slanted Plane

In this thesis, the geometry of a scene is modelled using the framework presented in [17, 68]. This framework uses functional surfaces and textures to model the scene. If we assume a horizontal slice of the scene, at a fixed height y_0 , then the framework comprises three functions: $z(x)$ the depth of the scene relative to x , $x(s)$ the mapping of the curvilinear coordinate s onto x , and $g(s)$ the texture signal pasted to the scene surface. Collectively, we shall term these the scene geometry equations \mathcal{G}_s .

For the slanted plane, illustrated in Figure 3.1, the functions $z(x)$ and $x(s)$ are determined using trigonometry, and the texture signal is assumed to be bandlimited.

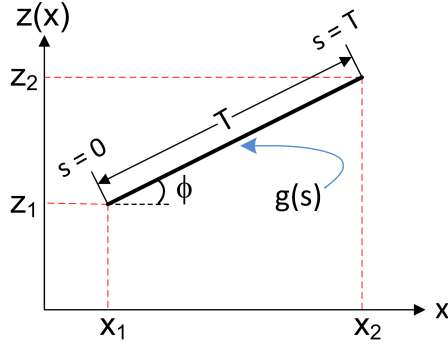


Figure 3.1: Diagram illustrating the scene geometry for a slanted plane with a texture signal, $g(s)$, pasted to the surface. The plane is dictated by three elements; the slant of the plane, ϕ , the width of the plane, T , and the starting position of the surface, (x_1, z_1) . The texture signal is defined in terms of a curvilinear coordinate $s \in [0, T]$. Notice that if $\phi > 0$ then $z_2 > z_1$, however, if $\phi < 0$ then $z_2 < z_1$.

Consequently the scene geometry equations are

$$\mathcal{G}_s = \begin{cases} z(x) = (x - x_1) \tan(\phi) + z_1 \\ x(s) = s \cos(\phi) + x_1 \text{ for } s \in [0, T] \\ g(s) \xrightarrow{\mathcal{F}} G(\omega) \text{ and } G(\omega) = 0 \text{ for } |\omega| > \omega_s \end{cases} \quad (3.1)$$

where ω_s is the maximum frequency of the texture signal, ϕ is the angle between the plane and the x -axis and T is the length of the plane. The spatial coordinate (x_1, z_1) indicates the starting point of the plane, in other words the origin for the curvilinear coordinate s , and, due to FSW, (x_2, z_2) indicates the end point of the plane. At (x_2, z_2) the curvilinear coordinate is equal to the width of the plane, i.e. $s = T$, which leads to the following relationship

$$T = \frac{x_2 - x_1}{\cos(\phi)} = \frac{z_2 - z_1}{\sin(\phi)}. \quad (3.2)$$

Notice that this model of the slanted plane allows both positive and negative values of ϕ to occur, leading to the following

$$\begin{aligned} z_1 < z_2 & \text{ if } \phi > 0, \\ z_1 > z_2 & \text{ if } \phi < 0, \\ z_1 = z_2 & \text{ if } \phi = 0. \end{aligned}$$

The absolute magnitude of ϕ , however, is restricted as we impose the no-occlusion constraint from [17] on the scene. The constraint, defined in (2.8), limits the magnitude of the slope of the surface, $z'(x)$, which is equal to $\tan(\phi)$ for a slanted plane. The resulting bound on ϕ is

$$|\phi| < \tan^{-1} \left(\frac{f}{v_m} \right). \quad (3.3)$$

In the discussion so far we have only considered a scene comprising a single object, a single slanted plane. However, in general, a scene can consist of an arbitrary number of such objects, each separate from one another. For such a scene, the imposition of the no-occlusion constraint results in the objects being sufficiently far apart to avoid occluding each other. Consequently we can treat each object independently and determine their plenoptic spectra separately. Therefore, the plenoptic spectrum of a scene, comprising multiple separate objects, is the sum of the individual spectra corresponding to each object. In view of this, we shall focus on scenes that only comprise a single finite object. This finite object can comprise multiple slanted planes, but its surface will be continuous. In practise, this constraint is restrictive as most real scenes contain occlusions. However, Zhang and Chen [68] proposed a method for dealing with occlusions and extended the framework in Section 2.3.1 to cover such scenes. Their method involved modelling occlusions as masks applied to the plenoptic function and using the properties of the Fourier transform to determine the plenoptic spectrum. Therefore, using a similar method, we could extend any results we derive to scenes with occlusions.

3.3 Spectral Analysis under FFoV and FSW

To illustrate the effects of FFoV and FSW, we first examine the simpler case when $\phi = 0$, i.e. a fronto-parallel plane. We initially assume that the scene is not constrained by

FFoV or FSW, thus the scene geometry equations are

$$\mathcal{G}_{FPP} = \begin{cases} z(x) = z_c \\ x(s) = s \\ g(s) \xrightarrow{\mathcal{F}} G(\omega) \text{ and } G(\omega) = 0 \text{ for } |\omega| > \omega_s \end{cases} \quad (3.4)$$

where z_c is a constant depth. An example of a synthetic EPI generated by an unconstrained fronto-parallel plane is illustrated in Figure 3.2(a). As highlighted in this example, we also assume that the surface is Lambertian. This assumption, as noted earlier, means that $l_x(x, \bar{v}) = l_x(x)$ and, when applied to (2.10), results in the following

$$l_x(x(s)) = l(s) = g(s). \quad (3.5)$$

Having defined the scene, its plenoptic spectrum is determined by introducing (3.4) and (3.5) into the general equation defined in (2.12). Therefore the spectrum is

$$\begin{aligned} P_{FPP}(\omega_t, \omega_v) &= \int_{-\infty}^{\infty} l_x(x) e^{-j\omega_t x} \int_{-\infty}^{\infty} f(1 - z'(x)\bar{v}) e^{-j(\omega_v f + z(x)\omega_t)\bar{v}} d\bar{v} dx \\ &= \int_{-\infty}^{\infty} g(s) e^{-j\omega_t s} ds \int_{-\infty}^{\infty} f e^{-j(\omega_v f + z_c\omega_t)\bar{v}} d\bar{v}, \end{aligned} \quad (3.6)$$

since $z'(x) = 0$. By solving (3.6) we arrive at the result first presented in [10], namely that the spectrum is a line in the (ω_t, ω_v) -domain, given by

$$P_{FPP}(\omega_t, \omega_v) = G(\omega_t) \delta\left(\omega_v - \frac{z_c \omega_t}{f}\right). \quad (3.7)$$

This type of plenoptic spectrum is shown in Figure 3.3(a). It is the corresponding plenoptic spectrum for the synthetic EPI illustrated in Figure 3.2(a).

Now if we constrain the cameras to a FFoV, such that $\bar{v} \in [-\bar{v}_m, \bar{v}_m]$ and $\bar{v}_m = v_m/f$, then the second integral in (3.6) has a finite range. Intuitively, this finite range can be viewed as applying a rectangular window function in the EPI domain. The effect of this windowing in the EPI domain is illustrated in Figure 3.2(b) using the previous synthetic EPI from part (a). The integral in \bar{v} , therefore, is the Fourier transform of

the window function, hence the spectrum becomes

$$P_{FPP}(\omega_t, \omega_v) = 2v_m G(\omega_t) \operatorname{sinc} \left(\omega_v v_m - \omega_t \frac{z_c v_m}{f} \right), \quad (3.8)$$

where the sinc function¹ is the Fourier transform of the window function. Consequently the FFoV constraint results in spectral spreading along the ω_v -axis. Figure 3.3(b) illustrates the extent of this spectral spreading on the plenoptic spectrum corresponding to the EPI in Figure 3.2(b).

A similar result occurs if we constrain the scene to be of finite width T , such that $s \in [0, T]$. In this case it is the first integral in (3.6) that has a finite range, which can be expressed as another rectangular window function. Using the same synthetic EPI, this new windowing in the EPI domain is shown in Figure 3.2(c). As a result, the integral in s is the Fourier transform of the product of the two functions, the window and the texture signal, which leads to

$$P_{FPP}(\omega_t, \omega_v) = T \left(G(\omega_t) * \operatorname{sinc} \left(\frac{\omega_t T}{2} \right) e^{-j\omega_t \frac{T}{2}} \right) \delta \left(\omega_v - \frac{z_c \omega_t}{f} \right), \quad (3.9)$$

where $*$ is the convolution operator. This time the spectral spreading in the plenoptic spectrum is along the line $\omega_v = \omega_t z_c / f$. The corresponding spectral illustration of this effect is shown in Figure 3.3(c).

Finally if we apply both constraints, as illustrated in Figure 3.2(d), then the corresponding plenoptic spectrum is a combination of the two previous cases, so we obtain

$$P_{FPP}(\omega_t, \omega_v) = 2v_m T \operatorname{sinc} \left(\omega_v v_m - \omega_t \frac{z_c v_m}{f} \right) \left(G(\omega_t) * \operatorname{sinc} \left(\frac{\omega_t T}{2} \right) e^{-j\omega_t \frac{T}{2}} \right). \quad (3.10)$$

Similar to the previous cases, the corresponding spectral illustration of (3.10) is shown in Figure 3.3(d). The importance of this analysis is that, even assuming a constant depth and Lambertian surface, the FFoV and FSW constraints introduce spectral spreading in the frequency domain, which lead to a band-unlimited plenoptic spectrum. Consequently, if we now introduce depth variation by analysing a slanted plane,

¹We use the definition $\operatorname{sinc}(h) = \sin(h)/h$ when $h \in \mathbb{R}$.

we can expect the finite constraints to result in similar spectral spreading. In other words its plenoptic spectrum will be band-unlimited as well.

3.3.1 Analysis of a Slanted Plane

Using the geometry defined in (3.1), we will now examine the plenoptic spectrum for a slanted plane assuming FFoV and FSW. To determine an exact expression for this spectrum, we need to specify the bandlimited signal, $g(s)$, used to model the scene's texture. For example, in [17,68], the authors use a sinusoidal signal comprising a single sine wave as texture in their analysis. This type of texture signal, however, can be decomposed into complex exponentials. Therefore, due to the linearity of the Fourier transform, the corresponding plenoptic spectrum is the sum of the individual spectra relating to each complex exponential. In view of this, we assume that the texture signal is a complex exponential, $g(s) = e^{j\omega_s s}$. Under this assumption, we are able to determine an exact closed-form expression for the plenoptic spectrum, which can then be extended to more complicated texture signals using the linearity of the Fourier transform. The key stages of its derivation are outlined below and the full derivation is included as Appendix A.

Similar to the fronto-parallel case, we start with the general equation in (2.12). From this we assume a Lambertian surface and then apply the FSW and FFoV constraints, to obtain

$$P(\omega_t, \omega_v) = \int_{x=x_1}^{x=x_2} l_x(x) e^{-j\omega_t x} \int_{\bar{v}=-\bar{v}_m}^{\bar{v}=\bar{v}_m} (1 - z'(x)\bar{v}) f e^{-j(\omega_v f - z(x)\omega_t)\bar{v}} d\bar{v} dx. \quad (3.11)$$

In contrast to the case of the fronto-parallel plane, (3.6), these integrals are no longer separable as $z(x)$ is not a constant and $z'(x) = \tan(\phi)$. Therefore we solve the integral in \bar{v} first, to obtain

$$P(\omega_t, \omega_v) = 2v_m \int_{x_1}^{x_2} l_x(x) \left(\text{sinc}(\omega_I) - j \frac{z'(x)v_m}{f} \text{sinc}'(\omega_I) \right) e^{-j\omega_t x} dx, \quad (3.12)$$

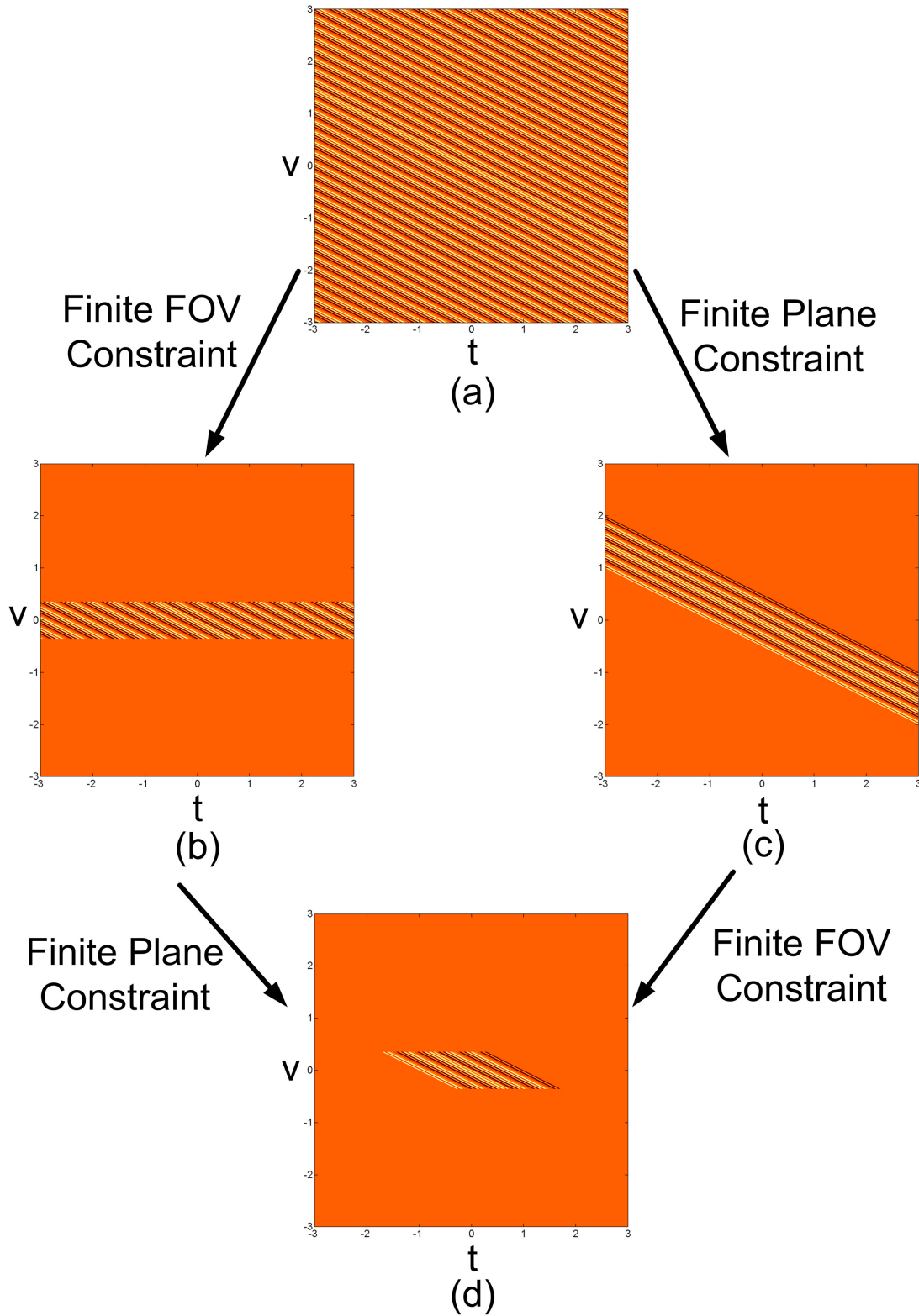


Figure 3.2: Diagram illustrating the windowing effects of FSW and FFoV in the EPI domain for a synthetic EPI with bandlimited texture. The constraints are as follows; (a) unconstrained EPI, (b) only FFoV, (c) only FSW, and (d) both FFoV and FSW.

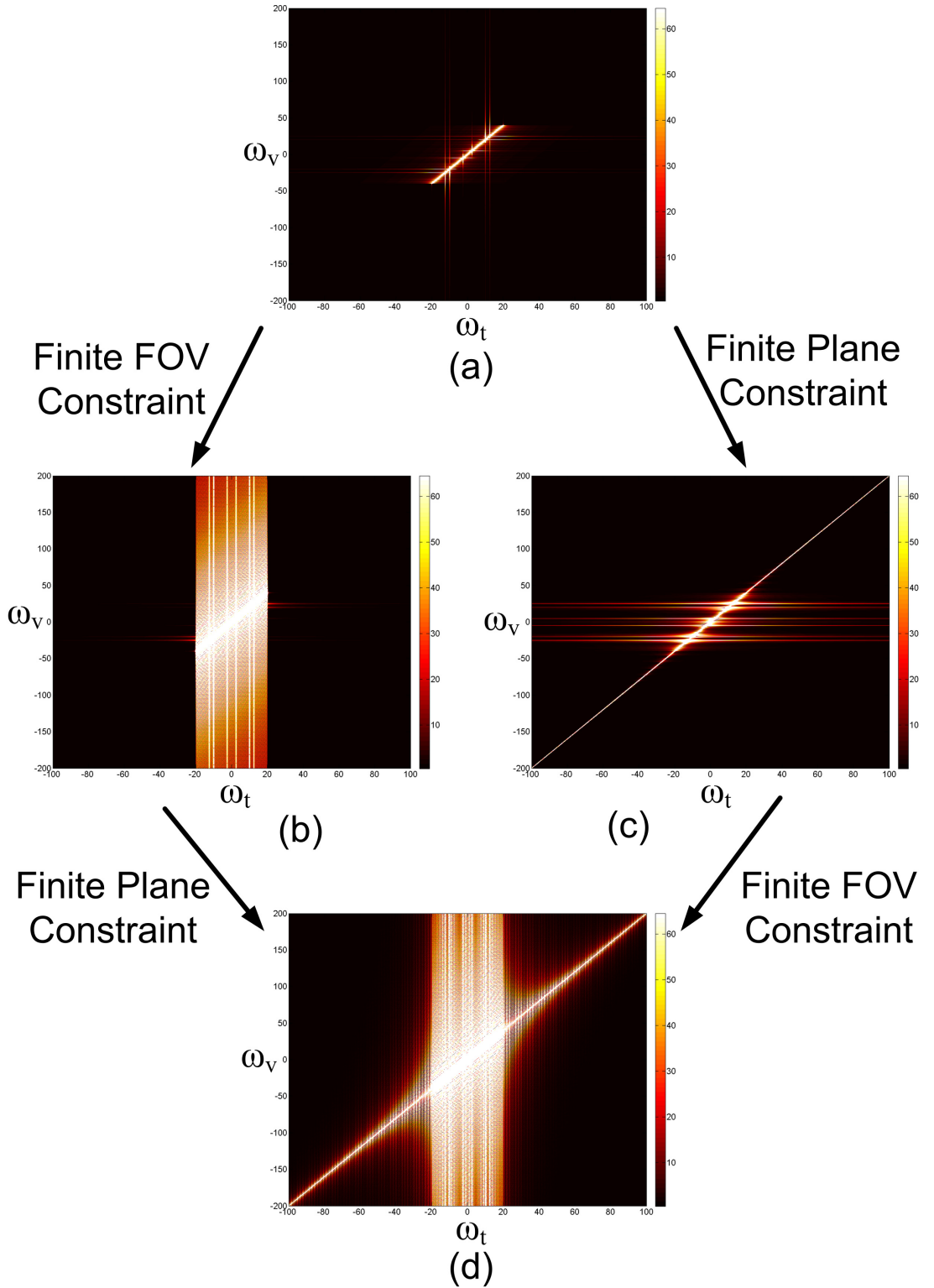


Figure 3.3: Diagram illustrating the windowing effects of FSW and FFoV on the plenoptic spectrum for a synthetic EPI with bandlimited texture. The constraints are as follows; (a) unconstrained EPI, (b) only FFoV, (c) only FSW, and (d) both FFoV and FSW. The corresponding EPI's are shown in Figure 3.2.

where $\text{sinc}'(\omega_I)$ is the first derivative of the sinc function with respect to its argument², and ω_I depends on x as follows

$$\omega_I = \omega_v v_m - z(x) \frac{\omega_t v_m}{f}.$$

At the moment, however, (3.12) defines the plenoptic spectrum for a general Lambertian scene with no occlusion, in other words it is independent of scene geometry. The expression illustrates the complex nature of the finite constraints when analysing scenes with depth variation. For the specific case of a slanted plane with complex exponential texture, defined in (3.1), the equation becomes

$$P_S(\omega_t, \omega_v) = M_1 \int_0^T g(s) \left[\text{sinc}(\hat{\omega}_I) - j \frac{v_m \tan(\phi)}{f} \text{sinc}'(\hat{\omega}_I) \right] e^{-j\omega_t \cos(\phi)s} \cos(\phi) ds \quad (3.13)$$

where $M_1 = 2v_m e^{-j\omega_t x_1}$ and $\hat{\omega}_I = \omega_v v_m - (s \sin(\phi) + z_1) \frac{v_m}{f} \omega_t$.

Finally, by solving the integral in s , we obtain a closed-form expression for the plenoptic spectrum of a slanted plane. Before presenting this expression, we define the following three quantities

$$a = \omega_v v_m - \omega_t \frac{z_2 v_m}{f}, \quad b = \omega_v v_m - \omega_t \frac{z_1 v_m}{f}, \quad \text{and} \quad c = \frac{\omega_s f - f \omega_t \cos(\phi)}{\sin(\phi) \omega_t v_m}.$$

Notice that a and b depend respectively on the depths z_2 and z_1 , shown in Figure 3.1, and c depends on the frequency of the texture signal ω_s . Consequently the expression for the spectrum, in terms of these quantities, is

$$P_S(\omega_t, \omega_v) = \left(\frac{j 2 v_m}{\omega_t} \left[\text{sinc}(a) e^{-jT(\omega_t \cos(\phi) - \omega_s)} - \text{sinc}(b) \right] + \frac{j \omega_s f}{\sin(\phi) \omega_t^2} \left[\zeta \{jb(c-1)\} - \zeta \{ja(c-1)\} - \zeta \{jb(c+1)\} + \zeta \{ja(c+1)\} \right] e^{jbc} \right) e^{-j\omega_t x_1}, \quad (3.14)$$

² $\text{sinc}'(h) = \cos(h)/h - \sin(h)/h^2$ when $h \in \mathbb{R}$.

if $\omega_t \neq 0$, else

$$P_S(0, \omega_v) = 2v_m T \operatorname{sinc}\left(\frac{\omega_s T}{2}\right) \left[\cos(\phi) \operatorname{sinc}(\omega_v v_m) - j \frac{\sin(\phi) v_m}{f} \operatorname{sinc}'(\omega_v v_m) \right] e^{j\omega_s \frac{T}{2}}.$$

The function ζ is defined as

$$\zeta\{jh\} = \begin{cases} E_1(jh) + \ln|h| + j\frac{\pi}{2} + \gamma & \text{if } h > 0, \\ E_1^*(j|h|) + \ln|h| - j\frac{\pi}{2} + \gamma & \text{if } h < 0, \\ 0 & \text{if } h = 0, \end{cases}$$

where $h \in \mathbb{R}$, γ is Euler's constant, $E_1(jh)$ is the exponential integral [1] and $E_1^*(jh)$ is its complex conjugate. Figure 3.4(a) shows an example of a plenoptic spectrum obtained using this expression. The spectrum in the figure corresponds to a slanted plane with a texture signal $g(s) = \cos(\omega_s s) = e^{j\omega_s s} + e^{-j\omega_s s}$, hence (3.14) is calculate for both exponentials and then combined.

To give some insight into this result, let us examine the effect of the quantities a , b and c on the expression in (3.14). In particular, we are interested in the situations when $a = 0$, $b = 0$, $c + 1 = 0$ and $c - 1 = 0$. Starting with $a = 0$ and $b = 0$, these conditions occur when the following occurs in frequency

$$\omega_v = \omega_t \frac{z_2}{f} \quad \text{and} \quad \omega_v = \omega_t \frac{z_1}{f},$$

respectively. Therefore, when applied to (3.14), each condition yields a diagonal line in the plenoptic spectrum that relates to either z_1 or z_2 . Given that z_1 and z_2 are the limits of the plane's depth, these diagonal lines form the basis of the 'bow-tie' shape shown in Figure 2.5(b). Note, however, that the spectrum is no longer bound between these lines. If we now consider $c + 1 = 0$ and $c - 1 = 0$, these conditions occur when the following occurs in frequency

$$\omega_t = \frac{\omega_s f}{f \cos(\phi) + \sin(\phi) v_m} \quad \text{and} \quad \omega_t = \frac{\omega_s f}{f \cos(\phi) - \sin(\phi) v_m},$$

respectively. Therefore, when applied to (3.14), each condition yields a vertical line the plenoptic spectrum. We shall examine these characteristics in more detail in Section 3.4.

On a final note, the expression defined in (3.14) is also consistent with the spectrum for a fronto-parallel plane defined in (3.10). In other words, if we assume a complex exponential texture in (3.10), then the following is true:

$$\lim_{\phi \rightarrow 0} \left\{ P_S(\omega_t, \omega_v) \right\} = P_{FPP}(\omega_t, \omega_v). \quad (3.15)$$

For proof of this relationship see Appendix B.1.

3.3.2 Extending to Multiple Slanted Planes

Finally, we generalise the previous spectral analysis to scenes comprising multiple slanted planes. In this case, assuming the surface is continuous, the scene geometry equations are the set of the individual geometry equations for each plane. Therefore, given a scene comprising L_s slanted planes with complex exponential texture, the geometry is described as follows

$$\mathcal{G}_{MP} = \mathcal{G}_{s,i}, \text{ if } x \in [x_i, x_{i+1}] \forall i = 1, \dots, L_s \quad (3.16)$$

where x_i and x_{i+1} are the starting and ending spatial positions of the i th plane in the surface and $\mathcal{G}_{s,i}$ are the corresponding scene geometry equations. These individual scene geometry equations, $G_{s,i}$, are defined as follows

$$\mathcal{G}_{s,i} = \begin{cases} z(x) = (x - x_i) \tan(\phi_i) + z_i \\ x(s) = (s - s_i) \cos(\phi_i) + x_i \\ g(s) = e^{j\omega_s s} \end{cases} \quad (3.17)$$

where $s_i \forall i = 1, \dots, L_s$ is an offset parameter in order to preserve the consistency in the curvilinear coordinate. These offsets parameters are defined as

$$s_i = \begin{cases} 0 & \text{if } i = 1, \\ \sum_{k=1}^{i-1} T_k & \text{if } i = 2, \dots, L_s. \end{cases} \quad (3.18)$$

Consequently the overall range of the curvilinear coordinate is $s \in [0, s_{L_s}]$.

The inclusion of this offset parameter results in a phase shift occurring in the corresponding plenoptic spectrum. As a result, using the same procedure as detailed in Appendix A, the plenoptic spectrum for the i th plane in the surface is

$$P_{S,i}(\omega_t, \omega_v) = \left(\frac{j2v_m}{\omega_t} \left[\text{sinc}(a_i) e^{-jT(\omega_t \cos(\phi_i) - \omega_s)} - \text{sinc}(b_i) \right] + \frac{j\omega_s f}{\sin(\phi_i)\omega_t^2} \left[\zeta \{jb_i(c_i - 1)\} \right. \right. \\ \left. \left. - \zeta \{ja_i(c_i - 1)\} - \zeta \{jb_i(c_i + 1)\} + \zeta \{ja_i(c_i + 1)\} \right] e^{jb_i c_i} \right) e^{-j(\omega_t x_i - s_i \omega_s)}, \quad (3.19)$$

if $\omega_t \neq 0$, else

$$P_{S,i}(0, \omega_v) = 2v_m T_i \text{sinc} \left(\frac{\omega_s T_i}{2} \right) \left[\cos(\phi_i) \text{sinc}(\omega_v v_m) \right. \\ \left. - j \frac{\sin(\phi_i) v_m}{f} \text{sinc}'(\omega_v v_m) \right] e^{j\omega_s (s_i + \frac{T_i}{2})}.$$

The parameters a_i , b_i and c_i in (3.19) are

$$a_i = \omega_v v_m - \omega_t \frac{z_{i+1} v_m}{f}, \quad b_i = \omega_v v_m - \omega_t \frac{z_i v_m}{f}, \quad \text{and} \quad c = \frac{\omega_s f - f \omega_t \cos(\phi_i)}{\sin(\phi_i) \omega_t v_m}.$$

Therefore, using the linear property of the Fourier transform, the overall plenoptic spectrum for a scene comprising L_s slanted planes with complex exponential texture is

$$P_{MP}(\omega_t, \omega_v) = \sum_{i=1}^{L_s} P_{S,i}(\omega_t, \omega_v). \quad (3.20)$$

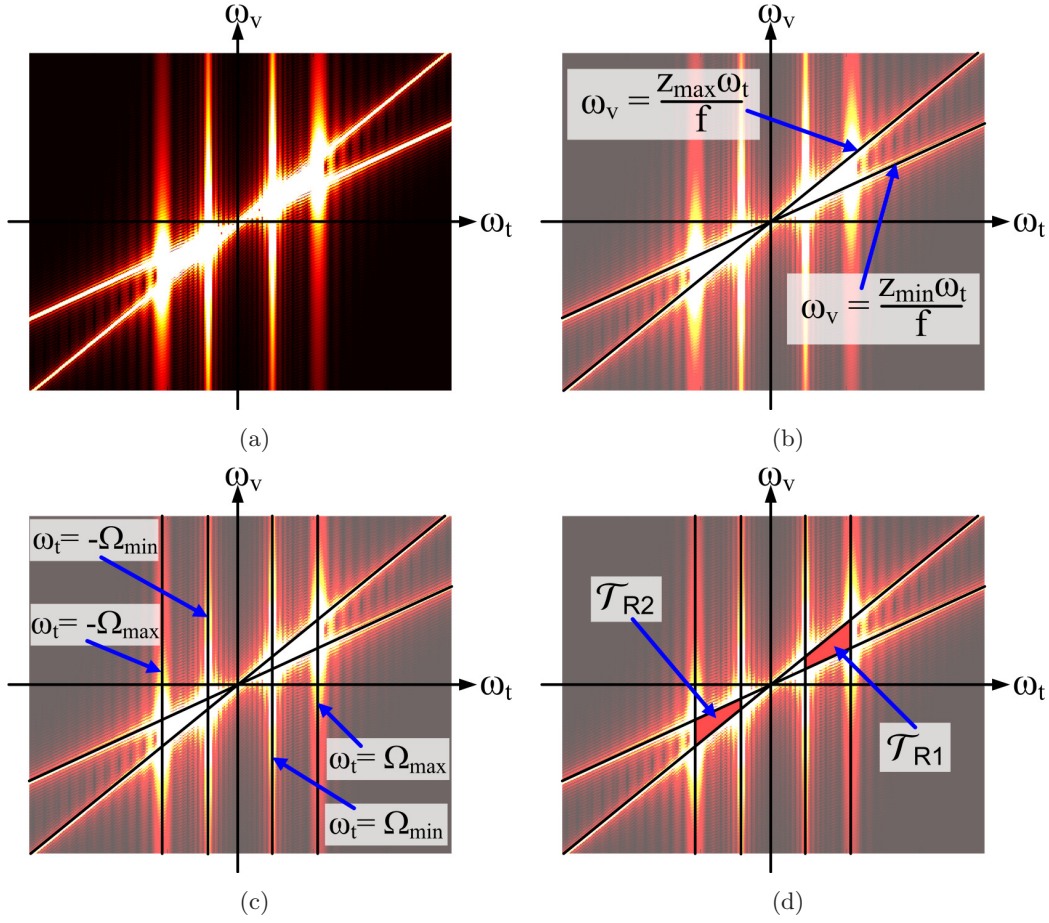


Figure 3.4: Diagram (a) shows the plenoptic spectrum for a slanted plane with a texture signal $g(s) = \cos(\omega_s s)$. The same spectrum is shown in (b), (c) and (d) each with different characteristics of the spectrum superimposed. These characteristics are: (b) the diagonal lines relating to the maximum and minimum depths of the plane, (c) the four vertical lines relating to the modulation of the texture signal when projected onto the image plane and (d) the two quadrilateral regions enclosed, \mathcal{T}_{R1} and \mathcal{T}_{R2} . Note that $\Omega_{\max} = \omega_s f / (f \cos(\phi) - \sin(\phi) v_m)$ and $\Omega_{\min} = \omega_s f / (f \cos(\phi) + \sin(\phi) v_m)$.

3.4 Behaviour of the Spectrum for a Slanted Plane

As predicted, the plenoptic spectrum for a slanted plane under FSW and FFoV is band-unlimited. However, by analysing our closed-form expression, we can characterise the magnitude of the spectrum. This characterisation results in a structural model of the plenoptic spectrum that can qualitatively describe how the spectrum changes with the properties of the scene. As most scenes have non-complex texture, we shall assume a real, sinusoidal texture signal, comprising a positive and negative complex exponential, when examining this structural model. Under this assumption, the structural model

comprises two bounded regions constructed from six lines. The first two lines are diagonal and relate to the maximum and minimum depths of the slanted plane, i.e.

$$\omega_v = \omega_t \frac{z_{\max}}{f} \quad \text{and} \quad \omega_v = \omega_t \frac{z_{\min}}{f},$$

where $z_{\max} = \max\{z_1, z_2\}$ is the maximum depth and $z_{\min} = \min\{z_1, z_2\}$ is the minimum. In terms of the plenoptic spectrum in (3.14), these lines correspond to the situation when $a = 0$ or $b = 0$. Using the same plenoptic spectrum as shown in Figure 3.4(a), these diagonal lines are illustrated in Figure 3.4(b).

The other four lines are vertical and relate to the modulation of the texture signal when projected into the image plane. Using the rotational symmetry of the plenoptic spectrum, these four form two pairs $\omega_t = \pm\Omega_{\max}$ and $\omega_t = \pm\Omega_{\min}$, where Ω_{\max} represents the maximum frequency of the modulated texture signal and Ω_{\min} represents its minimum. These quantities are defined as follows

$$\Omega_{\max} = \frac{\omega_s f}{f \cos(\phi) - |\sin(\phi)| v_m} \quad \text{and} \quad \Omega_{\min} = \frac{\omega_s f}{f \cos(\phi) + |\sin(\phi)| v_m},$$

see Figure 3.4(c) for an illustration. Notice that these quantities correspond to $c+1 = 0$ and $c-1 = 0$ in (3.14). In [17] this modulation of the texture signal is shown to be equivalent to time-warping the texture signal with a scene dependent warping function. Therefore the single sinusoidal texture pasted to the synthetic scene in Figure 3.4 is warped when projected into the image plane, thus smearing the single frequency component into a range of frequencies.

Finally, the two bounded regions are the areas enclosed by the intersection of these lines, see Figure 3.4(d) for an illustration. From the figure we observe that the two quadrilaterals, marked \mathcal{T}_{R1} and \mathcal{T}_{R2} , contain most of the energy of the plenoptic spectrum. Mathematically the quadrilateral regions are defined as

$$\mathcal{T}_{R1} = \left\{ \omega_t, \omega_v : \omega_t \in [\Omega_{\min}, \Omega_{\max}], \omega_v \in \left[\omega_t \frac{z_{\min}}{f}, \omega_t \frac{z_{\max}}{f} \right] \right\}, \quad (3.21)$$

and

$$\mathcal{T}_{R2} = \left\{ \omega_t, \omega_v : \omega_t \in [-\Omega_{\max}, -\Omega_{\min}], \omega_v \in \left[\omega_t \frac{z_{\max}}{f}, \omega_t \frac{z_{\min}}{f} \right] \right\}. \quad (3.22)$$

To illustrate the usefulness of this model we shall now use it to examine the effect on the plenoptic spectrum when the following four parameters vary; the slant of the plane ϕ , the maximum frequency of the texture signal ω_s , the minimum depth of the scene z_{\min} and the field of view of the cameras (i.e. the quantity v_m/f). To help visualise these effects we introduce an example of a plenoptic spectrum corresponding to a synthetic slanted plane, see Figure 3.5(b). This synthetic scene, with a texture $g(s) = \cos(\omega_s s)$, is initialised with the following parameters; $x_1 = 0\text{m}$, $z_{\min} = 1.5\text{m}$, $\phi = 0.6\text{rads}$, $T = 2.1\text{m}$, $\omega_s = 20\pi\text{rads/m}$ and $\text{FoV} = 40^\circ$. Note that we repeat this spectrum in Figure 3.5(e), (h) and (k) in order to highlight the effect of changing individual scene parameters.

3.4.1 Varying the Angle of Slant

Unsurprisingly, the slant of the plane, ϕ , controls several aspects of the plenoptic spectrum. For example the maximum depth of the scene depends on the slant of the plane as such $z_{\max} = z_{\min} + T |\sin(\phi)|$. Therefore varying ϕ will effect the diagonal line relating to the maximum depth in the spectrum. At the same time, the slant of the plane also controls the position of Ω_{\max} and Ω_{\min} ; an increase in ϕ will increase Ω_{\max} whilst decreasing Ω_{\min} . Therefore a change in ϕ will alter the overall size of the two bounded quadrilaterals \mathcal{T}_{R1} and \mathcal{T}_{R2} . At its most extreme, when $\phi = 0$, these regions disappear entirely as $\Omega_{\max} = \Omega_{\min}$ and $z_{\max} = z_{\min}$.

Using the synthetic example, we demonstrate the effect of varying ϕ on its plenoptic spectrum in Figure 3.5(a), (b) and (c). The slant of the plane is decreased to $\phi = 0.1\text{rads}$ in Figure 3.5(a), and increased to $\phi = 0.9\text{rads}$ in Figure 3.5(c). Notice that in both cases we have fixed the maximum and minimum depth of the plane thus limiting the expansion of the quadrilateral regions. A side effect of this is that the width of the plane, T , must also vary as $z_{\max} - z_{\min} = T |\sin(\phi)|$. This change in T slightly alters

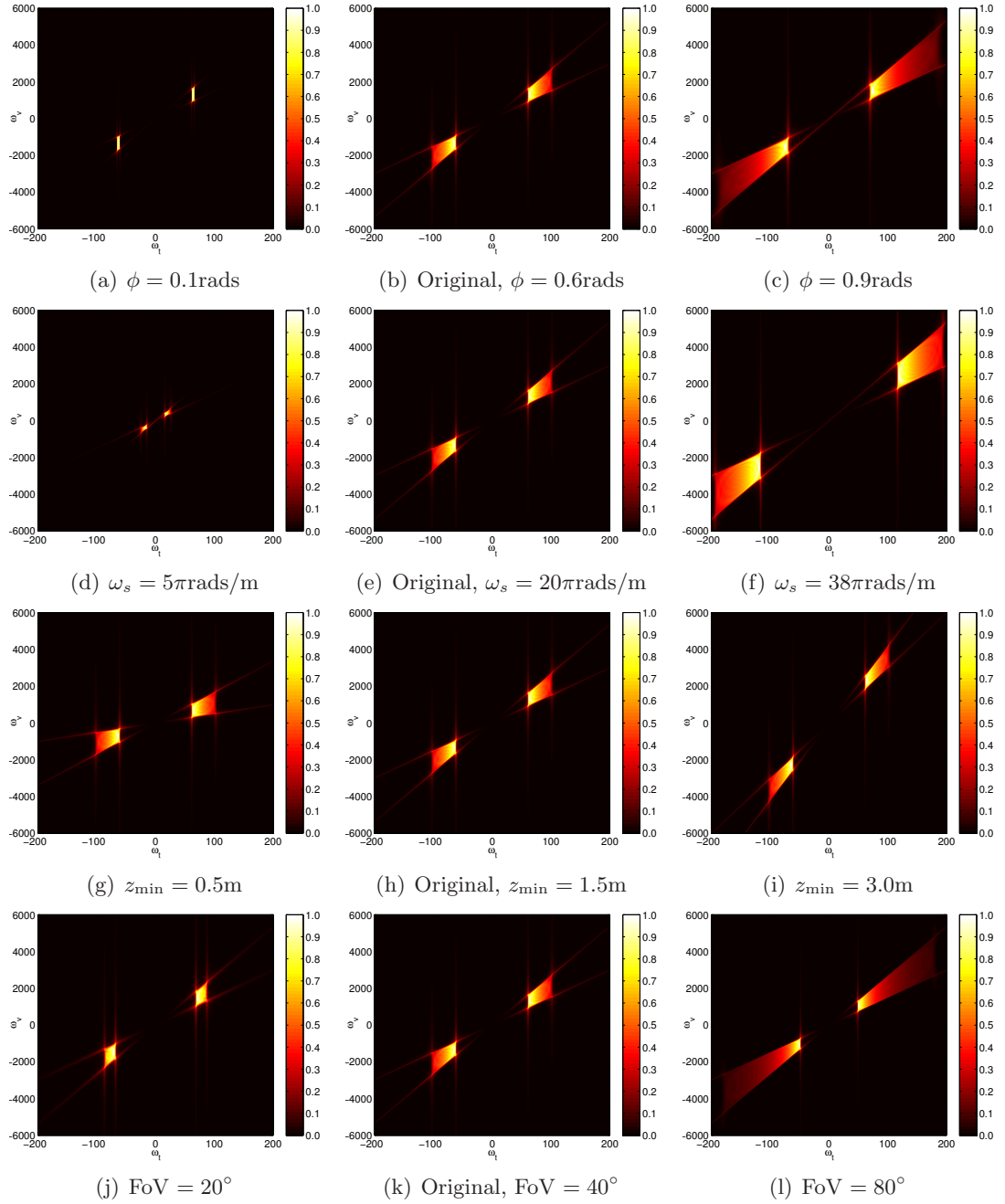


Figure 3.5: Diagrams showing the behaviour of the magnitude of the plenoptic spectrum for a slanted plane, $|P_S(\omega_t, \omega_v)|$, as four scene parameters are varied. Note that, for clarity, the magnitude of each spectrum is normalised.

the spectral spreading in the ω_t -axis.

3.4.2 Varying the Maximum Frequency of the Texture

The maximum frequency of the texture signal, ω_s , is directly proportional to the quantities Ω_{\max} and Ω_{\min} . This is to be expected as these quantities are the maximum and minimum frequencies of the warped texture signal, respectively. As a result increasing ω_s will increase both Ω_{\min} and Ω_{\max} , which shifts the quadrilateral regions, \mathcal{T}_{R1} and \mathcal{T}_{R2} , further from the ω_v -axis. The overall size of the regions also increases as they are only bounded in ω_v by the lines relating to the maximum and minimum depths of the scene. At the other extreme, both the quadrilateral regions will merge into a point at the origin if $\omega_s = 0$.

These effects are illustrated in Figure 3.5(d), (e) and (f) using the plenoptic spectrum corresponding to the synthetic scene. In the first figure the frequency is reduced to $\omega_s = 5\pi\text{rads/m}$, which causes the quadrilateral regions to decrease in size and move closer to the origin. The opposite occurs in Figure 3.5(f) when the frequency is increased to $\omega_s = 38\pi\text{rads/m}$.

3.4.3 Varying the Minimum Depth of the Scene

Assuming the relative depth variation is constant, then varying the minimum depth between the camera line and the slanted plane, z_{\min} , controls the gradient of the two diagonal lines illustrated in Figure 3.4(b). Therefore moving the camera line will result in the diagonal lines rotating around the frequency origin. If z_{\min} is decreased, i.e. the cameras are closer to the scene, then the rotation is towards the ω_t -axis whilst increasing z_{\min} will lead to a rotation away from the ω_t -axis. As the quadrilateral regions, \mathcal{T}_{R1} and \mathcal{T}_{R2} , are bounded by these diagonal lines, then varying z_{\min} will alter their position relative to the ω_t -axis and their width relative to the ω_v -axis. In other words increasing z_{\min} will move the regions further from the ω_t -axis and, at the same time, increase their width in ω_v .

We illustrate the effect of altering the minimum depth of the scene in Figure 3.5(g), (h) and (i). The camera line is moved closer to the synthetic scene in the first figure,

hence $z_{\min} = 0.5\text{m}$, and then moved further away in the third figure so that $z_{\min} = 3\text{m}$. Note that in both cases the relative depth variation is constant. As expected, the figures show how the diagonal lines relating to the maximum and minimum depths of the scene are rotated around the origin. This rotation causes the quadrilateral regions to move relative to the ω_t -axis; an increase in z_{\min} shifts the regions further from the ω_t -axis.

3.4.4 Varying the Field of View

The FoV of the camera directly controls the size of the quadrilateral regions, \mathcal{T}_{R1} and \mathcal{T}_{R2} ; an increase in the FoV elongates the regions in the ω_t -axis as Ω_{\max} increases and Ω_{\min} decreases. In particular, if ϕ is fixed, then Ω_{\max} will tend to infinite as the FoV becomes large. This is due to the no-occlusion constraint placed on the scene, i.e. as FoV increases the quantity f/v_m decreases until it invalidates the inequality in (3.3). Therefore the smaller the difference between $\tan(\phi)$ and f/v_m the larger the quadrilateral regions. Notice, at the other extreme, the regions reduce to a single line in ω_t as the FoV tends toward zero (i.e. each picture consists of a single pixel).

This effect is illustrated in Figure 3.5(j), (k) and (l) using the synthetic scene. In the first figure the cameras have a narrow FoV equal to 20° and in the third the cameras have a wide FoV equal to 80° . The figures show how the quadrilateral regions expand in ω_t as the FoV is increased. The figures also highlight how the lines $\omega_t = \pm\Omega_{\max}$ are more pronounced when the FoV of the cameras is small. The reason for this is that a narrow FoV is equivalent to a narrow window in the EPI domain, which causes more pronounced spectral spreading along the ω_v -axis.

3.5 Rotation of the Camera Line

In the discussion so far we have determined the plenoptic spectrum for a slanted plane and examined how it reacts when parameters of the scene are changed. However, instead of changing the scene, we could rotate the camera line around a point x_r to an angle α . This rotation of the camera line is shown in Figure 3.6. In this new scenario,

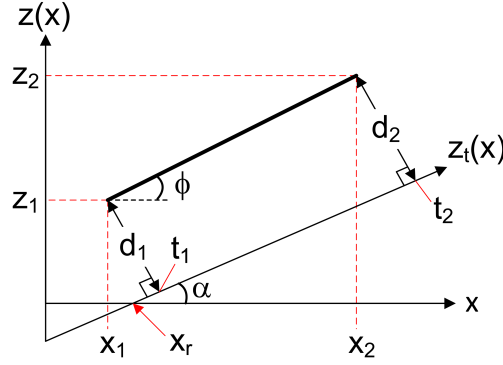


Figure 3.6: Diagram illustrating the rotation of the camera line around a point x_r at an angle α . Note that d_1 and d_2 are the distances between the camera line and plane at x_1 and x_2 , respectively.

the scene can be redefined relative to the rotated camera line. The slant of the plane relative to the camera line is $\phi - \alpha$, the perpendicular distance between the camera line and the spatial coordinate (x_1, z_1) is

$$d_1 = \cos(\alpha) \left(z_1 - (x_1 - x_r) \tan(\alpha) \right), \quad (3.23)$$

and the perpendicular distance between the camera line and (x_2, z_2) is

$$d_2 = \cos(\alpha) \left(z_2 - (x_2 - x_r) \tan(\alpha) \right). \quad (3.24)$$

Notice that $d_1 = z_1$ and $d_2 = z_2$ when $\alpha = 0$. Lastly, the no-occlusion constraint for the scene is

$$|\tan(\phi - \alpha)| < \frac{f}{v_m}. \quad (3.25)$$

Therefore the plenoptic spectrum for this scenario is equivalent to the spectrum corresponding to a plane with a slant $\phi - \alpha$, a minimum depth $z_{\min} = \min\{d_1, d_2\}$ and a maximum depth $z_{\max} = \max\{d_2\}$.

In view of this, the plenoptic spectrum for a slanted plane observed from a camera

line rotated around a point x_r at an angle α is

$$P_{Sr}(\omega_t, \omega_v) = e^{-j\omega_t t_1} \left(\frac{j2v_m}{\omega_t} \left[\text{sinc}(a_r) e^{-jT(\omega_t \cos(\phi-\alpha)-\omega_s)} - \text{sinc}(b_r) \right] + \frac{j\omega_s f e^{jb_r c_r}}{\sin(\phi-\alpha)\omega_t^2} \right. \\ \left. \cdot \left[\zeta \{jb_r(c_r-1)\} - \zeta \{ja_r(c_r-1)\} - \zeta \{jb_r(c_r+1)\} + \zeta \{ja_r(c_r+1)\} \right] \right), \quad (3.26)$$

if $\omega_t \neq 0$, else

$$P_{Sr}(0, \omega_v) = 2v_m T \text{sinc}\left(\frac{\omega_s T}{2}\right) \left[\cos(\phi-\alpha) \text{sinc}(\omega_v v_m) - j \frac{\sin(\phi-\alpha)v_m}{f} \text{sinc}'(\omega_v v_m) \right] e^{j\omega_s \frac{T}{2}},$$

where t_1 is the point on the camera line from which d_1 is measured (see Figure 3.6), and

$$a_r = \omega_v v_m - \omega_t \frac{d_2 v_m}{f}, \quad b_r = \omega_v v_m - \omega_t \frac{d_1 v_m}{f}, \quad \text{and} \quad c_r = \frac{\omega_s f - f \cos(\alpha - \phi) \omega_t}{\omega_t v_m \sin(\phi - \alpha)}.$$

A full derivation of this plenoptic spectrum is presented in Appendix C.

The main point to take from this analysis is that, as we are unlikely to be able to alter the scene in practice, we can rotate the camera line in order to obtain a suitable plenoptic spectrum. In particular, if we were to rotate the camera line such that $\alpha = \phi$ then the plenoptic spectrum would correspond to a fronto-parallel plane, which, as we shall see in the next chapter, requires less cameras to sample.

3.6 Summary

In this chapter we have re-examined the spectral analysis for the plenoptic function generated by a slanted plane. The novelty of our approach is that we have incorporated two realistic constraints, FSW and FFoV, directly into the spectral analysis. The imposition of these finite constraints leads to spectral spreading in the frequency domain, which results in band-unlimited plenoptic spectra. The nature of this spectral spreading

is shown to be complex when dealing with scenes that have depth variation. However, we are able to derive an exact closed-form expression for the plenoptic spectrum of a Lambertian slanted plane with complex exponential texture. This expression can be extended to scenes with more complicated textures, and scenes comprising multiple slanted planes, using the linearity of the Fourier transform.

We examined the behaviour of the plenoptic spectrum for a slanted plane as scene and camera parameters varied. We characterised this behaviour using a structural model derived from the expression of the spectrum. The structural model comprised six lines and two bounded quadrilateral regions. The six lines in question depended on the maximum and minimum scene depth, the projection of the texture signal into the image plane and the camera parameters. Finally, we showed that rotation of the camera line around a point alters the plenoptic spectrum of a slanted plane as if altering the scene parameters.

The key points to take from this chapter are as follows:

- The FSW and FFoV constraints result in band-unlimitedness in the plenoptic spectrum, which is non-trivial for scenes with depth variation.
- Using the FSW and FFoV constraints, an exact closed-form expression for the plenoptic spectrum of a Lambertian slanted plane with complex exponential texture is derived.
- Analysis of the closed-form expression allows the structure of the plenoptic spectrum to be characterised using a model comprising six lines and two bounded quadrilateral regions.
- The structure of the plenoptic spectrum for a slanted plane is altered by either varying the scene geometry directly or rotating the camera line around a point.

Chapter 4

Uniform Plenoptic Sampling

4.1 Introduction

In this chapter we examine the uniform sampling and reconstruction of the plenoptic function for a slanted plane with bandlimited texture. The plenoptic spectrum for the slanted plane, however, is band-unlimited. In view of this, we utilise the concept of the essential bandwidth proposed in [35]. This bandwidth is a region in the frequency domain that contains approximately 90% of the signal's energy. We choose this percentage because it coincides with the percentage of energy within the main lobe of the sinc function [35]. This chapter, therefore, analyses the sampling and reconstruction of the plenoptic function for a slanted plane assuming it to be bandlimited to its essential bandwidth.

The idea of calculating the essential bandwidth for the plenoptic function has been proposed in [17], however they do not use knowledge of the essential bandwidth to sample the plenoptic function. Also, their analysis is based on fixing either t or v and then studying the bandwidth. Consequently they determine two separate 1D essential bandwidths, one in ω_t and the other in ω_v . In contrast, we determine a non-separable 2D essential bandwidth in the (ω_t, ω_v) -domain using our spectral analysis of the plenoptic function.

This chapter is organised as follows: in Section 4.2 we re-examine the plenoptic spectrum of a slanted plane and derive its essential bandwidth using a parametric

model. We then use this essential bandwidth in Section 4.3 to sample and reconstruct the corresponding plenoptic function. In particular, we derive a new expression for the maximum spacing between adjacent cameras and a new reconstruction filter. Sections 4.3.1 and 4.3.2 examine the results of using these expressions to sample and reconstruct several synthetic EPIs. This analysis is then extended in Section 4.4, where we examine the reconstruction of an EPI-volume consisting of real images. Section 4.4 also explores the improvement in reconstruction quality when using different interpolating functions. We conclude the chapter in Section 4.5 with a summary.

4.2 The Essential Bandwidth for a Slanted Plane

Using the analysis in Chapter 3, we now examine the spectral support of the plenoptic function for a slanted plane, with bandlimited texture, in order to determine its essential bandwidth EB_S . As the plenoptic spectrum in question is a 2D object, the corresponding essential bandwidth will be a 2D region in frequency that contains approximately 90% of its energy. To reduce the possible 2D regions that satisfy this criteria, we constrain the essential bandwidth EB_S to be a compact region in the frequency domain that is symmetrical around the origin.

Now, from the analysis of the plenoptic spectrum in Section 3.4, we observe that the spectral energy is concentrated in the two quadrilateral regions, \mathcal{T}_{R1} and \mathcal{T}_{R2} , shown in Figure 3.4. Thus we further constrain the essential bandwidth EB_S such that it contains the majority of these quadrilateral regions. Consequently, we propose a parametric model, shaped like a parallelogram, for the essential bandwidth EB_S that is centred around the origin. The model comprises four parameters: Ω_v , the maximum frequency in ω_v ; Ω_t , the maximum frequency in ω_t ; z_G/f , a parameter that controls the skew of the model relative to the ω_v -axis, and A , the width of the model in ω_t . Consequently the parametric essential bandwidth for the plenoptic spectrum of a slanted plane is

$$EB_S = \left\{ \omega_t, \omega_v : \omega_t \in [-\Omega_t, \Omega_t], \omega_v \in \left[\frac{z_G}{f} \left(\omega_t - \frac{A}{2} \right), \frac{z_G}{f} \left(\omega_t + \frac{A}{2} \right) \right] \right\}. \quad (4.1)$$

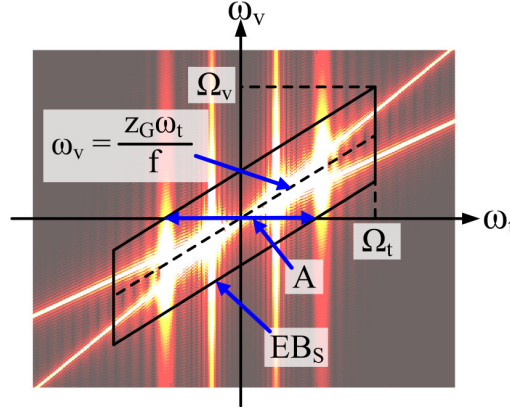


Figure 4.1: Diagram illustrating the parametric essential bandwidth for the plenoptic spectrum corresponding to a slanted plane with bandlimited texture. The essential bandwidth, EB_S , is superimposed on an example spectrum in which the texture signal is $g(s) = \cos(\omega_s s)$. The model of the essential bandwidth comprises four parameters: Ω_t and Ω_v , the maximum values in ω_t and ω_v respectively, z_G/f the parameter that controls skew of the model relative to the ω_v -axis, and A the width of the essential bandwidth in ω_t .

An illustration of this model is shown in Figure 4.1. The figure presents the essential bandwidth superimposed on an actual plenoptic spectrum. In the following discussion we will determine each parameter in the model.

4.2.1 The Essential Bandwidth Parameters

To determine the four parameters defined in (4.1), we first revisit the spectral analysis of the plenoptic function for a slanted plane. In the previous chapter we focused on deriving an exact closed-form expression for this spectrum assuming the texture was a complex exponential. It could, therefore, be extended to more complicated texture using the linear property of the Fourier transform. However, in this situation, we are interested in determining an expression for the plenoptic spectrum in terms of the Fourier transform of the bandlimited texture signal. It is not necessary that this new expression be closed-form.

In view of this, we start with the equation for the plenoptic spectrum given in (3.13). If we impose the finite integration limits with the following rectangular window

function

$$\text{rect}\left(\frac{s}{T} - \frac{1}{2}\right) = \begin{cases} 1, & \text{if } 0 \leq s \leq T \\ 0, & \text{else,} \end{cases} \quad (4.2)$$

then the plenoptic spectrum in (3.13) becomes

$$P_S(\omega_t, \omega_v) = M_1 \cos(\phi) \left[\int_{-\infty}^{\infty} \text{rect}\left(\frac{s}{T} - \frac{1}{2}\right) g(s) \text{sinc}(\hat{\omega}_I) e^{-j\omega_t \cos(\phi)s} ds \right. \\ \left. - j \frac{v_m \tan(\phi)}{f} \int_{-\infty}^{\infty} \text{rect}\left(\frac{s}{T} - \frac{1}{2}\right) g(s) \text{sinc}'(\hat{\omega}_I) e^{-j\omega_t \cos(\phi)s} ds \right], \quad (4.3)$$

where $M_1 = 2v_m e^{-j\omega_t x_1}$ and $\hat{\omega}_I = \omega_v v_m - (s \sin(\phi) + z_1) \frac{v_m}{f} \omega_t$.

We now combine the rectangular function with the texture signal as follows

$$h(s) = \text{rect}\left(\frac{s}{T} - \frac{1}{2}\right) g(s), \quad (4.4)$$

hence the plenoptic spectrum becomes

$$P_S(\omega_t, \omega_v) = M_1 \cos(\phi) \int_{-\infty}^{\infty} h(s) \left[\text{sinc}(\hat{\omega}_I) - j \frac{v_m \tan(\phi)}{f} \text{sinc}'(\hat{\omega}_I) \right] e^{-j\omega_t \cos(\phi)s} ds. \quad (4.5)$$

At this point, we define $H(\Omega)$ as the Fourier transform of $h(s)$ with Ω as its frequency variable. Therefore $H(\Omega)$ is the Fourier transform of the texture signal convolved with a sinc function. By expressing $h(s)$ in terms of $H(\Omega)$, we can rearrange (4.5), as shown in Appendix A.2, to obtain the following expression for the plenoptic spectrum

$$P_S(\omega_t, \omega_v) = M_2 \int_{-\infty}^{\infty} H(\Omega) \frac{f\Omega}{\sin(\phi)\omega_t^2} \text{rect}\left(\frac{\Omega f - \omega_t \cos(\phi)f}{2 \sin(\phi)v_m \omega_t}\right) e^{-j\Omega\left(\frac{z_1}{\sin(\phi)} - \frac{\omega_v f}{\sin(\phi)\omega_t}\right)} d\Omega \quad (4.6)$$

where

$$M_2 = e^{-j\omega_t\left(x_1 - \frac{z_1}{\tan(\phi)}\right)} e^{-j\omega_v\left(\frac{f}{\tan(\phi)}\right)}. \quad (4.7)$$

Having derived (4.6), we now want to determine the parameters for the essential bandwidth EB_S . From the definition of EB_S in (4.1), we observe that the four parameters we require are not independent. In particular Ω_v is related to the other parameters as follows:

$$\Omega_v = \frac{z_G}{f} \left(\Omega_t + \frac{A}{2} \right).$$

Therefore, the problem we are trying to solve is multidimensional and non-separable. In view of this, to simplify the problem, we analyse the bandwidth of (4.6) along certain slices in frequency to obtain the values of Ω_t and Ω_v . As Ω_v is related to the other parameters, we first determine Ω_t and then use the result to obtain Ω_v . Given these values and the accompanying bandwidth analysis, we then determine the remaining parameters, A and z_G . The derivation of each parameter is as follows.

Note that we assume $\phi > 0$ for this analysis, hence $z_{\max} = z_2$ and $z_{\min} = z_1$. This assumption does not, however, result in a loss of generality due to the symmetry property of the plenoptic spectrum presented in Appendix B.2.

Determining the Ω_t Parameter

To determine the maximum value in ω_t , we examine the bandwidth of (4.6) along one of the two diagonal lines highlighted in Section 3.4. Assuming the worse case in ω_v , we select the line relating to the maximum depth of the scene, hence

$$\omega_v = \omega_t \frac{z_{\max}}{f}.$$

The resulting 1D plenoptic spectrum created by restricting (4.6) to this line is

$$P_S \left(\omega_t, \omega_t \frac{z_{\max}}{f} \right) = e^{-j\omega_t x_2} \int_{-\infty}^{\infty} H(\Omega) \frac{f\Omega}{\sin(\phi)\omega_t^2} \text{rect} \left(\frac{\Omega f - \omega_t \cos(\phi)f}{2\sin(\phi)v_m\omega_t} \right) e^{-j\Omega T} d\Omega. \quad (4.8)$$

From the definition of the rect function, the above integrand is only non-zero when the following inequality is true

$$-\frac{1}{2} \leq \frac{\Omega f - \omega_t \cos(\phi) f}{2 \sin(\phi) v_m \omega_t} \leq \frac{1}{2}. \quad (4.9)$$

Therefore, if $H(\Omega)$ is bandlimited to a frequency Ω_s , then (4.9) can be rearranged to show that $P_S(\omega_t, \omega_v)$ is zero outside the range

$$-\frac{\Omega_s f}{f \cos(\phi) - v_m \sin(\phi)} \leq \omega_t \leq \frac{\Omega_s f}{f \cos(\phi) - v_m \sin(\phi)}, \quad (4.10)$$

hence we have a bound on ω_t .

A problem, however, is that $H(\Omega)$ is not bandlimited since its definition involves a convolution with a sinc function. In view of this, we assume that $H(\Omega)$ is approximately bandlimited to its essential bandwidth. The essential bandwidth of $H(\Omega)$ is determined as follows; first we use the property that the essential bandwidth of a signal constructed from a convolution is equal to the sum of the individual essential bandwidths relating to the input signals. Secondly, the essential bandwidth of a sinc function is equal to the width of its main lobe [35]. Therefore given a function $\text{sinc}(k\Omega)$, where $k \in \mathbb{R}$ is a constant, its essential bandwidth is

$$\left\{ \Omega : |\Omega| \leq \frac{\pi}{k} \right\}. \quad (4.11)$$

In view of this, the essential bandwidth for $H(\Omega)$ is

$$\text{EB}_H = \left\{ \Omega : |\Omega| \leq \omega_s + \frac{2\pi}{T} \right\}, \quad (4.12)$$

which means

$$\Omega_s = \omega_s + \frac{2\pi}{T}. \quad (4.13)$$

As a result we obtain the following maximum limit in ω_t

$$\Omega_t = \frac{\Omega_s f}{f \cos(\phi) - v_m |\sin(\phi)|} = \frac{f}{f \cos(\phi) - v_m |\sin(\phi)|} \left(\omega_s + \frac{2\pi}{T} \right). \quad (4.14)$$

The absolute of $\sin(\phi)$ is introduced so that (4.14) is valid for planes with $\phi < 0$.

Determining the Ω_v Parameter

We now determine Ω_v by examining the bandwidth of (4.6) along the line $\omega_t = \Omega_t$.

Using (4.14), we set

$$\omega_t \triangleq \frac{\Omega_s f}{k_\phi},$$

where $k_\phi = f \cos(\phi) - v_m \sin(\phi)$, and substitute the definition into (4.6) to give

$$P_S \left(\frac{\Omega_s f}{k_\phi}, \omega_v \right) = \int_{-\infty}^{\infty} \Omega \operatorname{rect} \left(\frac{\Omega k_\phi}{2v_m \Omega_s \sin(\phi)} - \frac{f}{2v_m \tan(\phi)} \right) H(\Omega) e^{j\Omega \left(\frac{\omega_v k_\phi}{\Omega_s \sin(\phi)} - \frac{z_{\min}}{\sin(\phi)} \right)} d\Omega \\ \cdot \frac{k_\phi^2}{\sin(\phi) \Omega_s^2} e^{j \left(\frac{z_{\min}}{\tan(\phi)} - x_1 \right) \frac{\Omega_s f}{k_\phi}} e^{-j\omega_v \frac{f}{\tan(\phi)}}. \quad (4.15)$$

The integral in this equation is the inverse Fourier transform of the product of two functions. Therefore it can be expressed in terms of a convolution. If we name the integral I_1 , this convolution is

$$I_1 = \mathcal{F}_\Omega^{-1} \{ H(\Omega) \} * \mathcal{F}_\Omega^{-1} \left\{ \Omega \operatorname{rect} \left(\frac{\Omega k_\phi}{2v_m \Omega_s \sin(\phi)} - \frac{f}{2v_m \tan(\phi)} \right) \right\}, \quad (4.16)$$

where the variable for each inverse Fourier transform is

$$\frac{\omega_v k_\phi}{\sin(\phi) \Omega_s} - \frac{z_{\min}}{\sin(\phi)}. \quad (4.17)$$

As a result, the overall essential bandwidth is the sum of the individual bandwidths corresponding to each term in the convolution.

We start by determining the essential bandwidth for the first term in (4.16). Using the identity in (4.4), the solution to the first inverse Fourier transform is

$$\mathcal{F}_\Omega^{-1} \{ H(\Omega) \} = 2\pi g \left(\frac{\omega_v k_\phi}{\sin(\phi) \Omega_s} - \frac{z_{\min}}{\sin(\phi)} \right) \operatorname{rect} \left(\frac{\omega_v k_\phi}{T \sin(\phi) \Omega_s} - \frac{z_{\min}}{T \sin(\phi)} - \frac{1}{2} \right). \quad (4.18)$$

The presence of the rect function means that this solution is bounded in frequency.

Therefore it allows us to derive the following bounds on ω_v

$$\frac{z_{\min}}{f} \frac{\Omega_s f}{f \cos(\phi) - v_m \sin(\phi)} \leq \omega_v \leq \frac{z_{\max}}{f} \frac{\Omega_s f}{f \cos(\phi) - v_m \sin(\phi)}. \quad (4.19)$$

We now require the essential bandwidth for the second Fourier transform in I_1 . Using the derivative identity of the Fourier transform, we obtain the following closed-form expression for this transform

$$\begin{aligned} \mathcal{F}_\Omega^{-1} \left\{ \Omega \operatorname{rect} \left(\frac{\Omega k_\phi}{2v_m \Omega_s \sin(\phi)} - \frac{f}{2v_m \tan(\phi)} \right) \right\} &= M_3 \left[f \cos(\phi) \operatorname{sinc} \left(\omega_v v_m - z_{\min} \frac{\Omega_s v_m}{k_\phi} \right) \right. \\ &\quad \left. - j v_m \sin(\phi) \operatorname{sinc}' \left(\omega_v v_m - z_{\min} \frac{\Omega_s v_m}{k_\phi} \right) \right] e^{j \frac{f}{\tan(\phi)} \left(\omega_v - z_{\min} \frac{\Omega_s}{k_\phi} \right)}, \end{aligned} \quad (4.20)$$

where

$$M_3 = 2v_m \sin(\phi) \left(\frac{\Omega_s}{k_\phi} \right)^2.$$

The essential bandwidth is determined by examining the corresponding Energy Spectral Density (ESD) for this expression. The ESD for (4.20) is

$$\begin{aligned} \operatorname{ESD}(\omega_v) &= M_3^2 \left[f \cos(\phi) \operatorname{sinc} \left(\omega_v v_m - z_{\min} \frac{\Omega_s v_m}{k_\phi} \right) \right]^2 \\ &\quad + M_3^2 \left[v_m \sin(\phi) \operatorname{sinc}' \left(\omega_v v_m - z_{\min} \frac{\Omega_s v_m}{k_\phi} \right) \right]^2. \end{aligned} \quad (4.21)$$

From this expression, we observe that the ESD is the weighted combination of a squared sinc function and a squared derivative of a sinc function. The weighting of each function is dependent on the value of ϕ . If ϕ is very small then the ESD is approximately equal to the sinc function, and, if ϕ is equal to $\pi/2$, the opposite occurs. Given this structure, we could assume the worse case and approximate the essential bandwidth of (4.20) using the largest bandwidth of the two functions, which would be the bandwidth of the derivative. However, the size of ϕ is restricted by the no-occlusion constraint. For instance, a FoV of 45° would limit ϕ to just under $\pi/4$ rads. Therefore, to avoid over-estimating the essential bandwidth of (4.20), we propose a region that is the weighted combination of the individual essential bandwidths that belong to each function.

We weight each essential bandwidth based on the amount of energy contributed by that function to the total energy of (4.20). If we term E_{sinc} as the energy contribution related to the sinc function and E_{dsinc} as the corresponding contribution for the derivative of the sinc function, then the weights for the sinc function and the derivative sinc function are

$$\left(\frac{E_{\text{sinc}}}{E_{\text{sinc}} + E_{\text{dsinc}}} \right) \text{ and } \left(\frac{E_{\text{dsinc}}}{E_{\text{sinc}} + E_{\text{dsinc}}} \right), \quad (4.22)$$

respectively. The energies E_{sinc} and E_{dsinc} are defined as follows

$$E_{\text{sinc}} = M_3^2 (f \cos(\phi))^2 \frac{\pi}{v_m}$$

and

$$E_{\text{dsinc}} = M_3^2 (v_m \sin(\phi))^2 \frac{\pi}{3v_m}.$$

Having determined the weights, we require the essential bandwidth for the derivative of a sinc function. This essential bandwidth is estimated in Appendix D and results in the following for the derivative of the sinc function defined in (4.20)

$$\left\{ \omega_v : \omega_v \in \left[z_{\min} \frac{\Omega_s}{k_\phi} - \frac{3.5\pi}{v_m}, z_{\min} \frac{\Omega_s}{k_\phi} + \frac{3.5\pi}{v_m} \right] \right\}. \quad (4.23)$$

Therefore, the complete essential bandwidth for (4.20) is

$$\text{EB}_1 = \left\{ \omega_v : \omega_v \in \left[z_{\min} \frac{\Omega_s}{k_\phi} - n(\phi, \bar{v}_m) \frac{\pi}{v_m}, z_{\min} \frac{\Omega_s}{k_\phi} + n(\phi, \bar{v}_m) \frac{\pi}{v_m} \right] \right\}, \quad (4.24)$$

where

$$n(\phi, \bar{v}_m) = \frac{3 \cos^2(\phi) + 3.5 (\bar{v}_m \sin(\phi))^2}{3 \cos^2(\phi) + (\bar{v}_m \sin(\phi))^2}. \quad (4.25)$$

This essential bandwidth gives the following bound on ω_v

$$z_{\min} \frac{\Omega_s}{k_\phi} - n(\phi, \bar{v}_m) \frac{\pi}{v_m} \leq \omega_v \leq z_{\min} \frac{\Omega_s}{k_\phi} + n(\phi, \bar{v}_m) \frac{\pi}{v_m}. \quad (4.26)$$

Validation of this essential bandwidth is shown in Figure 4.2 using a synthetic scene with parameters $z_{\min} = 1.5 \text{ m}$ and $\omega_s = 1600\pi \text{ rad/m}$, and assuming cameras with

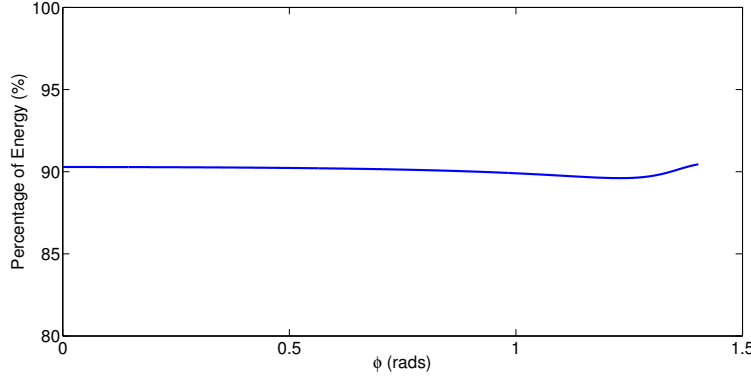


Figure 4.2: Graph validating the essential bandwidth for the signal expressed in (4.20). It shows the percentage of the signal’s energy that is contained within the region defined in (4.26) as the angle ϕ varies. The angle increases from zero to just under the no-occlusion constraint.

a FFoV $\approx 19^\circ$. The figure shows that approximately 90% of the energy of (4.20) is within the region (4.26) as the angle ϕ increases from zero to just under the no-occlusion constraint.

Finally, by adding (4.26) to (4.19), we obtain the following overall bound on ω_v for the plenoptic spectrum in (4.15)

$$\begin{aligned} \frac{z_{\min}\Omega_s}{f \cos(\phi) - v_m |\sin(\phi)|} - n(\phi, \bar{v}_m) \frac{\pi}{v_m} &\leq \omega_v \leq \frac{z_{\max}\Omega_s}{f \cos(\phi) - v_m |\sin(\phi)|} + n(\phi, \bar{v}_m) \frac{\pi}{v_m} \\ \Omega_t \frac{z_{\min}}{f} - n(\phi, \bar{v}_m) \frac{\pi}{v_m} &\leq \omega_v \leq \Omega_t \frac{z_{\max}}{f} + n(\phi, \bar{v}_m) \frac{\pi}{v_m}, \end{aligned} \quad (4.27)$$

where the second inequality follows from the definition of Ω_t in (4.14). Taking the maximum of this bound, we obtain

$$\Omega_v = \Omega_t \frac{z_{\max}}{f} + n(\phi, \bar{v}_m) \frac{\pi}{v_m}. \quad (4.28)$$

Notice that $n(\phi, \bar{v}_m) \in [1, 1.625]$, since $n(\phi, \bar{v}_m) = 1$ when $\phi = 0^\circ$ and approaches 1.625 as ϕ approaches the limit caused by the no-occlusion constraint.

Determining the z_G Parameter

The parameter z_G is determined as follows. From the definition of the parametric model in (4.1), if we set $\omega_t = \Omega_t$ and $\omega_v = \Omega_v$ we obtain the following

$$\Omega_v = \Omega_t \frac{z_{\max}}{f} + n(\phi, \bar{v}_m) \frac{\pi}{v_m} = \frac{z_G}{f} \left(\Omega_t + \frac{A}{2} \right). \quad (4.29)$$

Now, if we still assume $\omega_t = \Omega_t$ but set ω_v equal to the lower bound expressed in (4.27), we obtain

$$\Omega_t \frac{z_{\min}}{f} - n(\phi, \bar{v}_m) \frac{\pi}{v_m} = \frac{z_G}{f} \left(\Omega_t - \frac{A}{2} \right). \quad (4.30)$$

Therefore, by solving these equations, the parameter z_G is

$$z_G = \frac{z_{\max} + z_{\min}}{2}. \quad (4.31)$$

Determining the A Parameter

In a similar fashion, the parameter A is also determined from (4.30) and (4.29). By subtracting (4.30) from (4.29) we obtain

$$\frac{\Omega_t}{f} (z_{\max} - z_{\min}) + n(\phi, \bar{v}_m) \frac{2\pi}{v_m} = A \frac{z_G}{f}. \quad (4.32)$$

Therefore the parameter is

$$A = \frac{\Delta z}{z_G} \Omega_t + n(\phi, \bar{v}_m) \frac{2\pi f}{v_m z_G}, \quad (4.33)$$

where $\Delta z = z_{\max} - z_{\min} = T |\sin(\phi)|$ is the depth variation of the plane.

4.2.2 Validation of the Essential Bandwidth EB_S

Having defined each parameter for the essential bandwidth in (4.1), we want to verify that the region always contains approximately 90% of the signal's energy. For this validation, we use a synthetic scene comprising a single slanted plane with bandlimited texture. From Section 3.4, we observe that ϕ and ω_s have the greatest effect on the

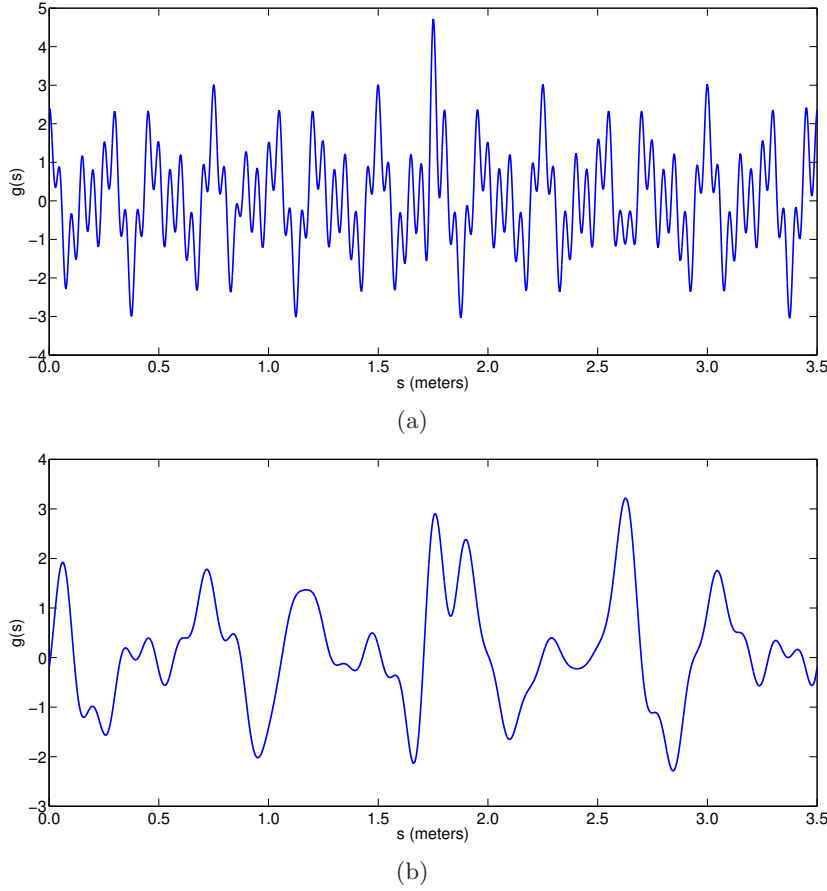


Figure 4.3: Graphs illustrating the bandlimited signal used as texture for the synthetic scenes. The signal in graph (a) has a maximum frequency of $\omega_s = 40\pi$ rad/m whereas the signal in graph (b) has a maximum frequency of $\omega_s = 50$ rad/m. In both cases the signals are plotted against a curvilinear coordinate $s \in [0, 3.5\text{m}]$.

structural model of the plenoptic spectrum. Therefore, the validation is performed by varying ϕ and ω_s whilst the other parameters remain fixed. The angle ϕ varies from 0 rads to just under the limit imposed by (3.3), and ω_s from 0 rad/m to 600 rad/m. Figure 4.3 illustrates an example of the texture signal for two different values of ω_s . The fixed parameters are $x_1 = 0$ m, $z_1 = 1.5$ m, $T = 3.5$ m, and cameras with a focal length of 32 mm and a FoV = 40° .

The results of the validation are shown in Figure 4.4. The figure presents a 2D plot of the percentage of energy inside the essential bandwidth, EB_S , as the angle ϕ varies along the y -axis and ω_s along the other axis. The validation shows that for all values of ω_s and ϕ the essential bandwidth defined in (4.1) always contains at least 89% of the energy of the plenoptic function, and that above 90% is achieved when $\phi > 0$ rads

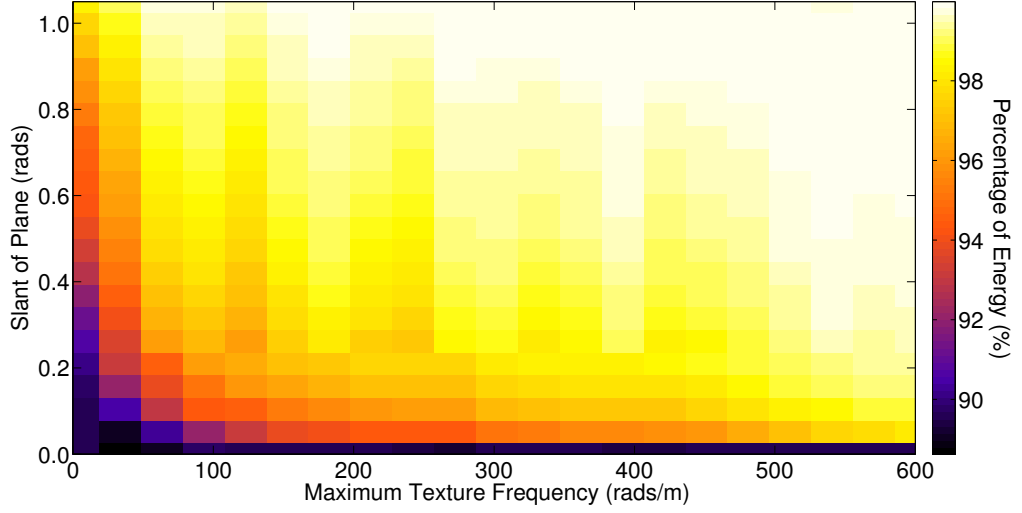


Figure 4.4: Graph showing the variation in the percentage of energy within the essential bandwidth for the plenoptic spectrum generated by a slanted plane with bandlimited texture. Changes in the angle of slant of the plane, ϕ , are illustrated along the y -axis and changes in the maximum frequency of the texture signal, ω_s , along the x -axis. The other parameters are as follows; $x_1 = 0$ m, $z_{\min} = 1.5$ m and $T = 3.5$ m. The cameras have a focal length of 32 mm, corresponding to a FoV $\approx 40^\circ$.

and $\omega_s > 0$ rads/m. The figure also illustrates that the essential bandwidth is very conservative for larger parameter values.

4.3 Sampling and Reconstructing the Plenoptic Function

In order to sample and reconstruct, we assume the bandwidth of the plenoptic function for a slanted plane is given by the parametric model defined in (4.1). By approximating the bandwidth in this way, we can determine an expression for the maximum spacing between adjacent cameras and a new parametrisation of the reconstruction filter.

Starting with the sampling process, the maximum camera spacing is inversely proportional to the Nyquist sampling density in the spatial domain (i.e. the density of the cameras). If we assume the bandwidth of the plenoptic function is equal to (4.1), then this Nyquist camera density is equivalent to A - the width of the (4.1) in ω_t . Therefore the expression for the maximum spacing between the cameras for a slanted plane is

$$\Delta t_G = \frac{2\pi}{A} = \frac{2\pi z_G v_m}{v_m \Omega_t \Delta z + 2\pi n(\phi, \bar{v}_m) f}. \quad (4.34)$$

This expression differs from that presented by Chai et al [10]. Using (4.28) as the maximum value in ω_v , their equivalent expression for the maximum camera spacing of a slanted plane is

$$\Delta t_C = \frac{2\pi z_{\max} z_{\min} v_m}{\Delta z(v_m z_{\max} \Omega_t + \pi n(\phi, \bar{v}_m)f)}. \quad (4.35)$$

To reconstruct the sampled plenoptic function, we use a filter similar to that presented in [10]. This filter comprises three elements: the width of the filter in ω_t , which is determined by the spacing between each camera Δt ; the width of the region in ω_v , which is determined by the finite pixel resolution of the cameras Δv ; and skew of the filter z_{tilt}/f . In frequency this filter is defined as follows

$$\Psi(\omega_t, \omega_v) = \begin{cases} 1, & \text{if } \omega_t, \omega_v \in \mathcal{R}_\psi \\ 0, & \text{else,} \end{cases} \quad (4.36)$$

where \mathcal{R}_ψ is the region of support in frequency given by

$$\mathcal{R}_\psi = \left\{ \omega_t, \omega_v : \omega_v \in \left[-\frac{\pi}{\Delta v}, \frac{\pi}{\Delta v} \right], \omega_t \in \left[\frac{\omega_v f}{z_{\text{tilt}}} - \frac{\pi}{\Delta t}, \frac{\omega_v f}{z_{\text{tilt}}} + \frac{\pi}{\Delta t} \right] \right\}. \quad (4.37)$$

Our filter differs from Chai et al's [10], however, in the value of the parameter z_{tilt} . In their analysis z_{tilt} is set equal to z_C , which is defined in (2.19). In our case, as we approximate the bandwidth of the plenoptic function with its essential bandwidth, z_{tilt} is determined by the parametric model in (4.1). Therefore we set z_{tilt} equal to z_G and obtain a new parametrisation for the corresponding reconstruction filter.

Having determined this camera spacing and this reconstruction filter, we now assess their validity by analysing the reconstruction of six synthetic EPIs at differing camera densities. The synthetic EPIs in question correspond to six scenes. The scenes are generated from a combination of three different geometries and two different textures. The three geometries are all related to a single slanted plane with the following parameters in common; $x_1 = 0$ m, $z_{\min} = 2.1$ m and $T = 3.5$ m. The difference between the three is the angle of slant. The first plane has a slant of $\phi = 3^\circ$, the second a slant of $\phi = 30^\circ$ and, finally, the third has $\phi = 60^\circ$. For the texture signals, one is bandlimited to

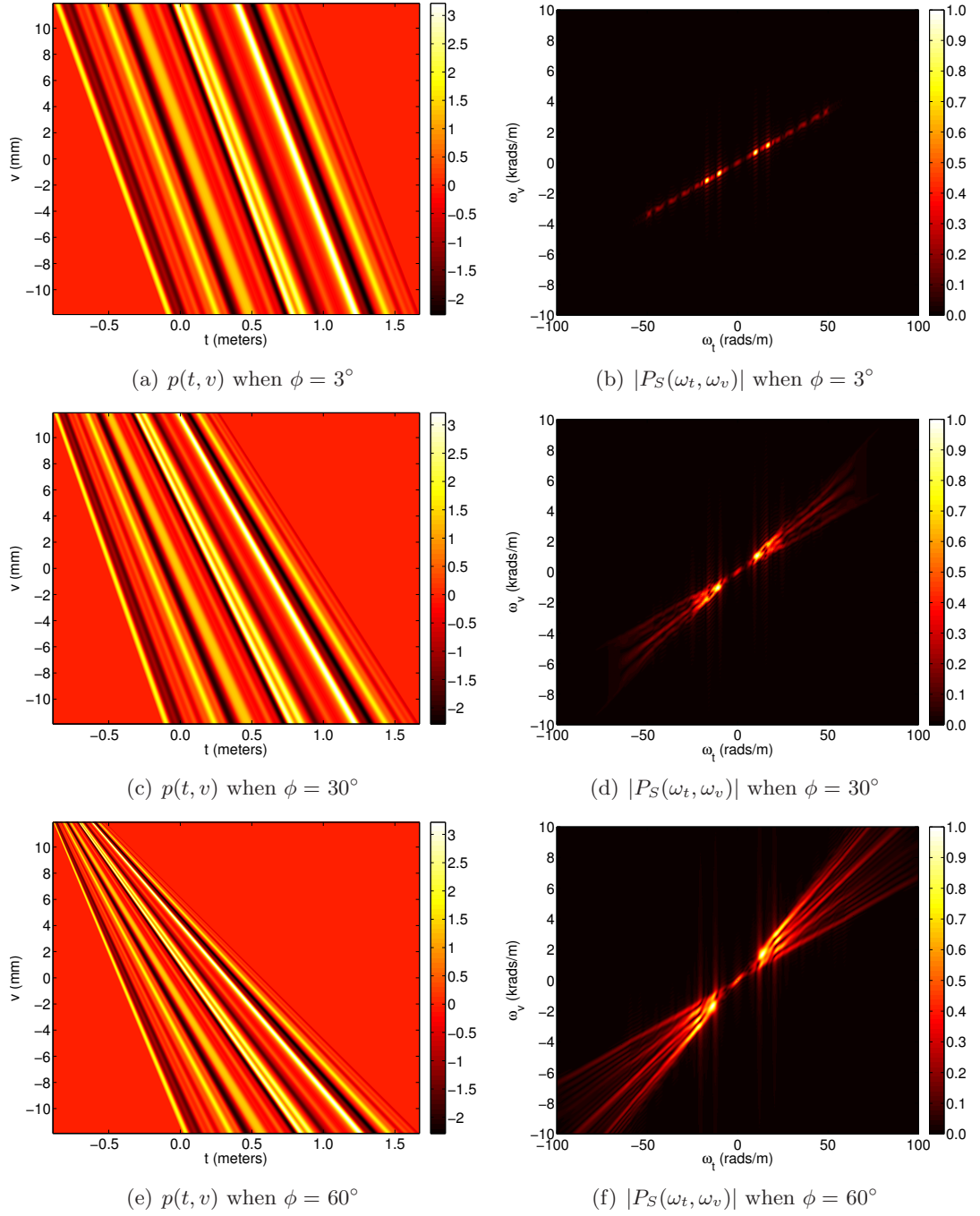


Figure 4.5: Diagrams showing the synthetic EPIs and plenoptic spectra for three different slanted planes. The EPIs, $p(t, v)$, are shown in (a), (c) and (e), and their plenoptic spectra, $|P_S(\omega_t, \omega_v)|$, in (b), (d) and (f), respectively. The three planes have the following parameters in common; $x_1 = 0$ m, $z_{\min} = 2.1$ m, $T = 3.5$ m and $\omega_s = 50$ rads/m. In (a) and (b) $\phi = 3^\circ$, (c) and (d) $\phi = 30^\circ$, and (e) and (f) $\phi = 60^\circ$.

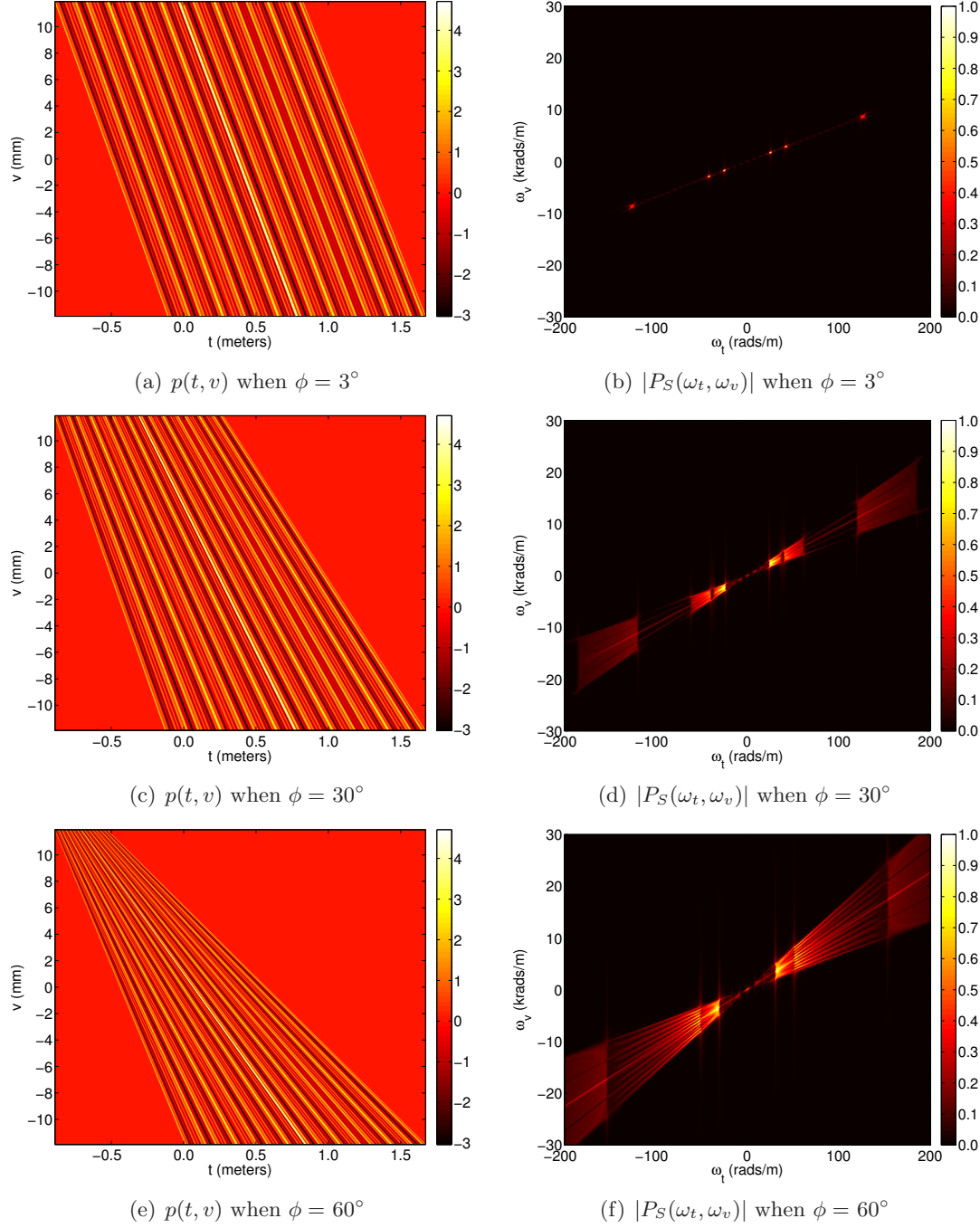


Figure 4.6: Diagrams showing the synthetic EPIs and plenoptic spectra for three different slanted planes. The EPIs, $p(t, v)$, are shown in (a), (c) and (e), and their plenoptic spectra, $|P_S(\omega_t, \omega_v)|$, in (b), (d) and (f), respectively. The three planes have the following parameters in common; $x_1 = 0$ m, $z_{\min} = 2.1$ m, $T = 3.5$ m and $\omega_s = 125$ rads/m. In (a) and (b) $\phi = 3^\circ$, (c) and (d) $\phi = 30^\circ$, and (e) and (f) $\phi = 60^\circ$.

$\omega_s = 50 \text{ rad/m}$ and the other to $\omega_s = 125 \text{ rad/m}$. They are both illustrated in Figure 4.3. Note that the texture bandwidth is uniform, i.e. it is the same for the entire scene. The cameras are defined with a focal length equal to 32mm and a FoV $\approx 40^\circ$. The synthetic EPIs generated when $\omega_s = 50 \text{ rad/m}$ are shown in Figure 4.5. Likewise, Figure 4.6 shows the synthetic EPIs generated when $\omega_s = 125 \text{ rad/m}$. Both figures also show the corresponding plenoptic spectrum for each EPI.

The PSNR curves for the reconstruction of these synthetic EPIs are shown in Figure 4.7 and Figure 4.8. The graphs in Figure 4.7 and 4.8 correspond respectively to the low and high frequency textures. The slant of the plane, and in turn the depth variation Δz , ascends from the smallest in graph (a) to the largest in (c) for each figure. Lastly, there are two PSNR curves plotted in each graph: the first, the solid blue line, represents the PSNR curve obtained using our reconstruction filter with a skew z_G/f . The second, the red dashed line, is the PSNR curve generated using the filter proposed in [10] with a skew z_C/f . In the following sections we analyse these graphs in terms of the maximum camera spacing and the reconstruction filter.

4.3.1 Analysing the Maximum Camera Spacing

The minimum number of cameras predicted by (4.34) and (4.35) are respectively N_G and N_C . These two values are indicated by vertical dashed lines in Figures 4.7 and 4.8. Comparing N_G to N_C , we see that our approach results in a more conservative sampling when the depth variation is small, see Figures 4.7(a) and 4.8(a). However this relationship is reversed as the depth variation Δz increases. The reason for this is that N_G is constrained by the FFoV of the cameras. Consequently, for small depth variation, N_G is restricted such that there are enough cameras to cover the entire scene. In contrast, N_C is not constrained in this manner thus it can tend to 1 as the depth variation tends to 0m. For larger depth variation, our estimation of the bandwidth of the plenoptic function is less conservative than that proposed in [10], hence N_G is less than N_C .

With the exception of the scenes where $\phi = 3^\circ$, the figures show that our prediction of the Nyquist number of cameras, N_G , occurs approximately at the ‘elbow’ of each

PSNR curve and that the PSNR at this point is high. In these cases, therefore, (4.1) represents a good approximation of the bandwidth for the plenoptic function. However, in Figures 4.7(a) and 4.8(a), N_G is slightly too small resulting in low PSNR values for the reconstruction of the EPIs. This suggests that a larger approximation of the bandwidth is required when the depth variation is small. It is worth noting, however, that this PSNR is better than that given by N_C in the same situation. As a result, our maximum camera spacing, Δt_G , offers a more accurate prediction of the critical sampling point than Δt_C in all cases.

4.3.2 Analysing the Reconstruction Filter

The graphs in Figure 4.7 and Figure 4.8 also compare the performance of two filters when reconstructing the synthetic EPIs. The filters both have the same parametric shape defined in (4.36). They differ, however, in their skew: our filter has a skew of z_G/f and the filter proposed in [10] has a skew of z_C/f . With the exception of Figures 4.7(a) and 4.8(a), our reconstruction filter outperforms that proposed in [10] for all camera densities. The gain in performance, moreover, increases with the depth variation within the scene. In Figures 4.7(a) and 4.8(a), the depth variation is so small that $z_G \approx z_C$. As a result, the PSNR curves are equivalent in these figures.

To explore the effect of the filter skew, we extend our analysis to examine the reconstruction of the plenoptic functions for a wider range of filter skew values. Using the three geometries defined earlier with the high frequency texture signal shown in Figure 4.3(a), we reconstruct the corresponding synthetic EPIs using a filter with a skew that varies from z_{\min}/f to z_{\max}/f . The results of this filter analysis are shown in Figure 4.9. The figure presents three 2D plots in which the number of cameras varies along the x -axis and the filter skew varies along the y -axis. The intensity in each image represents the difference in PSNR when that reconstruction is compared to the reconstruction achieved using z_G/f . Therefore a negative value indicates that our reconstruction filter, with skew z_G/f , outperforms any other filter. The slant of the plane, and in turn the depth variation Δz , ascends from the smallest in image (a) to the largest in (c).

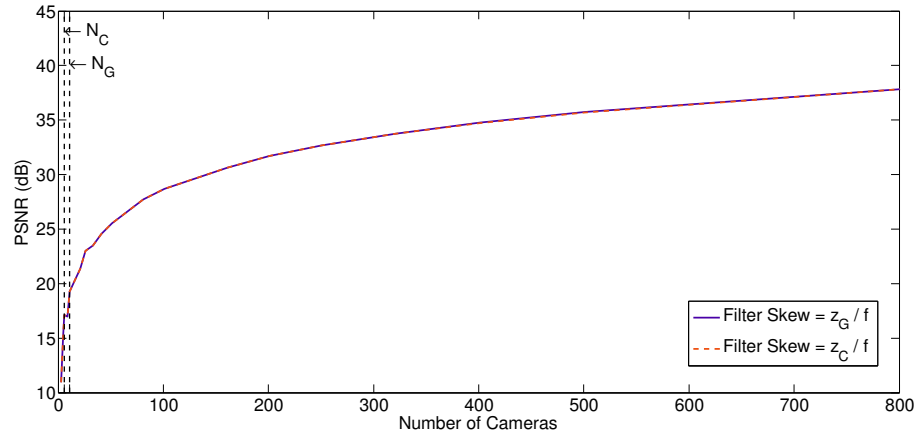
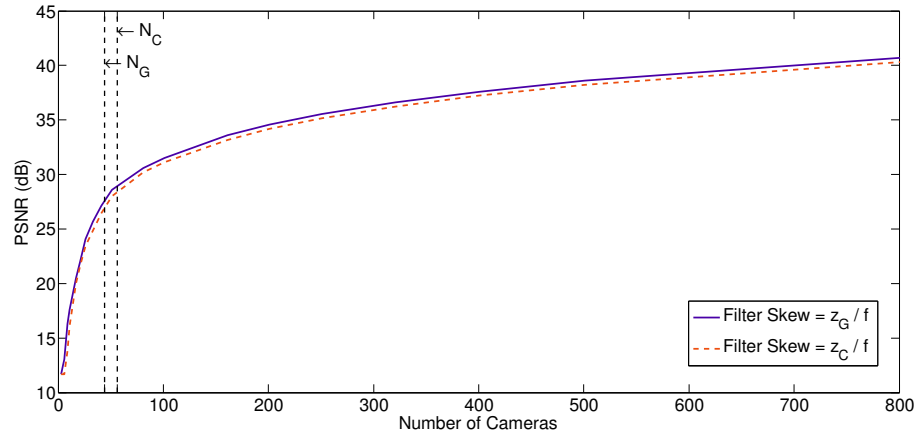
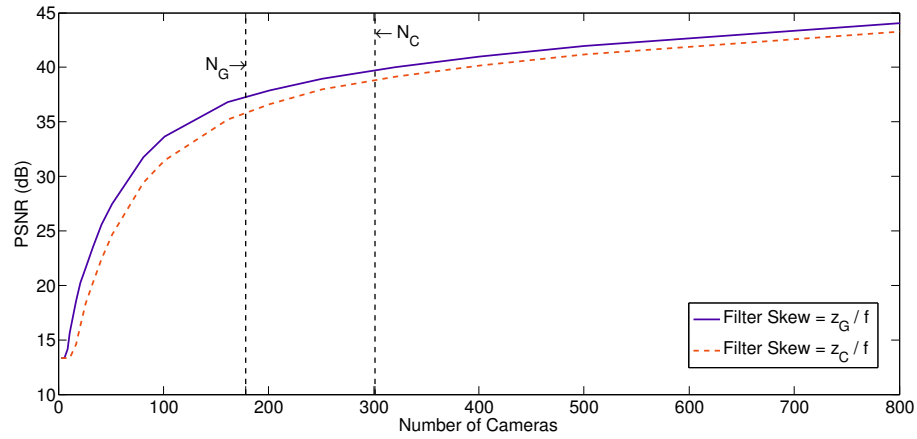
(a) $\phi = 3^\circ$ and $\Delta z = 0.18$ m(b) $\phi = 30^\circ$ and $\Delta z = 1.75$ m(c) $\phi = 60^\circ$ and $\Delta z = 3.03$ m

Figure 4.7: Graphs comparing the PSNR curves for the reconstruction of three EPIs using two different reconstruction filters. Each EPI relates to a slanted plane with the following parameters in common; $x_1 = 0$ m, $z_{\min} = 2.1$ m, $T = 3.5$ m and $\omega_s = 50$ rads/m. In (a) $\phi = 3^\circ$, (b) $\phi = 30^\circ$ and (c) $\phi = 60^\circ$. The filters differ in their skew; the first has a skew of z_G / f and the second z_C / f . N_G and N_C are the number of cameras required to achieve Nyquist sampling for each reconstruction method.

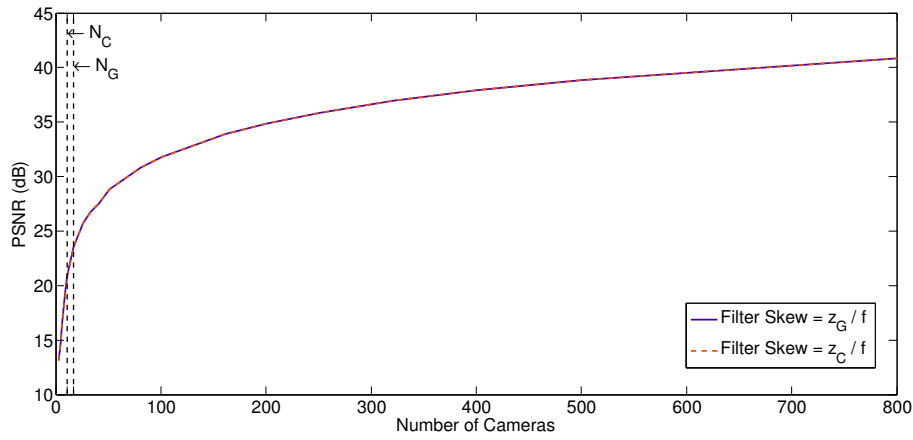
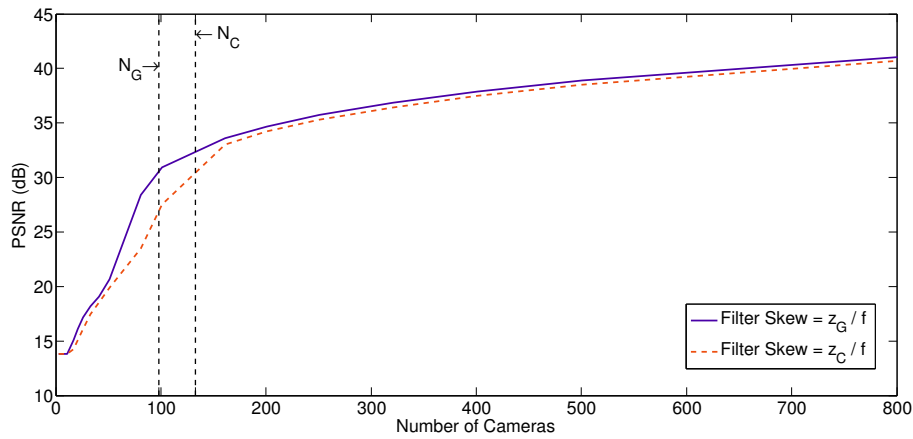
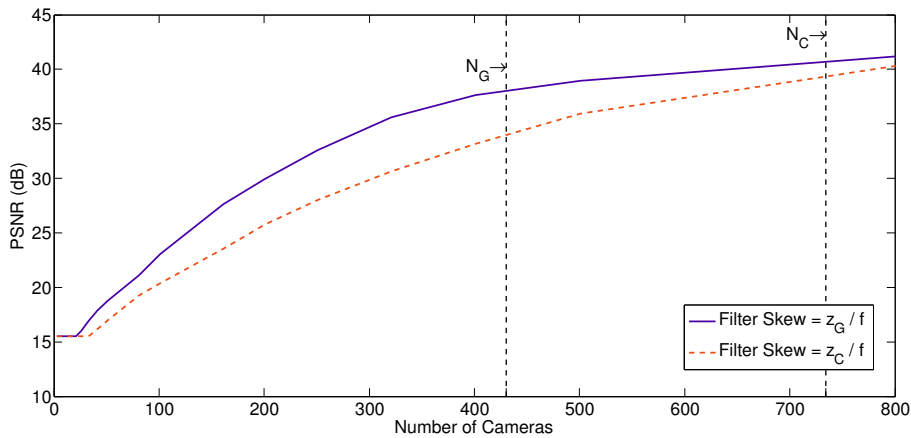
(a) $\phi = 3^\circ$ and $\Delta z = 0.18, \text{m}$ (b) $\phi = 30^\circ$ and $\Delta z = 1.75 \text{ m}$ (c) $\phi = 60^\circ$ and $\Delta z = 3.03 \text{ m}$

Figure 4.8: Graphs comparing the PSNR curves for the reconstruction of three EPIs using two different reconstruction filters. Each EPI relates to a slanted plane with the following parameters in common; $x_1 = 0 \text{ m}$, $z_{\min} = 2.1 \text{ m}$, $T = 3.5 \text{ m}$ and $\omega_s = 125 \text{ rads/m}$. In (a) $\phi = 3^\circ$, (b) $\phi = 30^\circ$ and (c) $\phi = 60^\circ$. The filters differ in their skew; the first has a skew of z_G/f and the second z_C/f . N_G and N_C the number of cameras required to achieve Nyquist sampling for each reconstruction method.

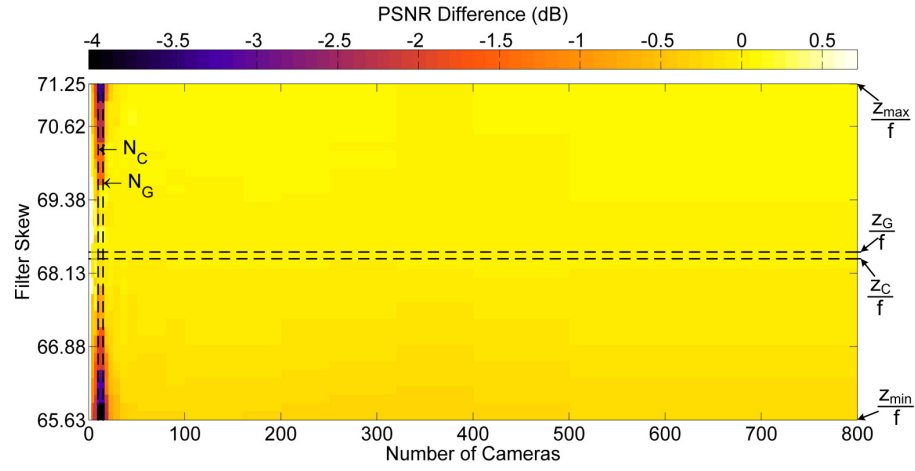
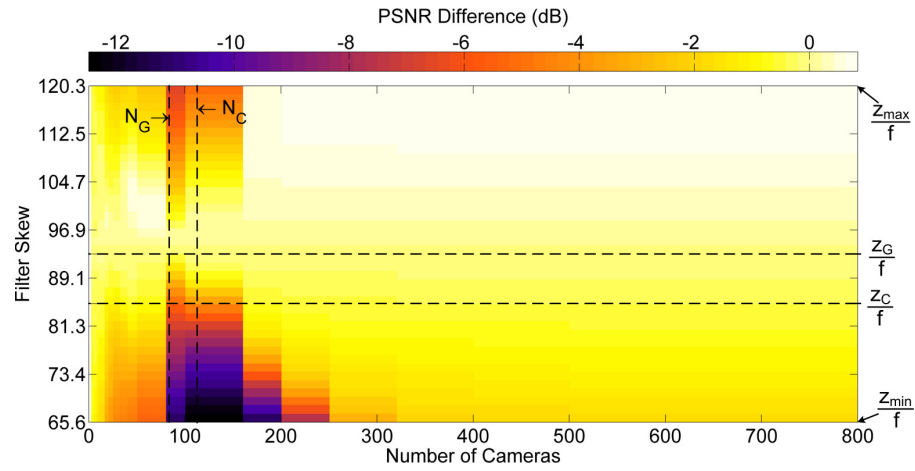
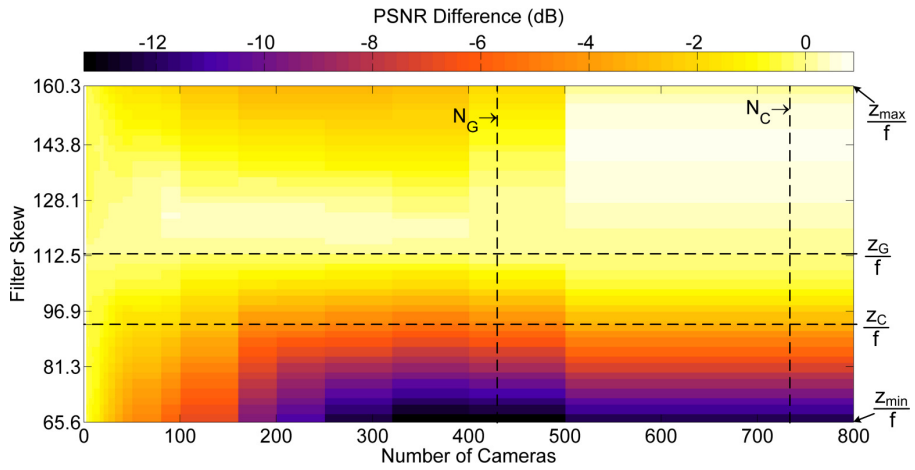
(a) $\phi = 3^\circ$ and $\Delta z = 0.18$ m(b) $\phi = 30^\circ$ and $\Delta z = 1.75$ m(c) $\phi = 60^\circ$ and $\Delta z = 3.03$ m

Figure 4.9: Graphs comparing the reconstruction of three EPIs as the skew of the reconstruction filter varies from z_{\min}/f to z_{\max}/f . The intensity in each image represents the difference in PSNR relative to the reconstruction when the skew is z_G/f . Each EPI relates to a slanted plane with the following parameters in common; $x_1 = 0$ m, $z_{\min} = 2.1$ m, $T = 3.5$ m and $\omega_s = 125$ rads/m.

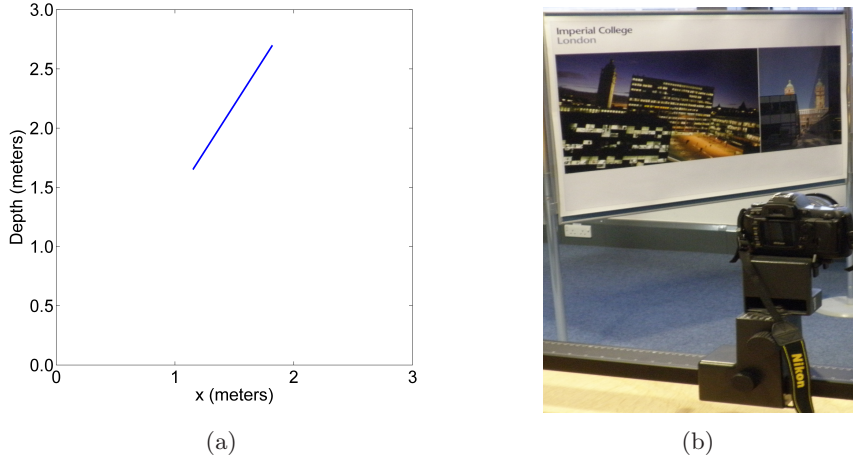


Figure 4.10: The set-up and data acquisition for the EPI-volume corresponding to a scene with a single slanted plane. Diagram (a) illustrates the bird’s eye view of the scene geometry and (b) shows the data acquisition. The resulting EPI-volume consists of 133 images spaced 1 cm apart. Each image is 3008 by 1888 pixels.

Figure 4.9 shows that across all three scenes our filter parametrisation is one of the best for most of the camera densities. However the figure also illustrates that oversampling will compensate a poor choice of filter skew. For instance when the depth variation is 0.18m, as shown in Figure 4.9(a), our filter is optimum up to 30 cameras after which oversampling compensates and there is little distinction between the filters. There is also little distinction when the scenes are undersampled. This is apparent in Figures 4.9(b) and 4.9(c) where the depth variation in the scene is greater. These figures also highlight an exception in which our filter is sub-optimal. This exception occurs when the camera density is high; above 150 cameras in Figure 4.9(b) and above 500 cameras in 4.9(c). In these cases the optimum filter skew is very close to or equal to z_{\max}/f . The reason this skew is optimal is that the width of the filter in frequency, $2\pi/\Delta t$, approaches $2\Omega_t$. Therefore a square reconstruction filter is preferable to a skewed filter. In practise, however, we aim to use fewer cameras hence our parametrisation of the reconstruction filter is the best in such cases.

4.4 Rendering Real Images

In Section 4.3 we determined a new expression for the maximum camera spacing for a slanted plane and a new parametrisation of the reconstruction filter. These expressions were then analysed using synthetic scenes. In this section we extend our analysis to real scenes. As such we examine the sampling and reconstruction of an EPI-volume generated from 133 images, each 3008 by 1888 pixels in size. The images are captured at 1cm intervals along the camera line. The scene in question comprises a single slanted plane with the following parameters; $x_1 = 1.15$ m, $z_{\min} = 1.65$ m, $T = 1.246$ m and $\phi = 57.5^\circ$. An urban picture is used as the texture pasted to the plane. The images are acquired using a camera with 70mm focal length, which corresponds to a FoV $\approx 20^\circ$. Figure 4.10 illustrates a bird's eye view of the scene geometry and data acquisition set-up used to obtain the images. This real scene is constrained such that the depth only varies with x , hence the EPI-volume can be treated as a set of 2D EPIs stacked together. Consequently we reconstruct the EPI-volume in 2D slices using symmetric extension.

Figure 4.11(a) compares the reconstruction of the whole EPI-volume using our filter and the reconstruction achieved using the filter proposed in [10]. The figure shows that our filter still outperforms the other for all camera densities, however the gain in PSNR is smaller than that shown in the results for the synthetic EPIs. A similar result is obtained when comparing a wider range of filter skews, see Figure 4.11(b). In a similar manner to Figure 4.9, the image in Figure 4.11(b) plots the difference in PSNR when that reconstruction of the EPI-volume is compared to the reconstruction achieved using z_G/f . Both figures also illustrate the Nyquist number of cameras relating to Δt_G and Δt_C for the EPI-volume. For this EPI-volume $N_C = 41$ cameras and $N_G = 35$ cameras.

So far, in all of the results we have examined, the reconstruction filter has the following expression in the EPI domain

$$\psi(t, v) = \eta\left(\frac{\pi}{\Delta t}t\right) \eta\left(\frac{\pi}{\Delta v}\left(v + \frac{f}{z_G}t\right)\right), \quad (4.38)$$

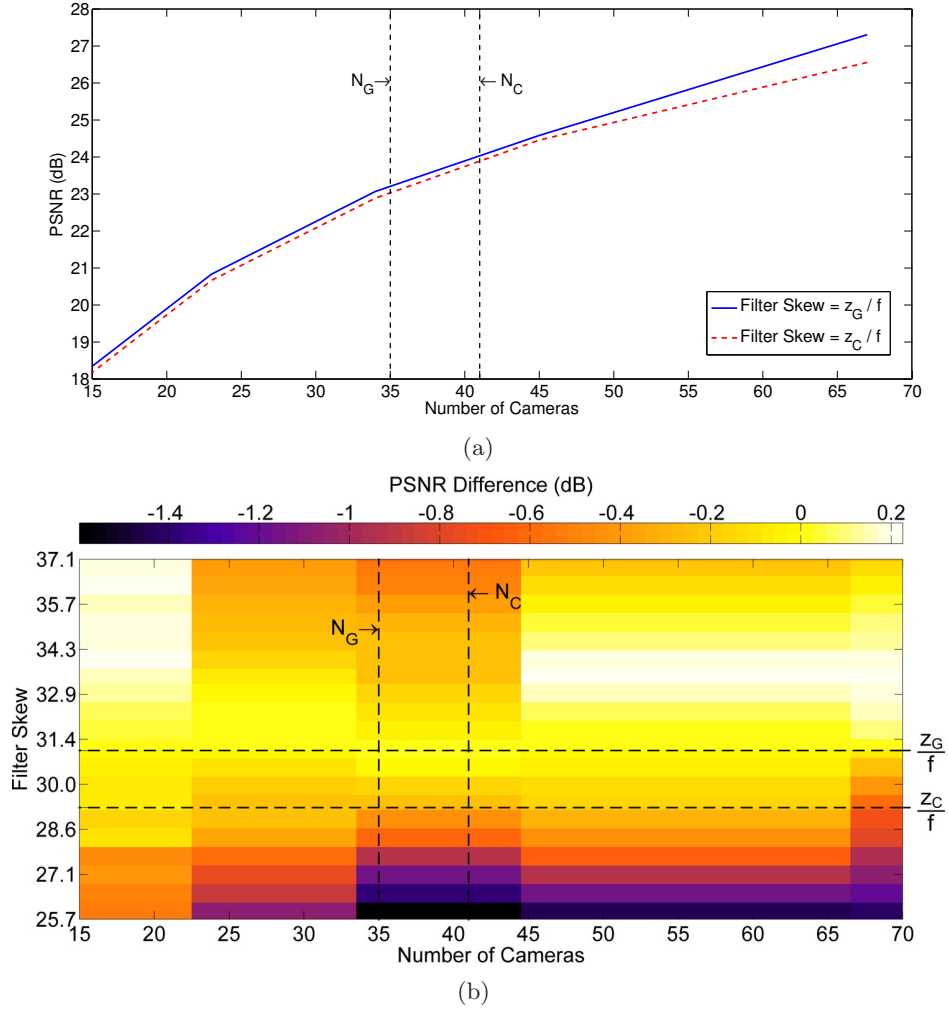


Figure 4.11: Graph (a) shows the PSNR curves for the reconstruction of an EPI-volume, relating to a slanted plane, using a filter skew of z_G/f and z_C/f . Graph (b) compares the reconstruction of the same EPI volume as the skew of the reconstruction filter varies from 25.7 to 37.1. The intensity in each image represents the difference in PSNR relative to the reconstruction when the skew is z_G/f . The parameters for this slanted plane are; $x_1 = 1.15$ m, $z_{\min} = 1.65$ m, $T = 1.246$ m and $\phi = 57.5^\circ$. The cameras have a focal length of 70 mm, which corresponds to a FoV $\approx 20^\circ$. The Nyquist number of cameras required for Δt_G and Δt_C are indicated with N_G and N_C , respectively.

where the interpolation function η is the sinc function. However the sinc function has infinite support and decays very slowly. Therefore Gibbs effects occur when reconstructing a signal with discontinuities. In terms of reconstructing an EPI, we highlighted in Section 2.2.2 the inherent structure caused by mapping a point in the scene to a line in the EPI. Consequently, reconstructing the EPI with a constant filter skew is equivalent to assuming all the lines have the same gradient. As a result trajectories within the

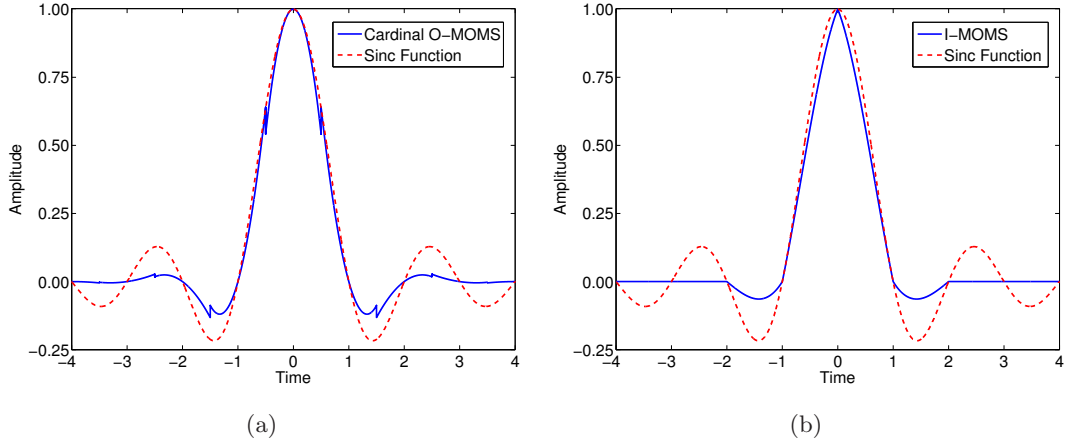


Figure 4.12: Graph (a) compares a 2nd order cardinal O-MOMS to the sinc function. Graph (b) compares a 3rd order I-MOMS to the sinc function. Notice that the O-MOMS has infinite support and discontinuities, whereas the I-MOMS is of compact support.

EPI relating to different depths are incorrectly reconstructed. The by-product of this incorrect reconstruction is the occurrence of ringing artefacts in rendered images. The solution to this problem is to use an interpolating function that decays more quickly than the sinc function.

The optimum function fitting this criteria is constructed from a class of functions known as Maximal-Order-Minimal-Support (MOMS) [6]. MOMS are functions that have minimal support for a given accuracy or approximation order. They are made from a weighted combination of a B-spline and its derivatives. Therefore a MOMS of order M has the highest approximation order, $M + 1$, for the smallest possible support size (the size of the support is $M + 1$). Its definition is

$$\varphi_{\text{MOMS}}^{\{M\}}(x) = \sum_{n=0}^{M+1} \lambda_n \frac{d^n}{dx^n} \left\{ \beta_M(x) \right\}, \quad (4.39)$$

where λ_n are the weights and $\beta_M(x)$ is a B-spline of order M . For a detailed survey of the properties of B-splines see [61]. Depending on the value of the weights, multiple MOMS functions can be constructed from this definition. Notice that if $\lambda_0 = 1$ and $\lambda_n = 0 \forall n > 0$ then we are left with a B-spline, hence B-splines themselves are also MOMS.

In [6], Blu et al. derived a set of weights that minimised the \mathbf{L}^2 error between an arbitrary function and its interpolation. They called the resulting MOMS function the Optimal MOMS (O-MOMS). Therefore, assuming a compact support of $M+1$, the M^{th} order O-MOMS is the optimum interpolator in terms of the \mathbf{L}^2 error. The O-MOMS, however, are designed to interpolate a function from a set of coefficients, which may not necessarily be equal to its samples. Consequently we alter the O-MOMS such that the coefficients always coincide with the samples. This altered O-MOMS is known as its cardinal. The cardinal O-MOMS is thus

$$\eta_O^{\{M\}}(x) = \mathcal{C} \left\{ \varphi_O^{\{M\}}(x) \right\}, \quad (4.40)$$

where \mathcal{C} is the cardinal operator, see Appendix E for details. A drawback of (4.40) is that it now has infinite support. To fix this, Blu et al. derived a MOMS of compact support that required the coefficients to be equal to the samples of the function, termed the Interpolating MOMS (I-MOMS). The trade off in designing the I-MOMS, however, is that it is a suboptimal interpolator when compared to the O-MOMS.

In view of this, we now compare the reconstruction of EPI-volume presented earlier using a 2nd order cardinal O-MOMS and a 3rd order I-MOMS. The cardinal O-MOMS is defined as follows

$$\eta_O^{\{2\}}(x) = \mathcal{C} \left\{ \beta_2(x) + \frac{1}{60} \frac{d^2}{dx^2} \left\{ \beta_2(x) \right\} \right\}, \quad (4.41)$$

and the I-MOMS is

$$\eta_I^{\{3\}}(x) = \beta_3(x) - \frac{1}{6} \frac{d^2}{dx^2} \left\{ \beta_3(x) \right\}. \quad (4.42)$$

The two interpolating functions are illustrated in Figure 4.12 together with the sinc function. The figure highlights that the cardinal O-MOMS is discontinuous (a feature of even ordered O-MOMS) and has infinite support although it decays much more rapidly than the sinc function. In contrast the I-MOMS has compact support and is continuous.

The PSNR curves for the reconstructed EPI-volume generated by these functions

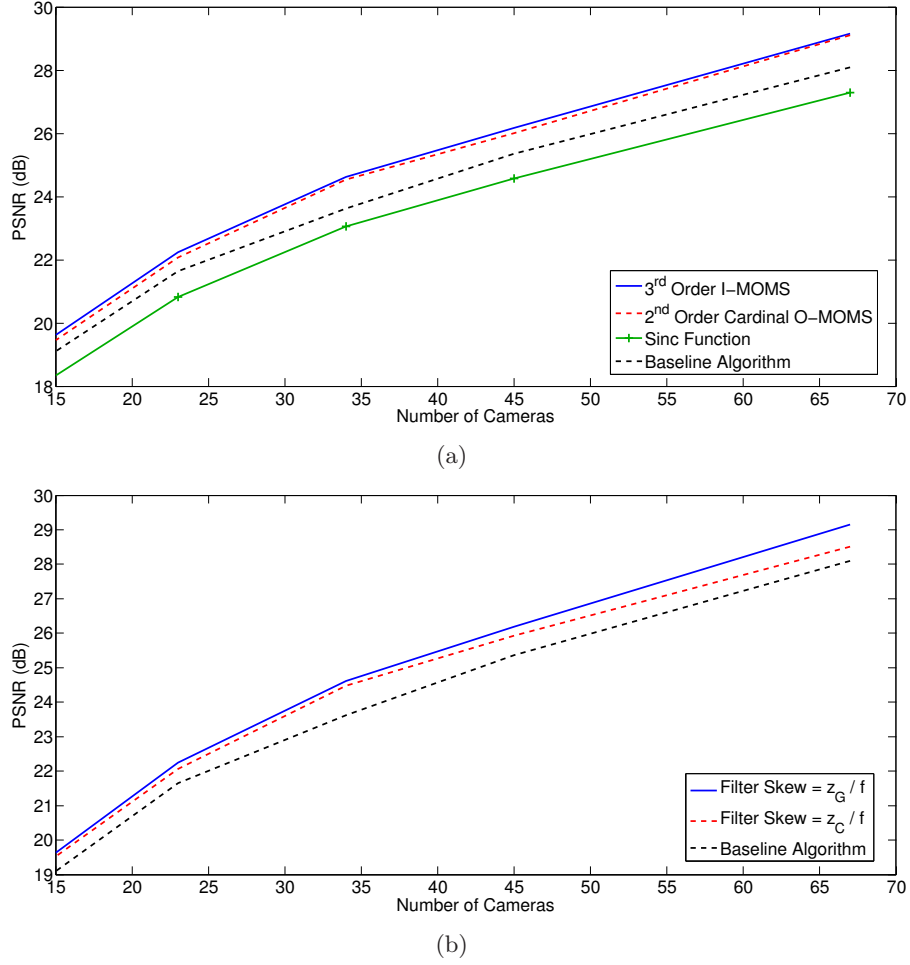


Figure 4.13: Graph (a) compares the PSNR curves for the reconstruction of the EPI-volume obtained using a reconstruction filter with a 3rd order I-MOMS, a 2nd order cardinal O-MOMS and a sinc function. The skew of the filter is z_G/f in all three cases. The graph also includes the PSNR curve obtained using the baseline algorithm with a constant depth of z_G . Graph (b) shows the PSNR curves for the reconstruction of the same EPI-volume using the I-MOMS but with differing filter skews, z_G/f and z_C/f . The graph also shows the PSNR curve for the baseline algorithm. The parameters for this slanted plane are; $x_1 = 1.15\text{m}$, $z_{\min} = 1.65\text{m}$, $T = 1.246\text{m}$ and $\phi = 57.5^\circ$. The cameras have a focal length of 70mm, which corresponds to a FoV $\approx 20^\circ$.

are shown in Figure 4.13(a). Using a filter skew of z_G/f , the figure illustrates that the reconstruction achieved using the two MOMS functions outperforms that achieved using the sinc function. The gain in performance is approximately 1dB for all camera densities. Comparing just the two MOMS functions, we see that the 3rd order I-MOMS marginally outperforms the 2nd order O-MOMS. To put these results in perspective, we also reconstruct the EPI-volume using a baseline algorithm and include the results

in Figure 4.13(a). This algorithm uses the linear interpolation method, outlined in Section 2.3.3, to reconstruct the EPI-volume. In other words, it simply performs linear interpolation along the EPI lines assuming a constant depth of z_G . Figure 4.13(a) shows that the reconstruction achieved using either of the MOMS functions outperforms the baseline algorithm. However, the algorithm is an improvement on the reconstruction achieved using the sinc function. This is to be expected as the baseline algorithm uses linear interpolation so does not suffer from Gibbs effects like the sinc function.

We illustrate the visually difference between the four reconstruction methods in Figure 4.14. The figure presents a section of an image rendered using the three different interpolating functions and the baseline algorithm. The original image is shown in Figure 4.14(a), its rendering using a sinc function in Figure 4.14(b), using the baseline algorithm in Figure 4.14(c), using the 2nd order O-MOMS in Figure 4.14(d) and the 3rd order I-MOMS in Figure 4.14(e). The images are rendered using only 34 of the original cameras and the corresponding PSNR values, for the whole rendered image, are (b) 23.6dB, (c) 23.9dB (d) 24.1db and (e) 24.2dB. Therefore, as predicted, the rendered images show that the MOMS interpolators suppress the ringing artefacts, thus improving the reconstruction results. The suppression, however, is slightly better when using the I-MOMS due to its compact support hence the slight gain in PSNR. The baseline algorithm also suppresses some of the ringing artefacts. Notice that, in all four cases, the ringing artefacts are worse on the right half of the rendered image. This effect is caused by the scene depth increasing as you move from the left to the right in the image. In the left side of the image, the scene is at a depth z_G ; hence the choice of z_G as a rendering depth (or z_G/f as a filter skew) matches the scene depth. However, the mismatch between the scene depth and the rendering depth gradually gets worse as you move to the right, which results in worse artefacts.

To complete this analysis, we compare z_G/f to z_C/f when using the MOMS interpolators. The results are presented in Figure 4.13(b) along with the baseline algorithm. They show that our filter skew results in a better reconstruction than z_C/f when using the I-MOMS. To illustrate the degradation in performance caused by using z_C/f , Figure 4.15 re-renders the images shown Figure 4.14 with z_C/f as the filter skew. In this



(a) Original



(b) Sinc Function, PSNR = 23.6dB



(c) Baseline Algorithm, PSNR = 23.9dB

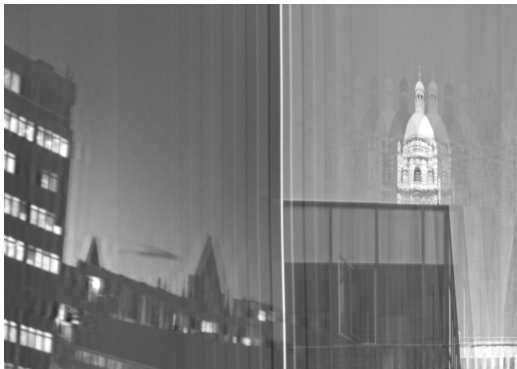
(d) 2nd Order O-MOMS, PSNR = 24.1dB(e) 3rd Order I-MOMS, PSNR = 24.2dB

Figure 4.14: An example of a rendered image from the EPI-volume using four different reconstruction methods. The original image is shown in (a), it's rendering using a sinc function in (b), using the baseline algorithm in (c), using a 2nd order O-MOMS in (d) and lastly, using a 3rd order I-MOMS in (e). The filter skew in each case is z_G/f . Each image is rendered using 34 cameras.

case the corresponding PSNR values for the whole rendered image are now (b) 22.4dB, (c) 22.9dB, (d) 23.1db and (e) 23.2dB. Consequently, the incorrect choice of the filter skew results in larger ringing artefacts in the rendered images.



(a) Original



(b) Sinc Function, PSNR = 22.4dB



(c) Baseline Algorithm, PSNR = 22.9dB

(d) 2nd Order O-MOMS, PSNR = 23.1dB(e) 3rd Order I-MOMS, PSNR = 23.2dB

Figure 4.15: An example of a rendered image from the EPI-volume using four different reconstruction methods. The original image is shown in (a), it's rendering using a sinc function in (b), using the baseline algorithm in (c), using a 2nd order O-MOMS in (d) and lastly, using a 3rd order I-MOMS in (e). The filter skew in each case is z_C/f . Each image is rendered using 34 cameras.

4.5 Summary

In this chapter we have studied the uniform sampling and reconstruction of the plenoptic function for a slanted plane with bandlimited texture. To perform this analysis, we

assume the plenoptic function is approximately bandlimited to its essential bandwidth (a region in frequency containing approximately 90% of the signal's energy). We defined this essential bandwidth using a parametric model derived from the plenoptic spectral analysis presented in Chapter 3. The model comprised four parameters; the maximum value in ω_t , the maximum value in ω_v , the width of the region in ω_t and the skew of the region in frequency. Using this parametric model, we were able to derive a new expression for the maximum spacing between two cameras when sampling a slanted plane. We also determined a new filter skew for the reconstruction filter.

Using synthetic and real scenes, we showed that our expression for the maximum camera spacing offered a more accurate prediction of the Nyquist camera density than that proposed in [10]. By comparing the reconstruction of the data sets using differing filter skews, we showed that our filter skew consistently outperformed the other skews. Therefore our filter skew is the most effective when reconstructing the plenoptic function for a slanted plane. Lastly, we showed that the PSNR of the reconstructed EPI-volume is improved when using a MOMS interpolator instead of the sinc function. We demonstrated that this improvement was caused by reducing the ringing artefacts within the rendered images. The optimum reconstruction of the EPI-volume was achieved using a 3rd order I-MOMS.

The key points to take from this chapter are as follows:

- We sample and reconstruct the plenoptic function for a slanted plane with bandlimited texture assuming it is bandlimited to its essential bandwidth.
- The essential bandwidth for the plenoptic function of a slanted plane can be expressed using a 2D model comprising four parameters.
- Using the parametric essential bandwidth, we derived a new expression for the maximum camera spacing relating to a slanted plane and a new value for the skew of the corresponding reconstruction filter.
- Our expression for the maximum camera spacing results in a more accurate Nyquist sampling density for both synthetic and real scenes than that proposed

in [10].

- Our new value for the skew of the reconstruction filter performs better than alternative skews on both synthetic and real EPI data sets.
- The reconstruction of the plenoptic function is improved using MOMS interpolators, such as O-MOMS and I-MOMS, instead of the sinc function.

Chapter 5

Non-Uniform Plenoptic Sampling

5.1 Introduction

In general most scenes encountered in IBR comprise a more complicated scene geometry than a slanted plane. Therefore, in order to avoid undersampling, a conservative camera spacing is employed when uniformly sampling the scene. This increases the likelihood that non-uniform sampling will offer a distinct improvement in the reconstruction of the plenoptic function and thus the rendering quality. In view of this, we examine non-uniform plenoptic sampling of scenes in this chapter.

We focus our analysis on a general class of scene that has a smoothly varying surface, with no-occlusions, and bandlimited texture. Note that the texture bandwidth is uniform across the entire scene. For such scenes, we develop a novel framework that combines the theoretical results from uniform plenoptic spectral analysis with adaptive camera placement. The basis of this framework is that we can model the local geometric complexity of the scene using a sequence of slanted planes. Then, if we assume the texture on each plane has a constant bandwidth (equal to the scene’s texture bandwidth), we can uniformly sample each plane in the model using the analysis presented in Chapters 3 and 4. Therefore, the end result is a piecewise constant sampling of the smoothly varying scene. The adaptive element comes from how we model the scene surface using the sequence of slanted planes. Using this framework, we present an algorithm that non-uniformly positions cameras by optimising the planar model of

the scene surface. The novelty of this algorithm is that the position of the cameras adapts depending on the local geometry of the scene and the total number of cameras available. Note that it belongs in the ARC category of non-uniform plenoptic sampling outlined in Section 2.3.2.

We structure this chapter as follows: in Section 5.2 we determine the framework to evaluate different planar models of the scene surface. This involves deriving a metric for the distortion caused in the EPI domain when sampling and reconstructing the plenoptic function assuming a certain geometry and a finite number of cameras. Using this framework, we present in Section 5.3 an adaptive sampling algorithm for the plenoptic function of a scene with a smoothly varying surface. This algorithm is based on optimising the planar model of the surface using a binary-tree. In Sections 5.4 and 5.5 we validate and analyse the adaptive sampling algorithm using synthetic and real scenes, respectively. Finally, we end with a summary of the chapter in Section 5.6.

5.2 Evaluating the Model of the Scene Surface

To determine a model of the scene surface, we require a distortion metric for the reconstruction of the plenoptic function. This distortion metric must take into account that we are trying to reconstruct the plenoptic function using incorrect geometric information (i.e. the planar model of the scene) and that we also have a limited number of cameras available. If we assume the worse case scenario, we can treat these two elements as independent sources of error in plenoptic domain. Therefore, using this assumption, we formulate the distortion metric as the sum of the two errors, which we call the geometric error and the aliasing error. We define these errors as follows.

The geometric error is caused by approximating the scene surface with a sequence of slanted planes. It is the difference between the actual plenoptic function and that corresponding to a scene made of a sequence of slanted planes. For a smoothly varying scene surface, with no-occlusions, the geometric error can be defined for the i th plane

in the sequence as follows

$$\gamma_i = \underbrace{\int \int}_{t,v \in \mathcal{P}_i} |p(t,v) - \tilde{p}_i(t,v)|^2 dt dv, \quad (5.1)$$

where $\tilde{p}_i(t,v)$ is the plenoptic function corresponding to the i th plane in the sequence and \mathcal{P}_i is its support in the EPI domain.

The aliasing error is caused by undersampling the plenoptic function that corresponds to a slanted plane. We estimate this error using the worse case scenario - twice the energy outside the reconstruction filter [8]. Therefore, like the geometric error, it can be defined for the i th plane in the sequence as follows

$$\Lambda_i(N_i) = 2 \underbrace{\int \int}_{\omega_t, \omega_v \notin \mathcal{R}_{\psi,i}} \left| \tilde{P}_{S,i}(\omega_t, \omega_v) \right|^2 d\omega_t d\omega_v, \quad (5.2)$$

where $\tilde{P}_{S,i}(\omega_t, \omega_v)$ is the plenoptic spectrum for the i th plane, N_i is the number of cameras used to reconstruct the corresponding plenoptic function and $\mathcal{R}_{\psi,i}$ is the support of the reconstruction filter. The aliasing error is related to the number of cameras N_i through the support of the reconstruction filter. In the frequency domain, this relationship is

$$\mathcal{R}_{\psi,i} = \left\{ \omega_t, \omega_v : \omega_v \in \left[-\frac{\pi}{\Delta v}, \frac{\pi}{\Delta v} \right], \omega_t \in \left[\frac{\omega_v f}{z_{G,i}} - \frac{\pi}{\Delta t_i}, \frac{\omega_v f}{z_{G,i}} + \frac{\pi}{\Delta t_i} \right] \right\}, \quad (5.3)$$

where $z_{G,i}$ is the local version of z_G for the i th plane and

$$\Delta t_i = \frac{t_{\text{range}}}{N_i}. \quad (5.4)$$

The parameter t_{range} is the overall support of \mathcal{P}_i in t . For the i th slanted plane in the sequence it is

$$t_{\text{range}} = x_{2,i} - x_{1,i} + (z_{2,1} + z_{1,i}) \bar{v}_m. \quad (5.5)$$

Having defined (5.1) and (5.2), the expression for the distortion caused by sampling

and reconstructing the plenoptic function, using a planar model of the scene surface that comprises L_s planes, is

$$D_T = \sum_{i=1}^{L_s} \gamma_i + \Lambda_i(N_i). \quad (5.6)$$

At the moment, this distortion does not take into account sampling the plenoptic function with a limited number of cameras, N_c . We impose this constraint by restricting the set of N_i such that

$$\sum_{i=1}^{L_s} N_i = N_c. \quad (5.7)$$

A consequence of this constraint is that we are implicitly assuming each camera can only sample one plane. This assumption means that our calculation of the overall aliasing error of a scene will be quite conservative. However, we want to optimise both errors jointly to determine the best model of the scene and in turn the best camera placement. Therefore, the advantage of this assumption is that we can treat the distortion for each plane independently, hence simplifying the joint optimisation. Accordingly, by restricting (5.6) with (5.7), we formulate the distortion caused by sampling and reconstructing the plenoptic function using N_c cameras and a model of the scene comprising L_s planes.

Unfortunately, an exact knowledge of the actual plenoptic function is required to determine both of the errors in (5.6) precisely. In view of this, we now estimate each error assuming only knowledge of the scene geometry and the maximum frequency of the texture signal.

5.2.1 Estimating the Geometric Error

Starting with the geometric error, we rewrite $p(t, v)$ and $\tilde{p}_i(t, v)$ in terms of the actual curvilinear coordinate, $s(t, v)$, and the curvilinear coordinate for the i th plane, $\tilde{s}_i(t, v)$. Therefore, we have $p(t, v) = p(s(t, v))$ and $\tilde{p}_i(t, v) = p(\tilde{s}_i(t, v))$. We now assume that the curvilinear coordinate for the i th plane can be written as $\tilde{s}_i(t, v) = s(t, v) + e_i(t, v)$, where $e_i(t, v)$ is an error term. Then, using Taylor series expansion, we obtain the

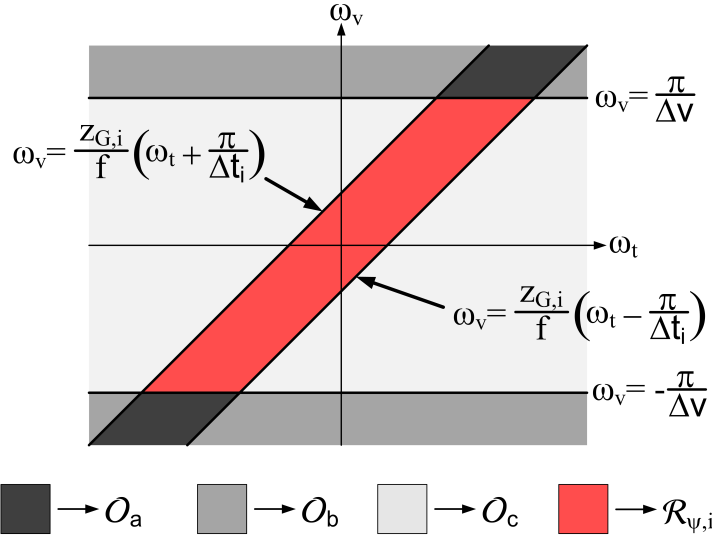


Figure 5.1: Diagram illustrating the regions \mathcal{O}_a , \mathcal{O}_b and \mathcal{O}_c used to approximate the energy of the plenoptic spectrum that resides outside the support of the reconstruction filter $\mathcal{R}_{\psi,i}$.

in which the ESD decays like $1/\omega^2$;

$$\mathcal{O}_b = \left\{ \omega_t, \omega_v : \omega_v \notin \left[-\frac{\pi}{\Delta v}, \frac{\pi}{\Delta v} \right], \omega_t \notin \left[\frac{\omega_v f}{z_{G,i}} - \frac{\pi}{\Delta t_i}, \frac{\omega_v f}{z_{G,i}} + \frac{\pi}{\Delta t_i} \right] \right\}, \quad (5.12)$$

in which the ESD decays like $1/\omega^4$; and

$$\mathcal{O}_c = \left\{ \omega_t, \omega_v : \omega_v \in \left[-\frac{\pi}{\Delta v}, \frac{\pi}{\Delta v} \right], \omega_t \notin \left[\frac{\omega_v f}{z_{G,i}} - \frac{\pi}{\Delta t_i}, \frac{\omega_v f}{z_{G,i}} + \frac{\pi}{\Delta t_i} \right] \right\}, \quad (5.13)$$

where the ESD decays like $1/\omega^2$. The approximate amount of energy within each region is determined as follows.

First, we note that, due to (5.4), all of these regions are dependent on the number of cameras N_i used to reconstruct the plenoptic function. However, the combined amount of energy within \mathcal{O}_a and \mathcal{O}_b is constant regardless of the number of cameras (i.e. it is a systematic error due to the finite pixel resolution of the cameras). In view of this, we estimate the combined energy by assuming the plenoptic function is sampled at the Nyquist density. This means that Δt_i in (5.11) and (5.12) is given by $\Delta t_i = 2\pi/A_i$, where A_i is the width of the essential bandwidth defined in (4.33). Based on this

assumption, we can safely model the decay of the ESD for \mathcal{O}_a and \mathcal{O}_b as $1/\omega_v^2$ and

$$\frac{1}{\omega_v^2 \left(\omega_t - \omega_v \frac{f}{z_{G,i}} \right)^2},$$

respectively. Using these models, the amount of energy within region \mathcal{O}_a is

$$\begin{aligned} E_a &= K_i \int_{\frac{\pi}{\Delta v}}^{\infty} \int_{\frac{\omega_v f}{z_{G,i}} - \frac{A_i}{2}}^{\frac{\omega_v f}{z_{G,i}} + \frac{A_i}{2}} \frac{1}{\omega_v^2} d\omega_t d\omega_v + K_i \int_{-\infty}^{-\frac{\pi}{\Delta v}} \int_{\frac{\omega_v f}{z_{G,i}} - \frac{A_i}{2}}^{\frac{\omega_v f}{z_{G,i}} + \frac{A_i}{2}} \frac{1}{\omega_v^2} d\omega_t d\omega_v \\ &= \frac{2A_i K_i \Delta v}{\pi}, \end{aligned} \quad (5.14)$$

and corresponding amount in \mathcal{O}_b is

$$\begin{aligned} E_b &= K_i \int_{\frac{\pi}{\Delta v}}^{\infty} \frac{1}{\omega_v^2} \left[\int_{\frac{\omega_v f}{z_{G,i}} - \frac{A_i}{2}}^{\infty} \frac{d\omega_t}{\left(\omega_t - \frac{\omega_v f}{z_{G,i}} \right)^2} + \int_{-\infty}^{\frac{\omega_v f}{z_{G,i}} - \frac{A_i}{2}} \frac{d\omega_t}{\left(\omega_t - \frac{\omega_v f}{z_{G,i}} \right)^2} \right] d\omega_v \\ &\quad + K_i \int_{-\infty}^{-\frac{\pi}{\Delta v}} \frac{1}{\omega_v^2} \left[\int_{\frac{\omega_v f}{z_{G,i}} + \frac{A_i}{2}}^{\infty} \frac{d\omega_t}{\left(\omega_t - \frac{\omega_v f}{z_{G,i}} \right)^2} + \int_{-\infty}^{\frac{\omega_v f}{z_{G,i}} - \frac{A_i}{2}} \frac{d\omega_t}{\left(\omega_t - \frac{\omega_v f}{z_{G,i}} \right)^2} \right] d\omega_v \\ &= \frac{8K_i \Delta v}{A_i \pi}. \end{aligned} \quad (5.15)$$

The parameter K_i in both (5.14) and (5.15) is a scaling factor and will be quantified once we have the whole approximation.

In contrast to the combined energy in \mathcal{O}_a and \mathcal{O}_b , the energy in region \mathcal{O}_c depends on the number of cameras and must be finite for all values of N_i . As a result, we model the decay of the ESD within this region using the following Lorentzian function

$$\frac{\Gamma_i}{\left(\omega_t - \omega_v \frac{f}{z_{G,i}} \right)^2 + \Gamma_i^2},$$

where Γ_i controls the width of the function. We choose this function because it decays like $1/\omega^2$ for large ω but has a finite maximum value of Γ_i . The amount of energy

within \mathcal{O}_c , therefore, is

$$\begin{aligned}
E_c &= K_i \int_{-\frac{\pi}{\Delta v}}^{\frac{\pi}{\Delta v}} \int_{\frac{\omega_v f}{z_{G,i}} + \frac{\pi}{\Delta t_i}}^{\infty} \frac{\Gamma_i}{\left(\omega_t - \omega_v \frac{f}{z_{G,i}}\right)^2 + \Gamma_i^2} d\omega_t d\omega_v \\
&\quad + K_i \int_{-\frac{\pi}{\Delta v}}^{\frac{\pi}{\Delta v}} \int_{-\infty}^{\frac{\omega_v f}{z_{G,i}} - \frac{\pi}{\Delta t_i}} \frac{\Gamma_i}{\left(\omega_t - \omega_v \frac{f}{z_{G,i}}\right)^2 + \Gamma_i^2} d\omega_t d\omega_v \\
&= \frac{4K_i\pi}{\Delta v} \tan^{-1} \left(\frac{t_{\text{range}}\Gamma_i}{\pi N_i} \right), \tag{5.16}
\end{aligned}$$

where the last step follows from solving the integration and using (5.4). In conclusion, the total approximate energy within the three regions is

$$E_a + E_b + E_c = K_i \left[\frac{2A_i\Delta v}{\pi} + \frac{8\Delta v}{A_i\pi} + \frac{4\pi}{\Delta v} \tan^{-1} \left(\frac{t_{\text{range}}\Gamma_i}{\pi N_i} \right) \right]. \tag{5.17}$$

Having determined the energy in (5.17), we now determine the scale factor K_i and width factor Γ_i . Starting with K_i , at critical sampling, we assume 90% of the signal's energy resides within the reconstruction filter, hence

$$\begin{aligned}
0.9 E_{T,i} &= K_i \int_{-\frac{\pi}{\Delta v}}^{\frac{\pi}{\Delta v}} \int_{\frac{\omega_v f}{z_{G,i}} - \frac{A_i}{2}}^{\frac{\omega_v f}{z_{G,i}} + \frac{A_i}{2}} \frac{\Gamma_i}{\left(\omega_t - \omega_v \frac{f}{z_{G,i}}\right)^2 + \Gamma_i^2} d\omega_t d\omega_v \\
&= \frac{4K_i\pi}{\Delta v} \tan^{-1} \left(\frac{A_i}{2\Gamma_i} \right), \tag{5.18}
\end{aligned}$$

where $E_{T,i}$ is the total energy for plenoptic function corresponding to the i th plane in the sequence. Directly following from (5.18), 10% of the energy must be outside the filter at critical sampling. Therefore, using (5.17), we obtain a second expression

$$\begin{aligned}
0.1 E_{T,i} &= E_a + E_b + E_c \\
&= \frac{2A_i K_i \Delta v}{\pi} + \frac{8K_i \Delta v}{A_i \pi} + \frac{4K_i \pi}{\Delta v} \tan^{-1} \left(\frac{t_{\text{range}}\Gamma_i}{\pi N_i} \right) \\
&= \frac{2A_i K_i \Delta v}{\pi} + \frac{8K_i \Delta v}{A_i \pi} + \frac{4K_i \pi}{\Delta v} \tan^{-1} \left(\frac{2\Gamma_i}{A_i} \right), \tag{5.19}
\end{aligned}$$

where the last step follows from $t_{\text{range}}/N_i = \Delta t_i = 2\pi/A_i$ at critical sampling. Now as

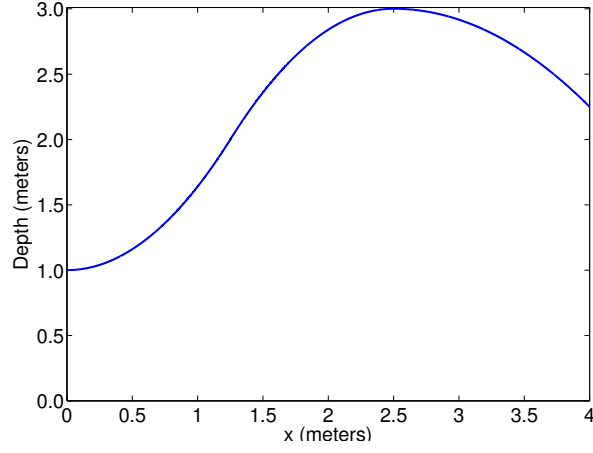


Figure 5.2: Diagram showing a piecewise quadratic scene, which consists of three quadratic pieces.

both A_i and Γ_i are greater than zero, the following is true

$$\tan^{-1}\left(\frac{2\Gamma_i}{A_i}\right) = \frac{\pi}{2} - \tan^{-1}\left(\frac{A_i}{2\Gamma_i}\right). \quad (5.20)$$

Thus, by applying the above expression to (5.19) and substituting (5.18), we obtain

$$K_i = \frac{E_{T,i} A_i \Delta v \pi}{2A_i^2 \Delta v^2 + 8\Delta v^2 + 2\pi^3 A_i}. \quad (5.21)$$

Using this definition of K_i , the width factor Γ_i is given by

$$\Gamma_i = \frac{A_i}{2} \cot\left(\frac{0.9(A_i^2 \Delta v^2 + 4\Delta v^2 + \pi^3 A_i)}{2\pi A_i}\right). \quad (5.22)$$

Note that with this definition of K_i and Γ_i the approximate energy outside the reconstruction filter, defined in (5.17), tends to $E_{T,i}$ as N_i tends to zero.

Finally, assuming the worst case, our estimate of the aliasing error is

$$\tilde{\Lambda}_i(N_i) = \frac{4A_i K_i \Delta v}{\pi} + \frac{16K_i \Delta v}{A_i \pi} + \frac{8K_i \pi}{\Delta v} \tan^{-1}\left(\frac{t_{\text{range}} \Gamma_i}{\pi N_i}\right), \quad (5.23)$$

where K_i is defined in (5.21) and Γ_i is (5.22).

5.2.3 Behaviour of the Distortion Metric

Using (5.10) and (5.23), the estimate of the distortion caused by sampling and reconstructing the plenoptic function, using N_c cameras and a L_s plane model of the scene surface, is

$$\tilde{D}_T = \sum_{i=1}^{L_s} \tilde{\gamma}_i + \tilde{\Lambda}_i(N_i), \quad (5.24)$$

where

$$\sum_{i=1}^{L_s} N_i = N_c.$$

Qualitatively, the expression in (5.24) behaves as follows; the distortion caused by the geometric error decreases as the planar model of the surface becomes more accurate. In contrast, the aliasing element of the distortion decreases as more cameras are allocated to each plane. However, as the sequence of slanted planes is unlikely to be uniform, an equal distribution of the cameras across the sequence is inefficient. An efficient approach would be to optimally allocate the cameras by minimising (5.24). Therefore, assuming N_c cameras and L_s slanted planes, this minimisation is defined as follows

$$\min \left\{ \sum_{i=1}^{L_s} \tilde{\gamma}_i + \tilde{\Lambda}_i(N_i) \right\} \text{ s.t. } N_c = \sum_{i=1}^{L_s} N_i. \quad (5.25)$$

We solve this minimisation using a Lagrangian multiplier λ . For the i th plane in the sequence the expression for this multiplier is

$$\lambda = -\frac{d}{dN_i} \left\{ \tilde{\Lambda}_i(N_i) \right\} = \left(\frac{8\pi K_i}{\Delta v} \right) \left(\frac{\pi \Gamma_i t_{\text{range},i}}{\pi^2 N_i^2 + \Gamma_i^2 t_{\text{range},i}^2} \right), \quad (5.26)$$

which can be rearranged to give

$$N_i = \sqrt{\frac{8K_i \Gamma_i t_{\text{range},i}}{\Delta v \lambda} - \left(\frac{\Gamma_i t_{\text{range},i}}{\pi} \right)^2}. \quad (5.27)$$

The above expression determines the camera allocation per plane but requires knowledge of λ . Using the finite camera constraint defined in (5.7), we obtain the following

expression in which λ is the only unknown

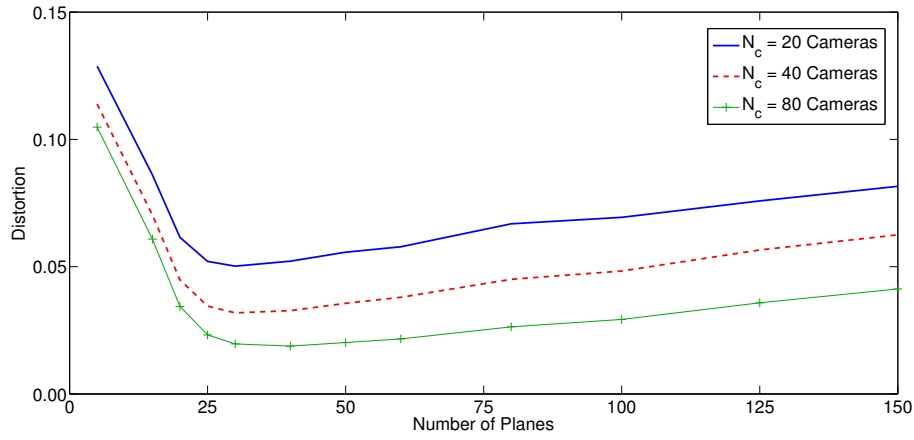
$$N_c = \sum_{i=1}^{L_s} \sqrt{\frac{8K_i\Gamma_i t_{\text{range},i}}{\Delta v \lambda} - \left(\frac{\Gamma_i t_{\text{range},i}}{\pi}\right)^2}. \quad (5.28)$$

Therefore, by determining λ from (5.28), we can optimally allocate the N_c cameras across L_s slanted planes using (5.27). Note that, as we uniformly sample each plane, this allocation results in a piecewise constant sampling of the scene.

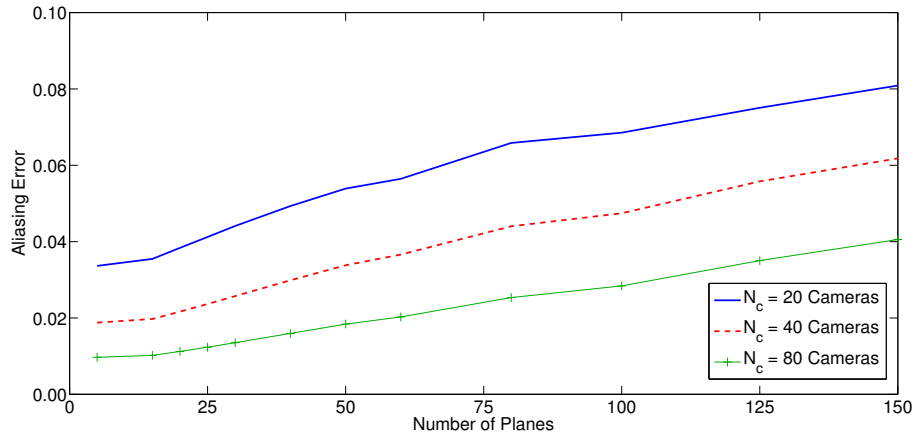
To analyse the minimum distortion caused by this camera allocation and to examine the effect of L_s , we calculate the distortion for a synthetic piecewise quadratic surface, shown in Figure 5.2, as the planar model of the surface becomes more accurate. The amount of distortion as a function of the number of planes in the model is shown in Figure 5.3(a). The figure plots the distortion for 20 cameras, 40 cameras and 80 cameras. The geometric and aliasing elements of each distortion curve are shown in Figure 5.3(b) and 5.3(c), respectively. Note that we set the cameras to have $\text{FoV} = 20^\circ$, $f = 32 \text{ mm}$ and $\omega_s = 125 \text{ rads/m}$.

First, the graphs confirm that increasing the number of cameras available decreases the overall distortion. Next, the graphs show that, for all three values of N_c , the distortion initially decreases to a minimum around 28 planes but it then rises as L_s increases beyond 30 planes. Figure 5.3(b) shows that this increase is due to the aliasing element of the distortion. In contrast, the geometric error shown in Figure 5.3(c) decreases for all values of L_s as the model becomes more accurate.

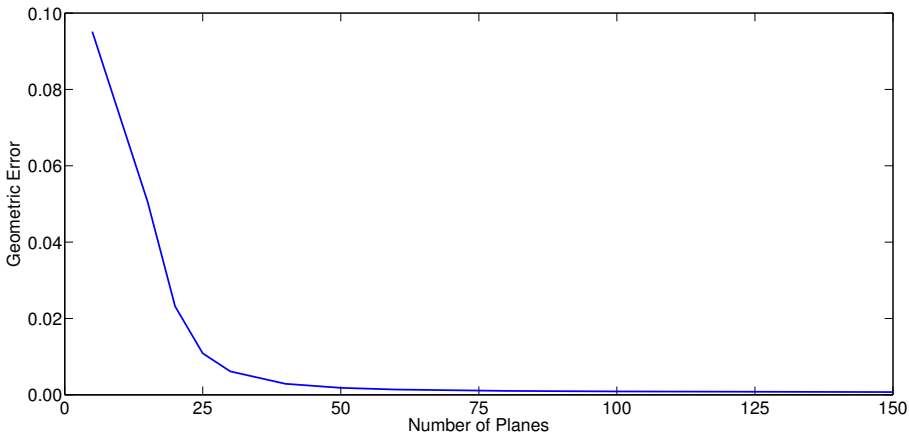
To understand why the aliasing error increases as the scene model becomes more accurate, first remember that the analysis in Chapters 3 and 4 predicted higher camera densities when the slant of a plane is large and that the gradient of a plane is equal to $|\tan(\phi)|$. Now, consider the gradient of the planar model in relation to the gradient of the scene's surface. As the model comprises a sequence of slanted planes and each plane within the model has a constant gradient, the model has a piecewise constant gradient that changes depending on L_s . In particular, as L_s increases, the gradient of the model changes to match the gradient of the scene's surface more accurately (i.e. the model changes to match the surface). This change results in the gradient of some



(a)



(b)



(c)

Figure 5.3: Graph (a) plots the estimated distortion for the plenoptic function as the number of planes used to model the surface increases from 5 to 150. The plenoptic function relates to the surface shown in Figure 5.2. The graph compares the distortion for 20, 40 and 80 cameras. Graphs (b) and (c) show respectively the amount of aliasing and geometric error that comprises the distortion in (a). Note that the cameras have a $\text{FoV} = 20^\circ$, in which $f = 32 \text{ mm}$, and $\omega_s = 125 \text{ rads/m}$.

planes increasing, which means an increase in their angle of slant. From the analysis in Chapter 3 and 4, we now require more cameras to avoid aliasing in these planes. Therefore, if we have a fixed number of cameras, increasing L_s will increase the aliasing error.

Using the scene in Figure 5.2 as an example, if $L_s = 1$ then the model of this scene will comprise a single slanted plane. This plane joins each end of the scene's surface together and has a slant of approximately 17° . If we now increase L_s to 2, thus split the first plane in half, we obtain a new model comprising two planes with approximate slants of 43° and -16° , respectively. From the analysis in Chapter 3 and 4, we observe that a higher camera density is required for a plane with a slant of 43° than a plane with a slant of 17° . Therefore, even though the scene model is more accurate, we will require more cameras to avoid aliasing.

5.3 Adaptively Sampling a Smoothly Varying Scene

To sample the plenoptic function of a smoothly varying surface with bandlimited texture, we want to position N_c cameras optimally based on the local geometry of the scene surface. In Section 5.2.3, we positioned these cameras non-uniformly, in a piecewise uniform pattern, by solving (5.25) for a given planar model of the surface. Therefore, given a planar model of the surface, we determined the optimum camera placement by minimising the aliasing error. This camera placement will change, however, depending on how we model the surface. For instance, if the model adapts to capture the local geometry of the scene, the camera positions will adapt as well. In view of this, we extend the minimisation from the previous section to allow the planar model of the scene surface to vary. We minimise the distortion caused in the plenoptic function by determining the optimal planar model of the surface and the optimum allocation of the cameras. Therefore, we are jointly minimising the geometric and aliasing error. In the following we outline the optimisation of the planar model of the surface and present the adaptive sampling algorithm based on this optimisation. Finally we examine the reconstruction of the plenoptic function from adaptive sample positions.

5.3.1 The Surface Optimisation

Taking inspiration from rate-distortion theory, we determine the optimum planar model of the scene using a binary tree framework. In this framework each leaf in the tree represents a plane in the model. The act of pruning leaves within the tree corresponds to merging planes in the model of the surface. Therefore, starting from an initial, fine-grain, model of the surface, we repeatedly prune the leaves of the tree until the overall distortion \tilde{D}_T is minimised. This pruning of the tree and its effects on the planar model of the surface are illustrated in Figure 5.4. Note that this framework can be scaled to a quad-tree allowing us to model 2D surfaces using a set of planar facets. Therefore our adaptive sampling algorithm can be scaled to sample 4D light fields, in which a quad-tree framework is more efficient.

The benefit of this framework is that the minimisation defined in (5.25) is reduced to a binary search in λ . In other words we search for the Lagrangian multiplier λ that minimises the overall distortion for the optimum planar model of the surface whilst obeying the constraint of N_c cameras. Therefore the process of pruning the binary tree is repeated many times, each time with a different value of λ , as we search for this optimum solution. Note that λ is progressively determined using bisection.

5.3.2 The Adaptive Sampling Algorithm

Drawing all the above together, the complete algorithm for adaptively sampling the plenoptic function of a smoothly varying surface is summarised as follows:

- Step 1.** Initially model the smoothly varying surface, $z(x)$, with a sequence of L_s slanted planes. The slanted planes are generated by joining $L_s + 1$ equidistant points $(x_i, z(x_i))$ where $i = 1, \dots, L_s + 1$. The points are equidistant in the x -axis. The number of planes, L_s , is equal to the largest power of two such that $x_2 - x_1 > z_{\min} v_m / f$.
- Step 2.** Given the initial planar model of the surface, minimise the distortion defined in (5.25). In doing so determine the starting value of the Lagrangian multiplier, λ , by solving (5.28).

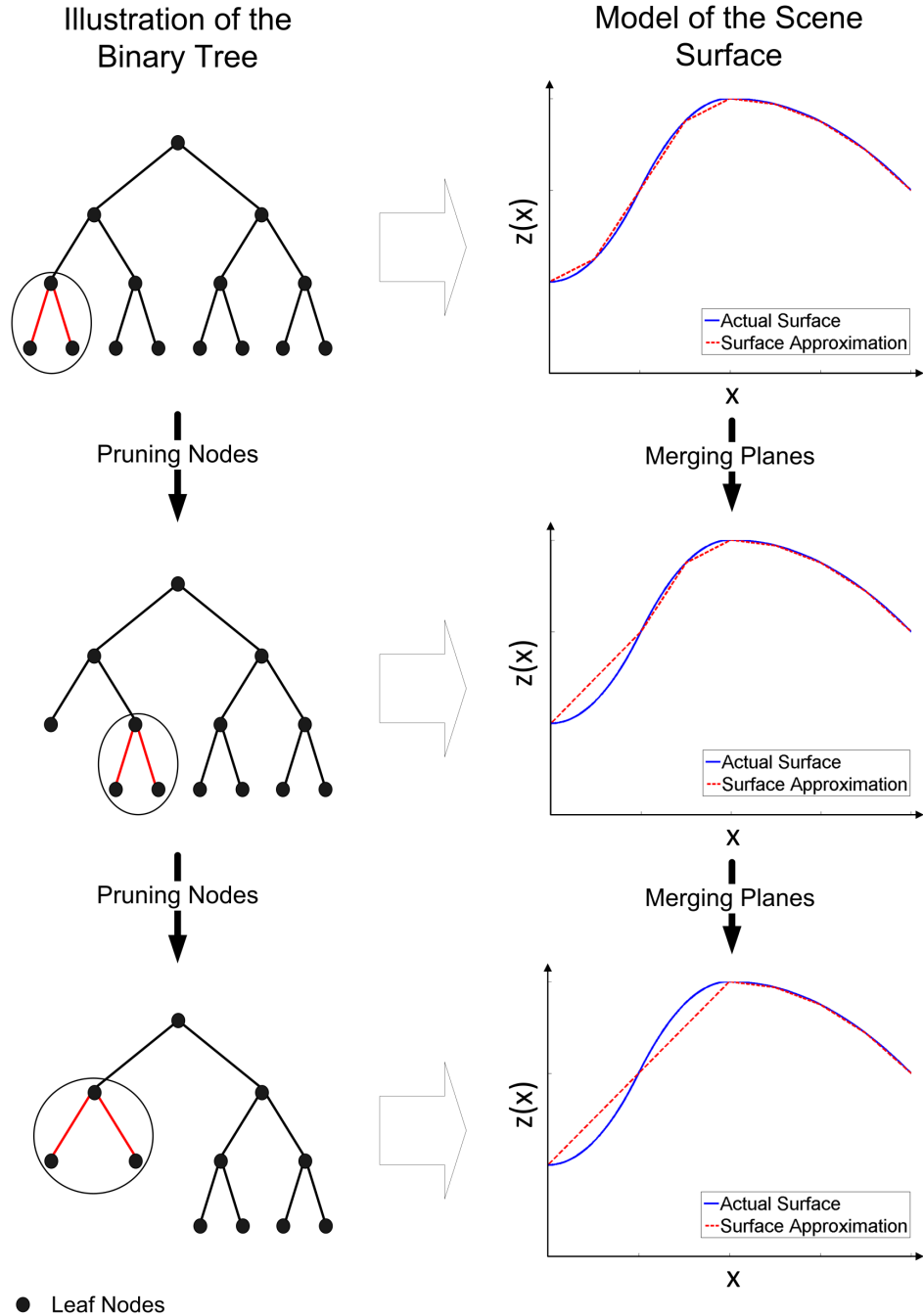


Figure 5.4: Diagram showing how the model of the scene surface, using a sequence of slanted planes, is optimised within a binary tree framework. Note that each leaf node in the tree represent a plane in the sequence, hence pruning the tree means merging two planes into one.

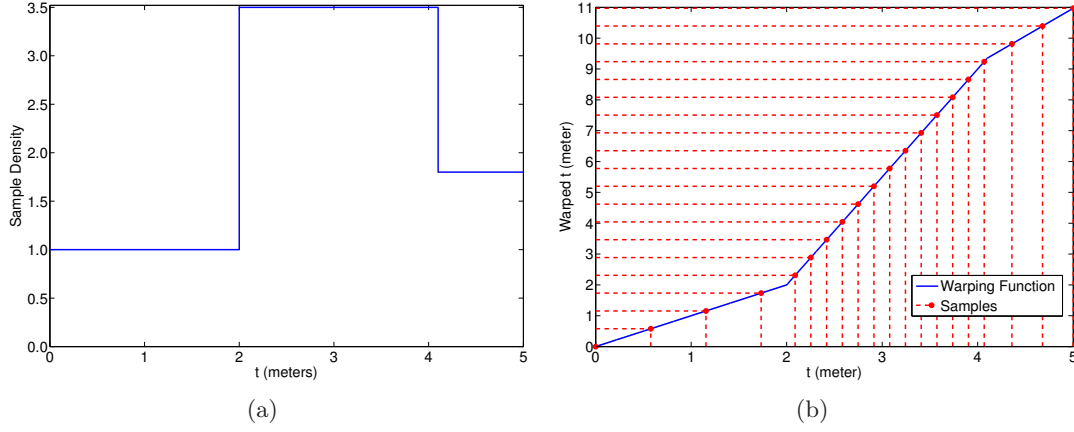


Figure 5.5: Diagrams illustrating the non-uniform placement of 20 cameras using a piecewise constant sample density in t . Graph (a) shows the piecewise constant sample density for the scene as a function of t . Graph (b) shows the warping function generated from the sample density. The function is constructed by integrating the sampling density, hence it is piecewise linear. The graph also shows how uniform sampling in the warped domain is converted to non-uniform sampling in t .

Step 3. Generate the binary-tree with L_s initial leaf nodes, one for each plane in the model. Using λ , determine the sample allocation, from (5.27), and local distortion for each node in the tree. Attempt to prune each pair of leaf nodes. Pruning is performed if the local distortion of the parent node is less than the joint local distortions of the leaves. This process is repeated until no possible merges occur.

Step 4. Calculate the total number of allocated cameras for the refined model of the surface, N_a , and compare to the constraint of N_c cameras. If the constraint is satisfied: $N_a = N_c$, stop. If the constraint is not satisfied, adjust λ and repeat Step 3. λ is determined using bisection. The direction of the search is as follows:

- If $N_a < N_c$, reduce λ .
- If $N_a > N_c$, increase λ .

Step 5. Determine the placement of the cameras to sample the scene. The process for determining these positions is illustrated in Figure 5.5. First, calculate a piecewise constant sample density for the scene using the optimal model from Step 4. An example density is shown in Figure 5.5(a). Once this density is obtained, construct a warping function by integrating the sampling density. Figure 5.5(b)

shows the corresponding warping function for the density in Figure 5.5(a). Finally, position the cameras by converting a set of uniform positions in the warped domain to positions in t , as shown in Figure 5.5(b) using 20 cameras. Note that, as the warping function is piecewise linear, the positions in t will be piecewise uniform in nature.

Therefore, the output of this algorithm is a set of camera positions distributed in a piecewise uniform pattern. These camera positions adapt depending on the total number of cameras available and the local geometric complexity of the scene. Note that, if the number of uniform pieces is greater than one, the global distribution of the cameras is non-uniform.

5.3.3 Reconstruction from Adaptive Sample Locations

In order to reconstruct the plenoptic function from these camera locations, we require adaptive filters. We, therefore, interpolate the plenoptic function using a warped version of the 3^{rd} order I-MOMS presented in Section 4.4. This interpolation has to be performed in the EPI domain as the warping is camera specific, hence we have a set of warped reconstruction filters. In view of this, we incorporate the depth information obtained from modelling the scene surface using a sequence of planes. The resulting set of reconstruction filters are defined as follows:

$$\phi_{i,k}(t, v) = \eta_I^{\{3\}} \left(\frac{\pi}{\Delta t} (\mathcal{W}(t) - \hat{t}_i) \right) \eta_I^{\{3\}} \left(\frac{\pi}{\Delta v} \left(v - v_k + \frac{f}{z(t, v)} (t - t_i) \right) \right), \quad (5.29)$$

where $v_k = k\Delta v \forall k = 1, \dots, N_p$ are the set of pixel locations, $t_i \forall i = 1, \dots, N_c$ are the set of camera positions, $\mathcal{W}(t)$ is the warped camera line generated using the warping function \mathcal{W} , and \hat{t}_i are the warped camera positions. Notice that the actual camera positions and the warped positions are related as follows $\hat{t}_i = \mathcal{W}(t_i)$. The varying depth information is introduced using the 2D variable $z(t, v)$, which is the depth of scene corresponding to the light ray (t, v) . This 2D variable, $z(t, v)$, is obtained using the planar model of the surface generated by our adaptive sampling algorithm. In terms of the discussion on depth information in Section (2.4), we are operating in the second

group as we have access to per pixel depth information. However, we use the depth information to reconstruct the whole plenoptic function rather than rendering single views.

To illustrate how the filter defined in (5.29) differs from a uniform filter, we examine the reconstruction of a sampled synthetic EPI. The synthetic EPI corresponds to the piecewise quadratic scene shown in Figure 5.2. This EPI is sampled uniformly and adaptively using our algorithm. Note that 30 cameras are used in both cases and each camera has $\text{FoV} = 20^\circ$ and $f = 70$ mm. Using this set up, we plot four different reconstruction filters in Figure 5.6. Each plot in the figure is constructed as follows. First, to indicate what the filter is trying to reconstruct, we have a low contrast version of the complete EPI. Then, on top of this low contrast EPI, we have the samples obtained from the cameras. Note that we focus on a small piece of the EPI, hence only samples from 5 cameras are visible. The position of the cameras, and the associated samples, are indicated in each plot. Finally, superimposed as the top most layer, we have the EPI structure of the filter.

The difference between the four plots in Figure 5.6 are as follows. In Figure 5.6(a), the filter has a constant skew of f/z_G and the cameras are uniformly placed to sample the EPI. Therefore, the structure of the filter is constant regardless of the camera or pixel position. In Figure 5.6(b), the camera placement is still uniform but the filter has a varying skew of $f/z(t, v)$. Accordingly, the structure of the filter is no longer constant as its skew varies depending on the camera and pixel position. Consequently, the filter structure is deformed by the varying depth information. In Figure 5.6(c), we introduce adaptive camera placement but fix the skew of the filter at f/z_G . The figure, therefore, illustrates the warping of the filter in order to reconstruct from adaptive samples. Finally, in Figure 5.6(d), the EPI is adaptively sampled but the filter has a varying skew of $f/z(t, v)$. Consequently, the structure of the filter is warped due to the sampling and deformed due to the depth information. Note that Figure 5.6(d) illustrates the one of the set of filters defined by (5.29).

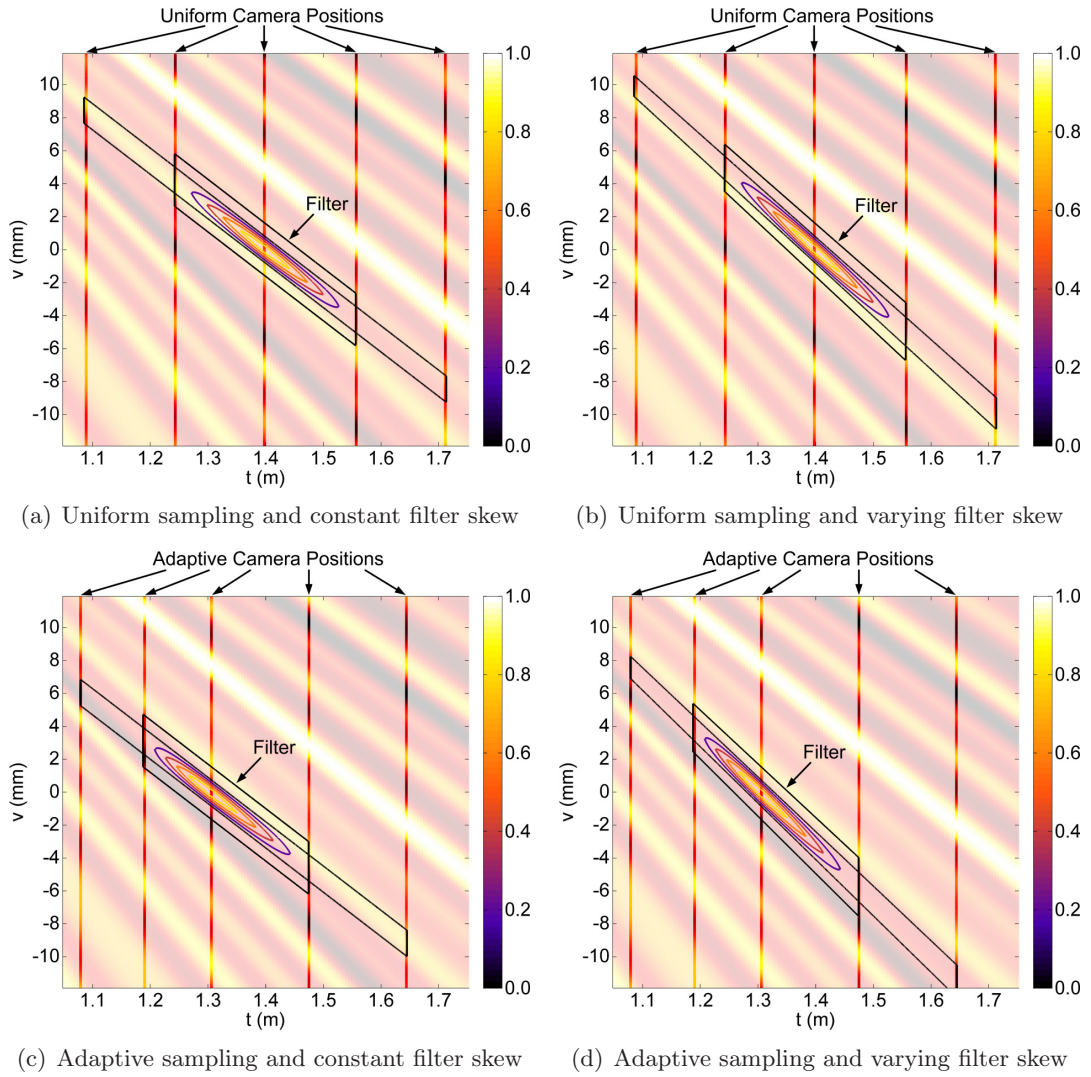


Figure 5.6: Graphs illustrating four different reconstruction filters for a sampled synthetic EPI. The EPI corresponds to the quadratic scene in Figure 5.2. It is sampled uniformly in (a) and (b), and adaptively in (c) and (d) using the algorithm in Section 5.3.2. Accordingly, in (a) the filter is uniform with a constant filter skew of f/z_G . In (b) the filter corresponds to uniform sampling but it has a varying filter skew of $f/z(t, v)$. In (c) the filter is adaptive but with a fixed filter skew of f/z_G . Finally, in (d) the filter is adaptive with a varying skew of $f/z(t, v)$. Note that in all of the graphs the filter is superimposed on the synthetic EPI and the actual sample are highlighted.

5.4 Results and Simulations for Synthetic Scenes

In this section we analyse and validate the proposed adaptive sampling algorithm using a synthetic scene. The scene comprises 5 quadratic pieces with bandlimited texture pasted to the surface. It is illustrated in Figures 5.7, 5.8 and 5.9. To sample this scene we use cameras with a focal length of 70 mm and a FoV $\approx 20^\circ$. For the following analysis, we initialise the binary tree with 16 planes.

5.4.1 Analysis of Camera Locations

Using this scene, we analyse the placement of the cameras by our algorithm as the maximum frequency of the texture signal varies. The camera placement is illustrated in Figures 5.7, 5.8 and 5.9 for $\omega_s = 60$ rads/m, 120 rads/m and 300 rads/m, respectively. Each figure shows the camera locations, along the x -axis, and the planar model of the surface when there are 20, 50 and 90 cameras in total. We also plot, just below each scene, the relative sampling density as a function of x that corresponds to the camera placement. The sampling density is relative to the density achieved using uniform sampling. In other words a value greater than 1 indicates oversampling relative to uniform sampling whereas below 1 indicates undersampling.

The figures show that the maximum frequency of the texture affects the number of planes used to model the scene surface. For a low frequency texture, as in Figure 5.7, the number of planes in the model increases from 5, to 8 and then 10 for the respective 20, 50 and 90 cameras available. However, when $\omega_s = 120$ rads/m, Figure 5.8 shows that the number of planes decreases from 8 to 6 as the number of cameras increased from 20 to 50. The model then remain constant as the number of cameras increased from 50 to 90. Finally, Figure 5.9 shows that for $\omega_s = 300$ rads/m the number of planes in the model is fixed at 10 regardless of the number of cameras.

The reason for this behaviour is that both the geometric error, defined in (5.10), and the aliasing error, defined in (5.23), depend on ω_s . The geometric error is proportional to ω_s^2 hence, for a large ω_s , the error is reduced by making the model more accurate. For the aliasing error, a large value of ω_s leads to more cameras being required to

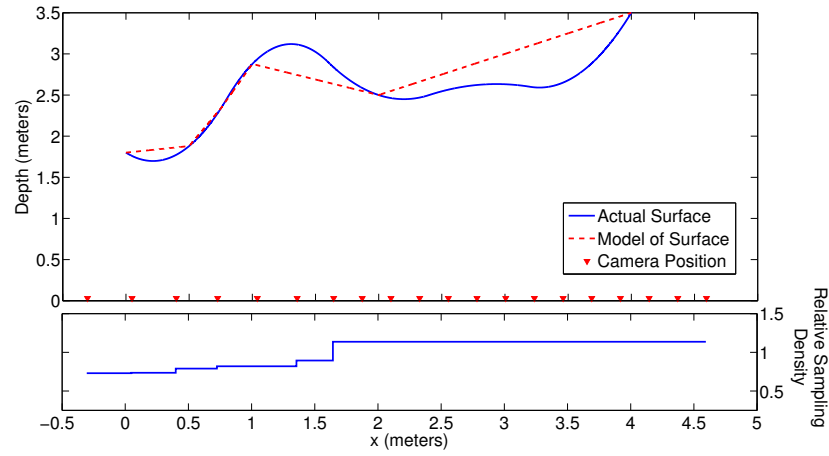
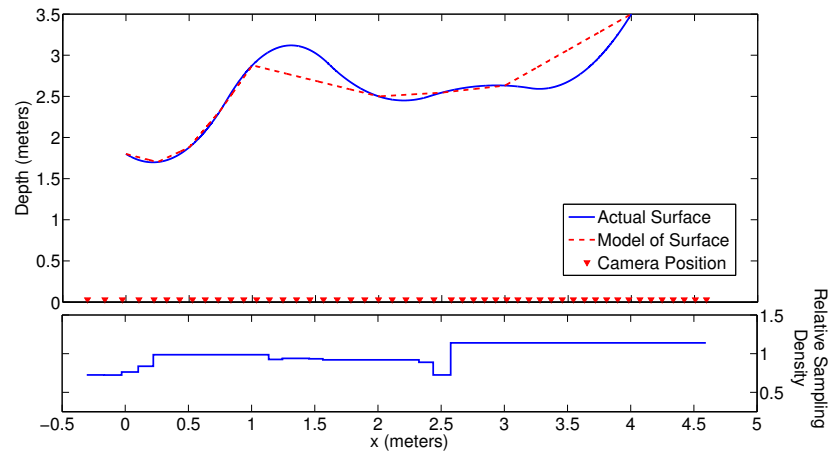
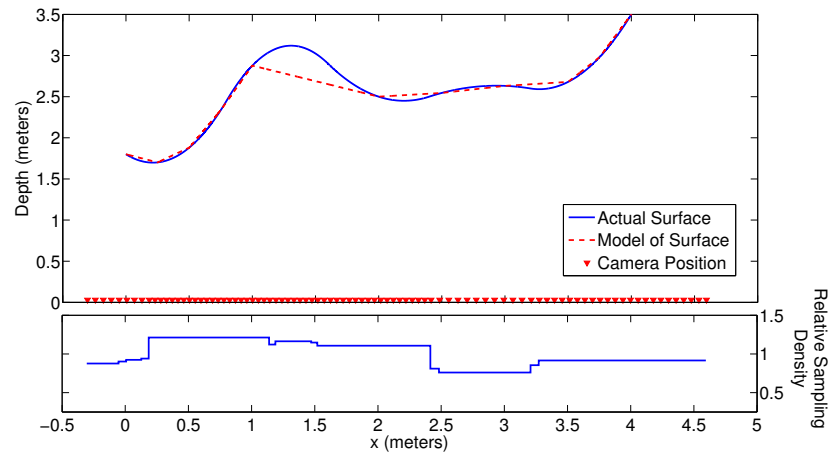
(a) $N_c = 20$ Cameras and $L_s = 5$ Planes(b) $N_c = 50$ Cameras and $L_s = 8$ Planes(c) $N_c = 90$ Cameras and $L_s = 10$ Planes

Figure 5.7: Diagrams showing the position of N_c cameras generated by the adaptive sampling algorithm for a piecewise quadratic surface, comprising 5 pieces, with texture bandlimited to 60rad/m. The positions of the cameras are indicated along the x -axis and, below each surface, is the corresponding sampling density as a function of x . The sampling density is relative to uniform sampling. Each graph also shows the planar model of the surface generated by the algorithm.

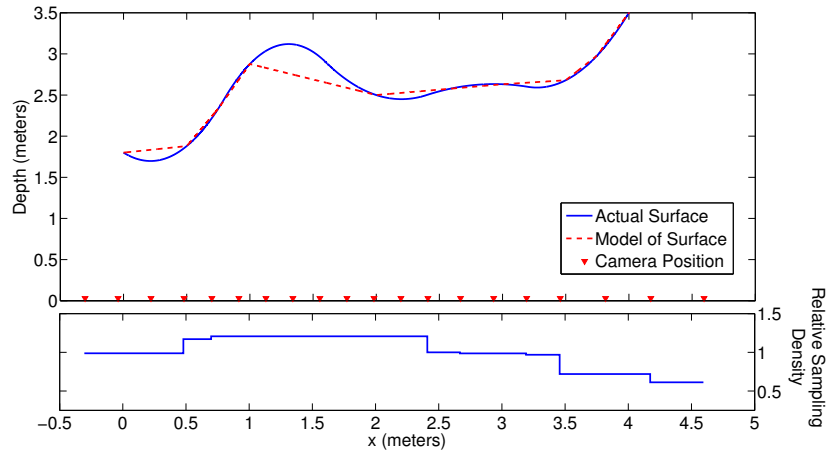
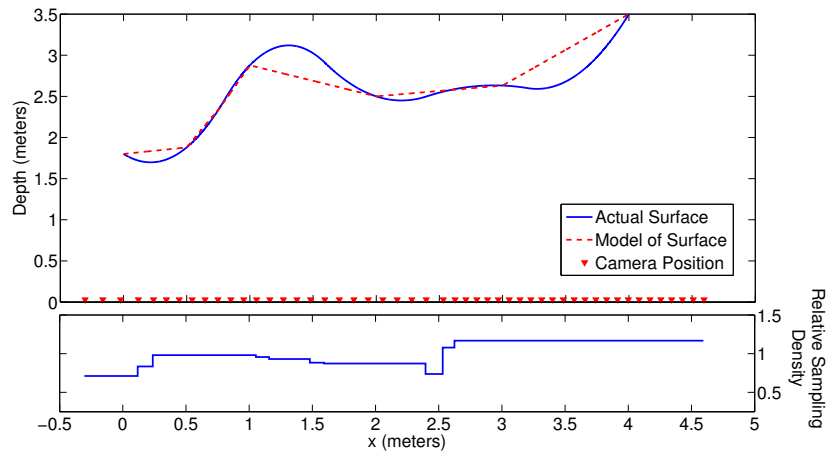
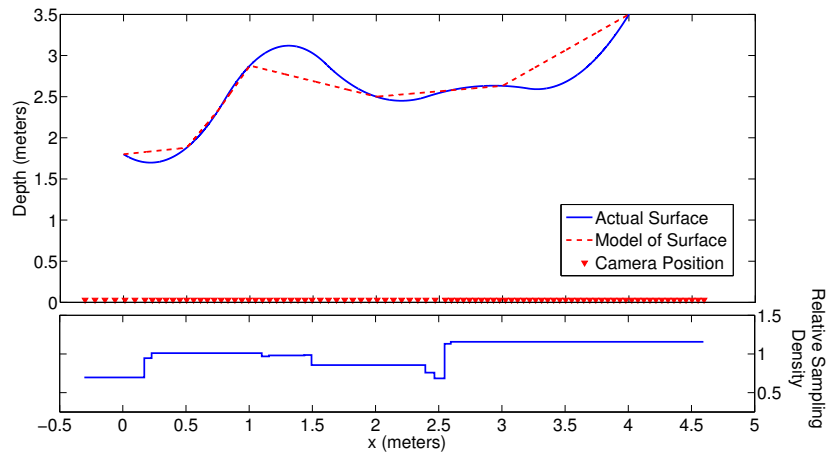
(a) $N_c = 20$ Cameras and $L_s = 8$ Planes(b) $N_c = 50$ Cameras and $L_s = 6$ Planes(c) $N_c = 90$ Cameras and $L_s = 6$ Planes

Figure 5.8: Diagrams showing the position of N_c cameras generated by the adaptive sampling algorithm for a piecewise quadratic surface, comprising 5 pieces, with texture bandlimited to 120 rads/m. The positions of the cameras are indicated along the x -axis and, below each surface, is the corresponding sampling density as a function of x . The sampling density is relative to uniform sampling. Each graph also shows the planar model of the surface generated by the algorithm.

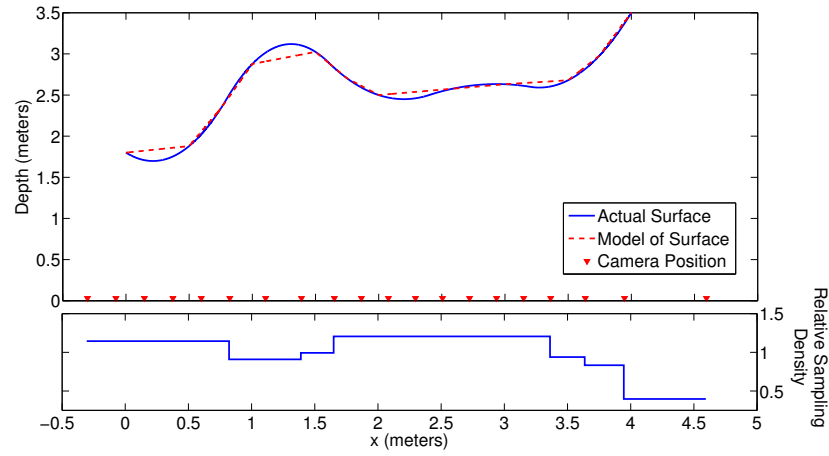
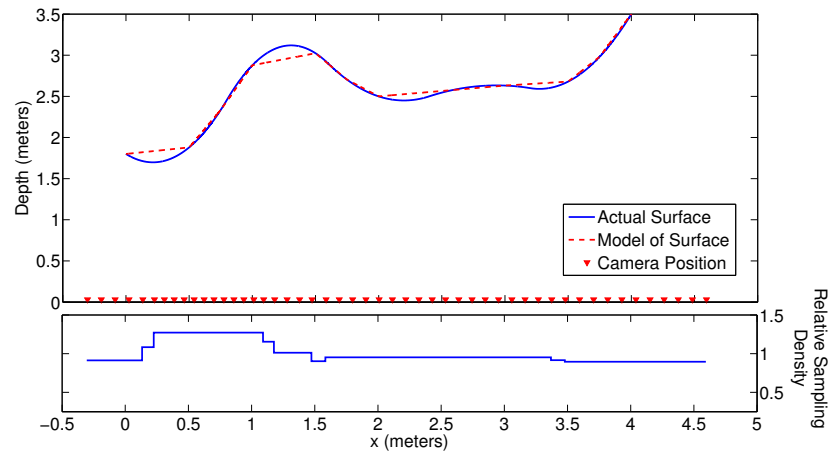
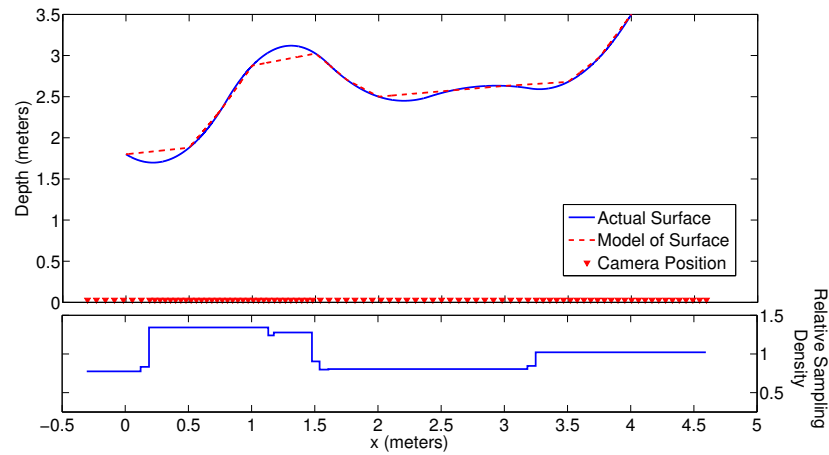
(a) $N_c = 20$ Cameras and $L_s = 10$ Planes(b) $N_c = 50$ Cameras and $L_s = 10$ Planes(c) $N_c = 90$ Cameras and $L_s = 10$ Planes

Figure 5.9: Diagrams showing the position of N_c cameras generated by the adaptive sampling algorithm for a piecewise quadratic surface, comprising 5 pieces, with texture bandlimited to 300 rads/m. The positions of the cameras are indicated along the x -axis and, below each surface, is the corresponding sampling density as a function of x . The sampling density is relative to uniform sampling. Each graph also shows the planar model of the surface generated by the algorithm.

sample the scene. In Figure 5.7 the value of ω_s is small, hence the algorithm focuses on the aliasing error. Accordingly, the planar model of surface gradually becomes more accurate as the number of cameras increases. However, in Figure 5.9, the value of ω_s is very large and too few cameras are available. Therefore the algorithm generates a more accurate model of the scene to reduce the geometric error regardless of the aliasing error. The point at which the algorithm switches from one to the other depends on the value of ω_s and it is this point that is captured in Figure 5.8.

If we now examine the algorithm's camera placement, the figures highlight three factors that contribute to high camera densities. The first factor is the gradient of the scene surface; the greater the gradient the higher the camera density, as shown in Figures 5.9(b) and 5.9(c). The reason for this is that the frequency content of the texture signal increases as the gradient of the surface increases, hence more cameras are required (especially if ω_s is high). The second factor is the distance between the scene and camera line; the closer the scene the higher the camera density due to the FFoV. This effect explains the differing camera densities for similar surface gradients in Figure 5.9(c). The third factor is the planar model of the surface itself. The cameras are allocated across the model using the minimisation in (5.25). Therefore, in situations where coarse models of the surface are generated, for example low frequency texture, the cameras will be allocated depending onto the size of the planes. This factor is seen at work in Figures 5.7(a) and 5.7(b), where the highest camera densities coincide with a large plane, which is not the steepest plane in the surface model.

5.4.2 Validating the Adaptive Sampling Algorithm

Using this synthetic scene, we validate the algorithm by comparing the reconstruction achieved using its non-uniform samples and the filters defined in (5.29) to uniform reconstruction. The PSNR curves for the reconstruction of the synthetic EPIs are shown in Figure 5.10. The figure compares the reconstruction using the two sampling schemes as the maximum frequency of the texture signal varies from $\omega_s = 60$ rads/m to $\omega_s = 120$ rads/m and then $\omega_s = 300$ rads/m.

The figures show that adaptive sampling using our algorithm results in a higher

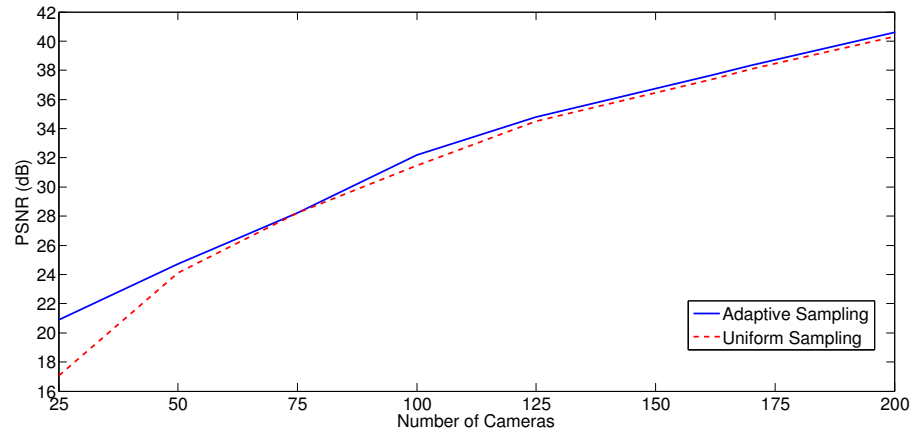
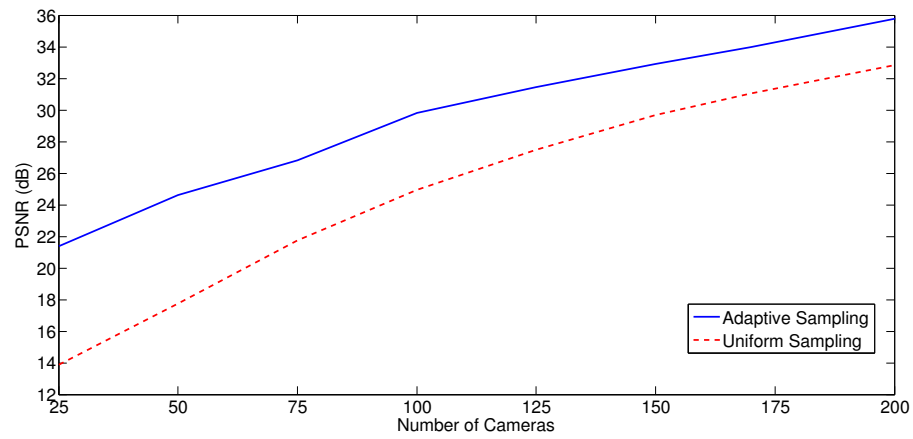
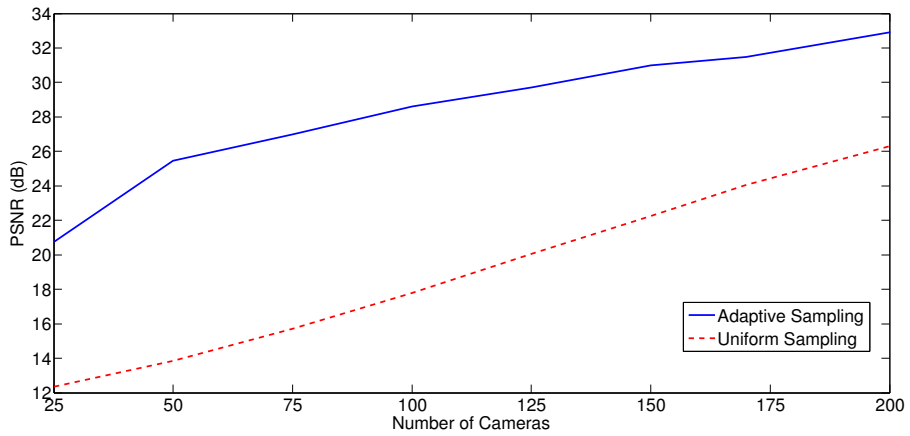
(a) $\omega_s = 60$ rads/m(b) $\omega_s = 120$ rads/m(c) $\omega_s = 300$ rads/m

Figure 5.10: Graphs comparing the PSNR curves for the reconstruction of three synthetic EPIs when sampled uniformly and adaptively. The adaptive sample positions are determined using our sampling algorithm. The EPIs correspond to a piecewise quadratic surface, comprising 5 pieces, with texture bandlimited to (a) 60 rads/m, (b) 120 rads/m and (c) 300 rads/m. Note that the cameras have a focal length equal to 70 mm and a FoV $\approx 20^\circ$.

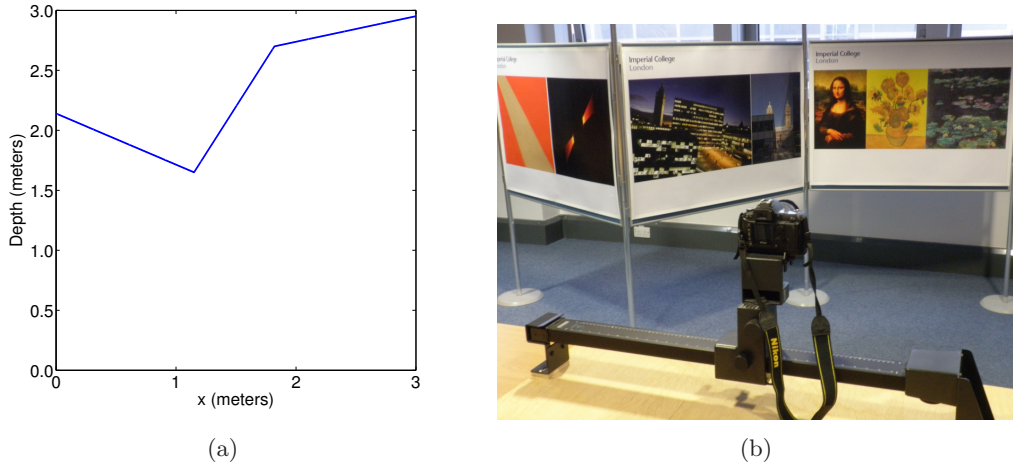


Figure 5.11: The set-up and data acquisition for the EPI-volume corresponding to the planar scene. Diagram (a) illustrates the bird’s eye view of the scene geometry and (b) shows the data acquisition. The resulting EPI-volume consists of 253 images spaced 1 cm apart. Each image is 3008 by 1888 pixels.

PSNR than uniform sampling for the same number of cameras. This gain in performance is most noticeable when the scene’s texture has a high frequency content (i.e. ω_s is large). For example, in Figure 5.10(c) the minimum gain in PSNR using our algorithm is 7 dB. In contrast, when $\omega_s = 60$ rads/m as shown in Figure 5.10(a), our algorithm only just outperforms uniform sampling. The reason for this behaviour is that more cameras are required for adequate uniform sampling as the maximum frequency of the texture signal increases. Therefore the scene is undersampled in Figures 5.10(b) and 5.10(c) when using a uniform sampling scheme. Our algorithm, however, positions the cameras in order to reduce undersampling, hence the reconstruction is improved. Note that this also occurs in Figure 5.10(a) but only when there are 25 cameras.

5.5 Results and Simulations for Real Scenes

Having verified the algorithm using synthetic data, we examine the sampling and reconstruction of two EPI-volumes generated from real scenes. The first scene comprises three planar facets with real texture pasted to the surface. The second scene comprises a smoothly varying surface with similar real texture pasted to the surface. Both scenes are constrained such that the depth only varies with x . To reconstruct the EPI-

volume we use the 3rd order I-MOM interpolator presented in Section 4.4. Also, similar to Section 4.4, we reconstruct the uniformly sampled EPI-volumes using the baseline algorithm to give perspective to our results.

5.5.1 Analysis of the Piecewise Planar Scene

A bird's eye view of the planar scene surface is illustrated in Figure 5.11(a). From this scene, we generated an EPI-volume comprising 253 images, each 3008 by 1888 pixels in size. The images are captured at 1 cm intervals along the camera line with a camera that has a focal length of 70 mm and a FoV $\approx 20^\circ$. The data acquisition set-up used to obtain the images is shown in Figure 5.11(b).

The PSNR curves for the reconstruction of the entire EPI-volume are shown in Figure 5.12. It compares the reconstruction achieved using adaptive sampling to that achieved using two types of uniform sampling and reconstruction; interpolation with I-MOMS and reconstruction with the baseline algorithm. The figure shows that adaptive sampling of the EPI-volume results in a higher PSNR than both uniform sampling methods for all camera densities. To highlight this gain in reconstruction, Figure 5.13 presents images rendered using each method. The original is shown in Figure 5.13(a), its rendering using uniform sampling in Figure 5.13(b), its rendering using the baseline algorithm in Figure 5.13(c), and its rendering using adaptive sampling in Figure 5.13(d). In all cases the rendered images are generated using 37 original images. The corresponding PSNR values for the rendered images are (b) 19.0 dB, (c) 20.6 dB and (d) 23.5 dB. Therefore our algorithm results in a 3.5 dB increase in rendering quality relative to uniform sampling and a 2.9 dB relative to the Baseline algorithm. More examples of rendered images are shown in Appendix F.

5.5.2 Analysis of the Smoothly Varying Scene

In this new scenario, we capture both colour images and depth images of the smoothly varying surface. The depth images are obtained using Microsoft's Xbox Kinect. From this depth data, we determine the bird's eye view of the surface shown in Figure 5.14(a). The EPI-volume generated from this scene comprises 73 colour images, each 3008 by

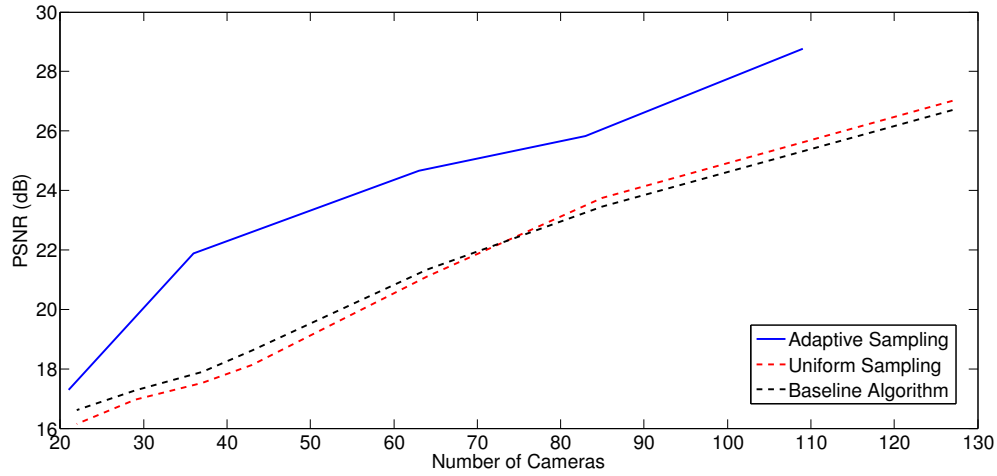


Figure 5.12: Graph comparing the PSNR curves for the reconstruction of an EPI-volume. The curves correspond to adaptive sampling, uniform sampling and uniform sampling with the baseline algorithm. The EPI-volume relates to the three plane scene shown in Figure 5.11. The adaptive samples are determined using our sampling algorithm. The cameras have a focal length of 70 mm, which corresponds to a FoV $\approx 20^\circ$.

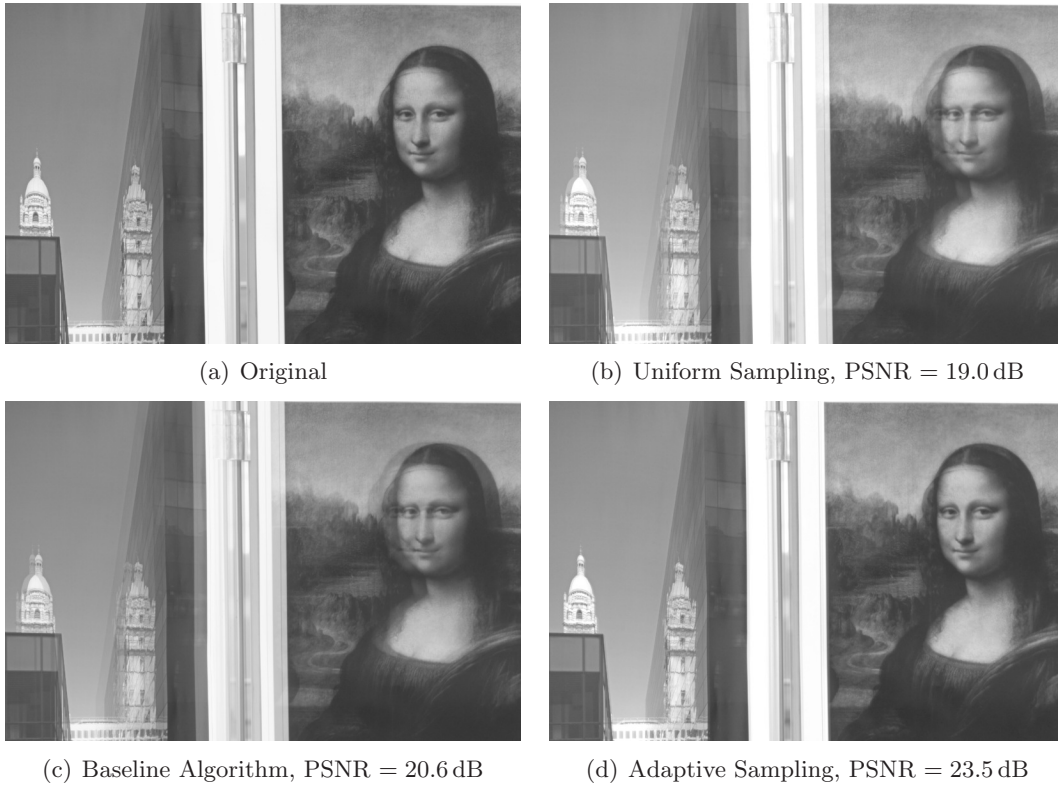


Figure 5.13: An example of a rendered image from an EPI-volume using three different sampling and reconstruction methods. The EPI-volume relates to the three plane scene shown in Figure 5.11. The original image is shown in (a), its rendering using uniform sampling in (b), its rendering using the baseline algorithm uniform in (c), and its rendering using adaptive sampling in (d). Each image is rendered using 37 cameras.

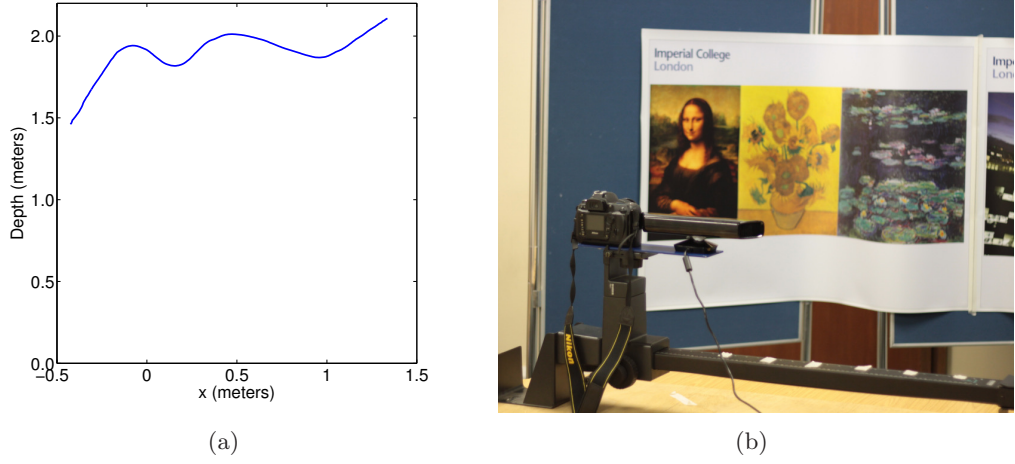


Figure 5.14: The set-up and data acquisition for the EPI-volume corresponding to the smoothly varying scene. Diagram (a) illustrates the bird’s eye view of the scene geometry. Diagram (b) shows the twin data acquisition of both colour and depth images using a standard camera and Microsoft’s Xbox Kinect. The resulting EPI-volume consists of 73 images spaced 1 cm apart. Each image is 3008 by 2000 pixels. We also obtain 73 depth images spaced 1 cm apart, but at a resolution of 374 by 248 pixels.

2000 pixels in size. Again, the images are captured at 1 cm intervals along the camera line but with a camera that has a focal length 34 mm and a $\text{FoV} \approx 39^\circ$. At the same time as the colour images, we also capture 73 depth images, each 374 by 248 pixels. The twin camera rig used for data acquisition is shown in Figure 5.14(b).

Figure 5.15 compares the reconstruction of the whole EPI-volume using adaptive sampling to that achieved using uniform. Again, we compare two types of uniform reconstruction, the first with I-MOMS and the second using the baseline algorithm. Similar to the piecewise planar scene, the adaptive sampling of the EPI-volume results in a higher PSNR than both uniform sampling methods for all camera densities. A comparison of rendered images from the EPI-volume is shown in Figure 5.16. The original is shown in Figure 5.16(a), its rendering using uniform sampling in Figure 5.16(b), its rendering using the baseline algorithm in Figure 5.16(c), and its rendering using adaptive sampling in Figure 5.16(d). The images are rendered using 10 original images and the corresponding PSNR values are (b) 21.0 dB, (c) 21.1dB and (d) 22.0 dB. Similar to Section 5.5.2, more examples of rendered images are shown in Appendix G.

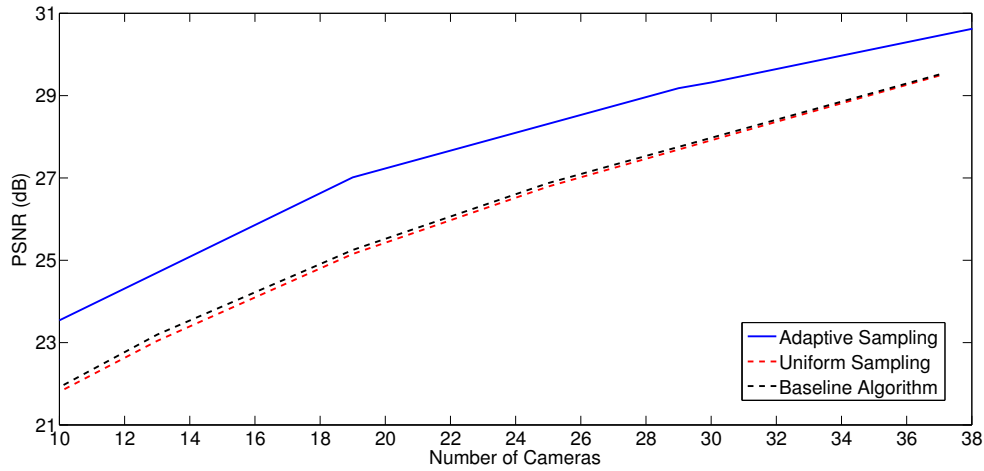


Figure 5.15: Graph comparing the PSNR curves for the reconstruction of a EPI-volume. The curves correspond to adaptive sampling, uniform sampling and uniform sampling with the baseline algorithm. The EPI-volume relates to the smoothly varying scene shown in Figure 5.14. The adaptive samples are determined using our sampling algorithm. The cameras have a focal length of 34 mm, which corresponds to a FoV $\approx 39^\circ$.

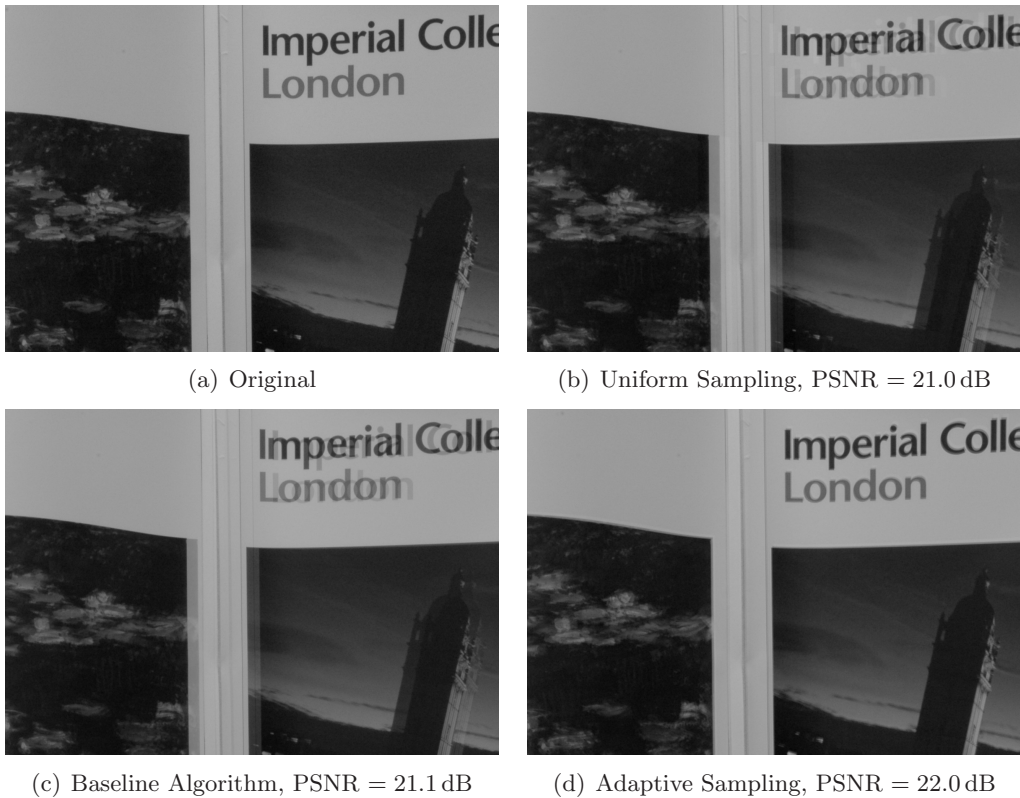


Figure 5.16: An example of a rendered image from a EPI-volume using three different sampling and reconstruction methods. The EPI-volume relates to the smoothly varying scene shown in Figure 5.14. The original image is shown in (a), it's rendering using uniform sampling in (b), it's rendering using the baseline algorithm in (c), and it's rendering using adaptive sampling in (d). Each image is rendered using 10 cameras.

5.6 Summary

In this chapter we examined the non-uniform sampling of the plenoptic function. We approached this problem by proposing a novel framework that combines uniform plenoptic sampling with adaptive camera placement. We proposed modelling the geometric complexity of the scene using a sequence of slanted planes. Once this model is acquired, the cameras are then positioned using our uniform sampling result for a slanted plane. Therefore the placement of the cameras adapts depending on the planar model of the surface.

Based on this framework, we presented an adaptive plenoptic sampling algorithm for a scene with a smoothly varying surface and bandlimited texture. The algorithm operates by determining the best planar model of the scene and positions the cameras accordingly. The best planar model is determined by minimising the distortion in the plenoptic function. This distortion comprises the geometric error, in the plenoptic function, caused by modelling the scene incorrectly and the aliasing error caused by undersampling the function. The minimisation was performed in a binary tree framework. To reconstruct the non-uniformly sampled plenoptic function, we presented an adaptive filter in the EPI-domain using 3rd order I-MOMS. Using this filter, we demonstrated that the adaptive sampling algorithm outperforms uniform sampling for both synthetic and real scenes.

The key points to take from this chapter are as follows:

- We present a framework that combines the theoretical results from uniform plenoptic sampling with adaptive camera placement.
- Within the framework, the scene geometry is modelled with a sequence of slanted planes and the cameras positioned based on our uniform plenoptic sampling of a slanted plane.
- Using this framework, we present an adaptive sampling algorithm that determines camera positions based on the scene geometry and the number of cameras available.

- The algorithm operates by determining a planar model of the surface that minimises the distortion in the plenoptic function. The distortion comprises the geometric and aliasing error caused when sampling and reconstructing the plenoptic function.
- We propose an adaptive filter in order to reconstruct the plenoptic function from non-uniform camera positions.
- Our algorithm results in improved reconstruction and rendering for both synthetic EPIs and real EPI-volumes when compared to uniform sampling.

Chapter 6

Conclusions

6.1 Summary of Thesis Achievements

The concept of the plenoptic function provides a natural framework for examining multi-view image processing algorithms. In particular, it allows image-based rendering to be posed in terms of sampling and reconstruction. The multi-view image set, in this scenario, represents the samples of the plenoptic function and the rendering of a new view its reconstruction. The minimum number of images required in IBR, and their optimum placement, can therefore be determined through sampling analysis of the plenoptic function. In this thesis we have explored plenoptic sampling in both a uniform and non-uniform framework with the aim of determining the optimum camera placement.

The core achievements of our thesis are summarised as follows. We derived theoretical results for uniform plenoptic sampling that improved the reconstruction of the plenoptic function. We designed a state-of-the-art adaptive plenoptic sampling algorithm that outperformed uniform plenoptic sampling. We proposed a new reconstruction filter for uniform plenoptic sampling and derived an adaptive version to handle non-uniform camera placement. In more detail, we expand upon each of these achievements in the following discussion.

In Chapter 3, we derived an exact closed-form expression for the plenoptic spectrum of a Lambertian slanted plane with complex exponential texture. This expression was

generalised to a slanted plane with bandlimited texture in Section 4.2.1. The novelty of our spectral analysis is that we incorporated two realistic conditions: finite scene width and cameras with a finite field of view. The imposition of these conditions leads to a band-unlimited plenoptic spectrum. Accordingly, in Section 3.4, we defined a structural model to characterise the plenoptic spectrum. The model comprised two quadrilateral regions bounded by six lines. Using this model, we studied the behaviour of the plenoptic spectrum when scene and camera parameters vary.

Assuming uniform camera placement, in Chapter 4, we derived a new expression for the maximum spacing between adjacent cameras for a slanted plane with bandlimited texture. This expression is derived assuming the plenoptic function for a slanted plane is adequately sampled using its essential bandwidth (a region in frequency containing at least 90% of the signal's energy). We defined this essential bandwidth using a 2D parametric model. The parameters of the model are determined from the plenoptic spectral analysis in Section 4.2.1 and the structural model defined in Section 3.4. The spatial Nyquist camera density is determined using this parametric model and leads to our expression for the maximum camera spacing. Using synthetic and real scenes, we showed that our expression results in a more accurate Nyquist camera density than the current state-of-the-art for a slanted plane.

In Chapter 5, we designed a novel algorithm to adaptively sample the plenoptic function for a smoothly varying surface with bandlimited texture. The algorithm adaptively positions cameras by using a slanted plane as an elementary element to construct more complicated scenes. Therefore the scene surface is modelled using a sequence of slanted planes and the cameras positioned using our spectral analysis of a slanted plane from Section 4.2.1. The algorithm functions by determining the best sequence of planes given the local geometric complexity of the scene and the number of cameras available. Therefore it determines the model of the surface that minimises the distortion in the plenoptic function. This distortion is caused by sampling and reconstructing the plenoptic function with a limited number of cameras and assuming a certain geometry. We defined this distortion as the combination of the geometric error - caused by modelling the scene incorrectly - and the aliasing error - caused by

undersampling the plenoptic function. The minimisation is performed in a binary tree framework. Using synthetic and real scenes, we showed that the non-uniform camera positioning generated by our algorithm outperforms uniform camera positioning for all camera densities.

Finally, we presented a new filter that improved the reconstruction of real and synthetic EPIs when compared to the current state-of-the-art. The parametrisation of the filter is based on the model of the essential bandwidth defined in Chapter 4. We showed that this change in parametrisation alone leads to a slight improvement in reconstruction when compared to the state-of-the-art. However, we further improved the filter by incorporating 3rd order I-MOMS in Section 4.4. Using an EPI-volume corresponding to a real scene, we showed that inclusion of these interpolators, as opposed to a sinc function, suppresses ringing artefacts when rendering new views. For the non-uniform plenoptic sampling in Chapter 5, we derived an adaptive version of these filters that also incorporated varying depth information. The depth information is determined from the planar model generated by our adaptive sampling algorithm.

6.2 Future Research

To conclude, we discuss some open questions and possible directions for future research. We split this discussion into two parts: direct extensions to the research presented in this thesis and research leading beyond this thesis.

6.2.1 Extensions to Presented Research

The possible extensions to the research presented in this thesis are as follows:

- **Extension to 4D Light Fields:** Throughout this thesis we restricted our analysis to a 2D light field representation. As a result we analysed the 2D plenoptic spectrum for a 1D slanted plane and used this analysis to adaptively sample plenoptic functions for 1D surfaces. In the instances when we examined real scenes, the depth of the scene had only one degree of freedom in the x -axis. An extension to this research, therefore, is to consider the full 4D light field obtained

from a 2D scene surface. For example one could analyse the plenoptic spectrum for a 2D planar facet and use the results to develop a 2D extension to our adaptive sampling algorithm. Therefore the aim of the algorithm would be to model the 2D scene surface with a set of planar facets and non-uniformly position cameras on the 2D camera plane to reconstruct the light field.

- **2D Camera Paths:** In the 2D light field representation, the spatial positions of the cameras are restricted to a 1D camera path, parallel to the scene, and their orientations are fixed such that they are perpendicular to this camera path. We derived in Section 4.3 the maximum camera spacing along this line for a slanted plane and showed it to be inversely proportional to the depth variation in the scene. In Section 3.5, however, we pointed out how rotation of the camera line could be used to reduce the depth variation in the scene. Future research, therefore, could be to consider 2D camera paths with the aim of reducing the number of cameras required to sample the plenoptic function. A possible solution to this problem would be a camera path that minimised the depth variation in the scene. However the problem becomes more difficult as occlusions would need to be taken into account. Aside from this consideration, the scenario also poses the following open questions: How should we reconstruct the plenoptic function in this situation? How should we assess the reconstruction? For instance, should we compare the reconstruction of the plenoptic function to that obtained from a 1D parallel camera path?

6.2.2 Beyond Our Research - Multi-View Depth Cameras

In Section 2.4 we highlighted the pivotal role of depth information in IBR and how such information is normally generated - off-line using multi-view stereo vision algorithms [42, 48]. Recent advances in sensing technologies, however, make large-scale deployment of 3D cameras using active depth sensing a distinct possibility. For example Microsoft's Xbox Kinect and, more recently, the compressive depth acquisition camera (CODAC) framework presented in [30, 31] - which promises low cost, high reso-

lution, depth acquisition using compressive sampling theory. This proliferation of depth cameras, therefore, leads to multi-view systems that contain both types of camera.

Such dual camera systems are starting to emerge in IBR and can be split into three categories: single depth and single colour camera systems [16, 46], single depth and multiple colour camera systems [36, 72], and multiple depth and multiple colour camera systems [29, 39, 64]. For example, in Section 5.5, we used a single depth and single colour camera system to sample a smoothly varying scene. The emergence of these systems, therefore, offers the prospect of analysing multi-view depth images as well as multi-view colour images.

For example we proposed in [26] analysing multi-view depth images using a framework akin to the plenoptic function, which we term the *pantelic function*¹. Within this framework, the multi-view depth images represent samples of the pantelic function. Therefore, using this pantelic function, we presented an initial spectral analysis on multi-view depth images. However the following open questions remain: How many depth cameras are necessary to infer the scene geometry? How do the number of depth cameras relate to the number of colour cameras? Can an excess of one be used to compensate a reduction of the other? How does the current low resolution of depth cameras affect this scenario?

¹The word pantelic is derived from a slight abuse of the Greek $\pi\alpha\nu$ meaning all, and $\tau\eta\lambda\epsilon$ meaning distance.

Appendix A

The Plenoptic Spectrum for a Slanted Plane

In this appendix we first derive the plenoptic spectrum for a slanted plane with complex exponential texture and then extend this to the more general case when the texture is bandlimited. The initial phase of both derivation is as follows: starting from the general equation in (2.12), we apply our finite constraints, FSW and FFoV, to the scene and assume the surface is Lambertian. Therefore $x \in [x_1, x_2]$, $\bar{v} \in [-\bar{v}_m, \bar{v}_m]$, where $\bar{v}_m = v_m/f$, and $l_x(x, \bar{v}) = l_x(x)$, which results in

$$P(\omega_t, \omega_v) = \int_{x=x_1}^{x=x_2} l_x(x) e^{-j\omega_t x} \underbrace{\int_{\bar{v}=-\bar{v}_m}^{\bar{v}=\bar{v}_m} (1 - z'(x)\bar{v}) f e^{-j(\omega_v f - z(x)\omega_t)\bar{v}} d\bar{v}}_{I_2} dx. \quad (\text{A.1})$$

In order to solve I_2 , the integral in \bar{v} , we define the FFoV constraint as a rectangular windowing function, as follows

$$\text{rect}\left(\frac{f}{2v_m}\bar{v}\right) = \begin{cases} 1 & , \text{ if } |\bar{v}| \leq \frac{v_m}{f} \\ 0 & , \text{ else} \end{cases} \quad (\text{A.2})$$

whose corresponding Fourier transform is

$$\mathcal{F}_{\bar{v}} \left\{ \text{rect} \left(\frac{f}{2v_m} \bar{v} \right) \right\} = 2 \frac{v_m}{f} \text{sinc} \left(\frac{v_m}{f} \omega_v \right).$$

By defining this window function, we evaluate I_2 as follows

$$\begin{aligned} I_2 &= f \int_{-\infty}^{\infty} (1 - z'(x)\bar{v}) \text{rect} \left(\frac{f}{2v_m} \bar{v} \right) e^{-j(\omega_v f - z(x)\omega_t)\bar{v}} d\bar{v} \\ &\stackrel{(i)}{=} f \mathcal{F}_{\bar{v}} \left\{ \text{rect} \left(\frac{f}{2v_m} \bar{v} \right) \right\} - j f z'(x) \frac{\partial}{\partial \omega_v} \left\{ \mathcal{F}_{\bar{v}} \left\{ \text{rect} \left(\frac{f}{2v_m} \bar{v} \right) \right\} \right\} \\ &= 2v_m \text{sinc}(\omega_I) - j 2 \frac{z'(x)v_m^2}{f} \text{sinc}'(\omega_I), \end{aligned} \quad (\text{A.3})$$

where step (i) follows from properties of the Fourier transform, $\text{sinc}'(\omega_I)$ is the first derivative of the sinc function with respect to ω_I , and

$$\omega_I = \omega_v v_m - z(x) \frac{\omega_t v_m}{f}.$$

Substituting the above expression into (A.1), we obtain

$$P(\omega_t, \omega_v) = 2v_m \int_{x=x_1}^{x=x_2} l_x(x) \left(\text{sinc}(\omega_I) - j \frac{z'(x)v_m}{f} \text{sinc}'(\omega_I) \right) e^{-j\omega_t x} dx, \quad (\text{A.4})$$

noting that ω_I is dependent upon x . At this point we use the scene geometry equations, (3.1), to change the variable of integration from x to the curvilinear coordinate s , and the Lambertian assumption to switch to the texture signal, i.e. $l(s) = g(s)$. Thus we obtain

$$P_S(\omega_t, \omega_v) = M_1 \int_0^T g(s) \left[\text{sinc}(\hat{\omega}_I) - j \frac{v_m \tan(\phi)}{f} \text{sinc}'(\hat{\omega}_I) \right] e^{-j\omega_t \cos(\phi)s} \cos(\phi) ds$$

where

$$\hat{\omega}_I = \omega_v v_m - (s \sin(\phi) + z_1) \frac{v_m}{f} \omega_t, \quad z'(x) = \tan(\phi), \quad \text{and} \quad M_1 = 2v_m e^{-j\omega_t x_1}.$$

At this point we need to define the texture signal, therefore the two derivations diverge.

A.1 Complex Exponential Texture

Assuming the texture signal is a complex exponential, $g(s) = e^{j\omega_s s}$, then the plenoptic spectrum becomes

$$P_S(\omega_t, \omega_v) = M_1 \int_0^T \left[\text{sinc}(\hat{\omega}_I) - j \frac{v_m \tan(\phi)}{f} \text{sinc}'(\hat{\omega}_I) \right] e^{-js(\omega_t \cos(\phi) - \omega_s)} \cos(\phi) ds. \quad (\text{A.5})$$

From here we change the variable of integration from s to $\hat{\omega}_I$, thus

$$s = \frac{\omega_v f - z_1 \omega_t}{\sin(\phi) \omega_t} - \hat{\omega}_I \frac{f}{\sin(\phi) \omega_t v_m}, \quad ds = d\hat{\omega}_I \frac{-f}{\sin(\phi) \omega_t v_m},$$

and the limits of integration become

$$\begin{aligned} s = 0 &\leftrightarrow \hat{\omega}_I = \omega_v v_m - \omega_t \frac{z_1 v_m}{f} = b, \\ s = T &\leftrightarrow \hat{\omega}_I = \omega_v v_m - \omega_t \frac{z_2 v_m}{f} = a. \end{aligned}$$

Consequently, (A.5) becomes

$$\begin{aligned} P_S(\omega_t, \omega_v) = M_1 &\left(\frac{-f e^{-j(\omega_t \cos(\phi) - \omega_s) \left(\frac{\omega_v f - z_1 \omega_t}{\sin(\phi) \omega_t} \right)}}{\tan(\phi) \omega_t v_m} \right) \int_b^a \text{sinc}(\hat{\omega}_I) e^{-j\hat{\omega}_I \left(\frac{\omega_s f - \omega_t \cos(\phi) f}{\sin(\phi) \omega_t v_m} \right)} d\hat{\omega}_I \\ &+ j M_1 \frac{\sin(\phi) v_m}{f} \left(\frac{f e^{-j(\omega_t \cos(\phi) - \omega_s) \left(\frac{\omega_v f - z_1 \omega_t}{\sin(\phi) \omega_t} \right)}}{\sin(\phi) \omega_t v_m} \right) \int_b^a \text{sinc}'(\hat{\omega}_I) e^{-j\hat{\omega}_I \left(\frac{\omega_s f - \omega_t \cos(\phi) f}{\sin(\phi) \omega_t v_m} \right)} d\hat{\omega}_I. \end{aligned} \quad (\text{A.6})$$

Note that change of variable is only valid for $\omega_t \neq 0$; the case when $\omega_t = 0$ is addressed below. To aid in the following manipulations, we define a new variable as

$$c = \frac{-f \omega_t \cos(\phi) + \omega_s f}{\sin(\phi) \omega_t v_m}.$$

As a result, (A.6) can be rewritten as

$$\begin{aligned}
P_S(\omega_t, \omega_v) &= \frac{jM_1}{\omega_t} e^{jbc} \int_b^a \text{sinc}'(\hat{\omega}_I) e^{-jc\hat{\omega}_I} d\hat{\omega}_I - \frac{fM_1 e^{jbc}}{\tan(\phi)\omega_t v_m} \int_b^a \text{sinc}(\hat{\omega}_I) e^{-jc\hat{\omega}_I} d\hat{\omega}_I \\
&\stackrel{(i)}{=} \frac{jM_1 e^{jbc}}{\omega_t} \left[\text{sinc}(a) e^{-jac} - \text{sinc}(b) e^{-jbc} \right] \\
&\quad - M_1 \left(\frac{c}{\omega_t} + \frac{f}{\tan(\phi)\omega_t v_m} \right) e^{jbc} \int_b^a \text{sinc}(\hat{\omega}_I) e^{-jc\hat{\omega}_I} d\hat{\omega}_I \\
&= \frac{jM_1 e^{jbc}}{\omega_t} \left[\text{sinc}(a) e^{-jac} - \text{sinc}(b) e^{-jbc} \right] \\
&\quad - \left(\frac{M_1 \omega_s f}{v_m \sin(\phi) \omega_t^2} \right) e^{jbc} \int_b^a \text{sinc}(\hat{\omega}_I) e^{-jc\hat{\omega}_I} d\hat{\omega}_I, \tag{A.7}
\end{aligned}$$

where (i) follows from integration by parts. The final step of the derivation is to rearrange the integral in (A.7) into four separate integrals as follows

$$\begin{aligned}
\int_b^a \text{sinc}(\hat{\omega}_I) e^{-jc\hat{\omega}_I} d\hat{\omega}_I &= \frac{1}{2j} \left(\int_b^a \frac{1 - e^{-j(c+1)\hat{\omega}_I}}{\hat{\omega}_I} d\hat{\omega}_I - \int_b^a \frac{1 - e^{-j(c-1)\hat{\omega}_I}}{\hat{\omega}_I} d\hat{\omega}_I \right) \\
&= \frac{1}{2j} \left(\int_0^{ja(c+1)} \frac{1 - e^{-\hat{\omega}_I}}{\hat{\omega}_I} d\hat{\omega}_I - \int_0^{jb(c+1)} \frac{1 - e^{-\hat{\omega}_I}}{\hat{\omega}_I} d\hat{\omega}_I \right. \\
&\quad \left. - \int_0^{ja(c-1)} \frac{1 - e^{-\hat{\omega}_I}}{\hat{\omega}_I} d\hat{\omega}_I + \int_0^{jb(c-1)} \frac{1 - e^{-\hat{\omega}_I}}{\hat{\omega}_I} d\hat{\omega}_I \right) \\
&= \frac{1}{2j} \left(\zeta \{ja(c+1)\} - \zeta \{jb(c+1)\} \right. \\
&\quad \left. - \zeta \{ja(c-1)\} + \zeta \{jb(c-1)\} \right). \tag{A.8}
\end{aligned}$$

Using identities from [1], the solution to ζ is

$$\zeta \{jh\} = \begin{cases} E_1(jh) + \ln|h| + j\frac{\pi}{2} + \gamma & \text{if } h > 0, \\ E_1^*(j|h|) + \ln|h| - j\frac{\pi}{2} + \gamma & \text{if } h < 0, \\ 0 & \text{if } h = 0. \end{cases}$$

where $h \in \mathbb{R}$, γ is Euler's constant, $E_1(jh)$ is the exponential integral (see [1]) and $E_1^*(jh)$ is its complex conjugate.

Therefore, the expression for the plenoptic spectrum of a finite slanted plane with

complex exponential texture, assuming $\omega_t \neq 0$, is

$$P_S(\omega_t, \omega_v) = \left(\frac{j2v_m}{\omega_t} \left[\text{sinc}(a) e^{-jT(\omega_t \cos(\phi) - \omega_s)} - \text{sinc}(b) \right] + \frac{j\omega_s f}{\sin(\phi)\omega_t^2} \left[\zeta \{jb(c-1)\} - \zeta \{ja(c-1)\} - \zeta \{jb(c+1)\} + \zeta \{ja(c+1)\} \right] e^{jbc} \right) e^{-j\omega_t x_1}. \quad (\text{A.9})$$

If $\omega_t = 0$, the integral in (A.5) may be evaluated directly to obtain

$$P_S(0, \omega_v) = 2v_m T \text{sinc} \left(\frac{\omega_s T}{2} \right) \left[\cos(\phi) \text{sinc}(\omega_v v_m) - j \frac{\sin(\phi) v_m}{f} \text{sinc}'(\omega_v v_m) \right] e^{j\omega_s \frac{T}{2}}.$$

A.2 Bandlimited Texture

Now assuming the texture signal is bandlimited, the spectrum for a slanted plane is derived as follows: we start from the equation for the plenoptic spectrum defined in (A.5) and represent both the texture signal, $g(s)$, and the finite limits of the integral with the function

$$h(s) = \text{rect} \left(\frac{s}{T} - \frac{1}{2} \right) g(s) = \begin{cases} g(s), & \text{if } 0 \leq s \leq T \\ 0, & \text{else.} \end{cases} \quad (\text{A.10})$$

Therefore we obtain the following equation for the plenoptic spectrum

$$P_S(\omega_t, \omega_v) = M_1 \cos(\phi) \int_{-\infty}^{\infty} h(s) \left[\text{sinc} \left(\omega_v v_m - (s \sin(\phi) + z_1) \omega_t \frac{v_m}{f} \right) - j \frac{v_m \tan(\phi)}{f} \text{sinc}' \left(\omega_v v_m - (s \sin(\phi) + z_1) \omega_t \frac{v_m}{f} \right) \right] e^{-j\omega_t \cos(\phi) s} ds. \quad (\text{A.11})$$

At this point, we define the Fourier transform of $h(s)$, with a frequency variable Ω , as follow

$$H(\Omega) = \mathcal{F}_s \{h(s)\} = \int_{-\infty}^{\infty} h(s) e^{-j\Omega s} ds, \quad (\text{A.12})$$

which leads to the following inverse transform

$$h(s) = \mathcal{F}_\Omega^{-1}\{H(\Omega)\} = \frac{1}{2\pi} \int_{-\infty}^{\infty} H(\Omega) e^{j\Omega s} d\Omega. \quad (\text{A.13})$$

By substituting this inverse transform into (A.11) the plenoptic spectrum becomes

$$P_S(\omega_t, \omega_v) = \frac{M_1}{2\pi} \cos(\phi) \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} H(\Omega) e^{j\Omega s} d\Omega \right] \left[\text{sinc} \left(\omega_v v_m - (s \sin(\phi) + z_1) \omega_t \frac{v_m}{f} \right) - j \frac{v_m \tan(\phi)}{f} \text{sinc}' \left(\omega_v v_m - (s \sin(\phi) + z_1) \omega_t \frac{v_m}{f} \right) \right] e^{-j\omega_t \cos(\phi)s} ds. \quad (\text{A.14})$$

From here we change the order of integration and define two new integrals,

$$K_1(\omega_t, \omega_v, \Omega) = \int_{-\infty}^{\infty} \text{sinc} \left(\omega_v v_m - (s \sin(\phi) + z_1) \frac{v_m}{f} \omega_t \right) e^{-j\omega_t \cos(\phi)s} e^{j\Omega s} ds, \quad (\text{A.15})$$

and

$$K_2(\omega_t, \omega_v, \Omega) = \int_{-\infty}^{\infty} \text{sinc}' \left(\omega_v v_m - (s \sin(\phi) + z_1) \frac{v_m}{f} \omega_t \right) e^{-j\omega_t \cos(\phi)s} e^{j\Omega s} ds, \quad (\text{A.16})$$

thus the plenoptic spectrum in (A.14) becomes

$$P_S(\omega_t, \omega_v) = \frac{M_1}{2\pi} \cos(\phi) \int_{-\infty}^{\infty} H(\Omega) \left[K_1(\omega_t, \omega_v, \Omega) - j \frac{\tan(\phi) v_m}{f} K_2(\omega_t, \omega_v, \Omega) \right] d\Omega. \quad (\text{A.17})$$

The integrals $K_1(\omega_t, \omega_v, \Omega)$ and $K_2(\omega_t, \omega_v, \Omega)$ are solved by rearranging them into Fourier transforms in which Ω is the frequency variable.

Starting with $K_1(\omega_t, \omega_v, \Omega)$, we rearrange the integral to obtain

$$K_1(\omega_t, \omega_v, \Omega) = \int_{-\infty}^{\infty} \text{sinc} \left(\frac{\sin(\phi) v_m \omega_t}{f} \left[s - \frac{z_1}{\sin(\phi)} + \frac{\omega_v f}{\sin(\phi) \omega_t} \right] \right) e^{j\omega_t \cos(\phi)s} e^{-j\Omega s} ds, \quad (\text{A.18})$$

which is equivalent to

$$K_1(\omega_t, \omega_v, \Omega) = \frac{1}{2\pi} \mathcal{F}_s \left\{ \text{sinc} \left(\frac{\sin(\phi) v_m \omega_t}{f} \left[s - \frac{z_1}{\sin(\phi)} + \frac{\omega_v f}{\sin(\phi) \omega_t} \right] \right) \right\} * \mathcal{F}_s \left\{ e^{j\omega_t \cos(\phi)s} \right\}. \quad (\text{A.19})$$

Using properties of the Fourier transform, we have

$$\begin{aligned} \mathcal{F}_s \left\{ \text{sinc} \left(\frac{\sin(\phi)v_m\omega_t}{f} \left[s - \frac{z_1}{\sin(\phi)} + \frac{\omega_v f}{\sin(\phi)\omega_t} \right] \right) \right\} \\ = \frac{f\pi}{\sin(\phi)v_m\omega_t} \text{rect} \left(\frac{\Omega f}{2\sin(\phi)v_m\omega_t} \right) e^{-j \left(\frac{z_1}{\sin(\phi)} - \frac{\omega_v f}{\sin(\phi)\omega_t} \right) \Omega}, \end{aligned} \quad (\text{A.20})$$

and

$$\mathcal{F}_s \left\{ e^{j\omega_t \cos(\phi)s} \right\} = \delta(\Omega - \omega_t \cos(\phi)), \quad (\text{A.21})$$

where δ is the delta Dirac. As a result (A.19) becomes

$$K_1(\omega_t, \omega_v, \Omega) = \frac{f\pi}{\sin(\phi)v_m\omega_t} \text{rect} \left(\frac{\Omega f - \omega_t \cos(\phi)f}{2\sin(\phi)v_m\omega_t} \right) e^{-j \left(\frac{z_1}{\sin(\phi)} - \frac{\omega_v f}{\sin(\phi)\omega_t} \right) (\Omega - \omega_t \cos(\phi))}. \quad (\text{A.22})$$

If we now examine $K_2(\omega_t, \omega_v, \Omega)$, it too is rearranged into a Fourier transform to give

$$K_2(\omega_t, \omega_v, \Omega) = \int_{-\infty}^{\infty} \text{sinc}' \left(\frac{\sin(\phi)v_m\omega_t}{f} \left[s - \frac{z_1}{\sin(\phi)} + \frac{\omega_v f}{\sin(\phi)\omega_t} \right] \right) e^{j\omega_t \cos(\phi)s} e^{-j\Omega s} ds. \quad (\text{A.23})$$

Therefore, similar to $K_1(\omega_t, \omega_v, \Omega)$, we have

$$K_2(\omega_t, \omega_v, \Omega) = \frac{1}{2\pi} \mathcal{F}_s \left\{ \text{sinc}' \left(\frac{\sin(\phi)v_m\omega_t}{f} \left[s - \frac{z_1}{\sin(\phi)} + \frac{\omega_v f}{\sin(\phi)\omega_t} \right] \right) \right\} * \mathcal{F}_s \left\{ e^{j\omega_t \cos(\phi)s} \right\}. \quad (\text{A.24})$$

The first Fourier transform in this case is

$$\begin{aligned} \mathcal{F}_s \left\{ \text{sinc}' \left(\frac{\sin(\phi)v_m\omega_t}{f} \left[s - \frac{z_1}{\sin(\phi)} + \frac{\omega_v f}{\sin(\phi)\omega_t} \right] \right) \right\} \\ \stackrel{(i)}{=} \left(\frac{f}{\sin(\phi)v_m\omega_t} \right) \int_{-\infty}^{\infty} \text{sinc}'(\hat{s}) e^{-j \frac{\Omega f}{\sin(\phi)v_m\omega_t} \hat{s}} d\hat{s} e^{-j \left(\frac{z_1}{\sin(\phi)} - \frac{\omega_v f}{\sin(\phi)\omega_t} \right) \Omega} \\ = j\pi \left(\frac{f}{\sin(\phi)v_m\omega_t} \right)^2 \Omega \text{rect} \left(\frac{\Omega f}{2\sin(\phi)v_m\omega_t} \right) e^{-j \left(\frac{z_1}{\sin(\phi)} - \frac{\omega_v f}{\sin(\phi)\omega_t} \right) \Omega}, \end{aligned} \quad (\text{A.25})$$

where step (i) follows from a change of variable. Substituting (A.21) and (A.25) into

(A.24) we obtain

$$K_2(\omega_t, \omega_v, \Omega) = j\pi \left(\frac{f}{\sin(\phi)v_m\omega_t} \right)^2 (\Omega - \omega_t \cos(\phi)) \operatorname{rect} \left(\frac{\Omega f - \omega_t \cos(\phi)f}{2 \sin(\phi)v_m\omega_t} \right) \cdot e^{-j \left(\frac{z_1}{\sin(\phi)} - \frac{\omega_v f}{\sin(\phi)\omega_t} \right) (\Omega - \omega_t \cos(\phi))}. \quad (\text{A.26})$$

Finally, if we substitute (A.22) and (A.26) into (A.17), we obtain

$$P_S(\omega_t, \omega_v) = \frac{M_1}{2} \int_{-\infty}^{\infty} H(\Omega) \left[\frac{f}{\tan(\phi)v_m\omega_t} + \frac{\sin(\phi)v_m}{f} \left(\frac{f}{\sin(\phi)v_m\omega_t} \right)^2 (\Omega - \omega_t \cos(\phi)) \right] \cdot \operatorname{rect} \left(\frac{\Omega f - \omega_t \cos(\phi)f}{2 \sin(\phi)v_m\omega_t} \right) e^{-j \left(\frac{z_1}{\sin(\phi)} - \frac{\omega_v f}{\sin(\phi)\omega_t} \right) (\Omega - \omega_t \cos(\phi))} d\Omega. \quad (\text{A.27})$$

Therefore, rearranging the above, the plenoptic spectrum for a slanted plane with bandlimited texture is

$$P_S(\omega_t, \omega_v) = \int_{-\infty}^{\infty} H(\Omega) \frac{f\Omega}{\sin(\phi)\omega_t^2} \operatorname{rect} \left(\frac{\Omega f - \omega_t \cos(\phi)f}{2 \sin(\phi)v_m\omega_t} \right) e^{-j\Omega \left(\frac{z_1}{\sin(\phi)} - \frac{\omega_v f}{\sin(\phi)\omega_t} \right)} d\Omega \cdot e^{-j\omega_t \left(x_1 - \frac{z_1}{\tan(\phi)} \right)} e^{-j\omega_v \left(\frac{f}{\tan(\phi)} \right)}. \quad (\text{A.28})$$

Appendix B

Properties of the $P_S(\omega_t, \omega_v)$

In this appendix we determine two properties of the plenoptic spectrum for a slanted plane with complex exponential texture. The first is that the plenoptic spectrum for a slanted plane, $P_S(\omega_t, \omega_v)$, tends to the spectrum for a fronto-parallel plane, $P_{FPP}(\omega_t, \omega_v)$, in the limit as ϕ tends to zero, which we term the consistency property. The second is that, assuming a maximum and minimum depth of z_{\max} and z_{\min} respectively, then the magnitude of the plenoptic spectrum for a slanted plane is independent of the sign of ϕ . This second relationship is termed the symmetry property.

B.1 The Consistency Property

This property states that the plenoptic spectrum for a slanted plane, $P_S(\omega_t, \omega_v)$, tends to the spectrum for a fronto-parallel plane, $P_{FPP}(\omega_t, \omega_v)$, in the limit as ϕ tends to zero, i.e.

$$\lim_{\phi \rightarrow 0} \left\{ P_S(\omega_t, \omega_v) \right\} = P_{FPP}(\omega_t, \omega_v). \quad (\text{B.1})$$

To proof this statement we start from (A.6), and take the limit as ϕ tends to zero, hence

$$\begin{aligned} \lim_{\phi \rightarrow 0} \left\{ P_S(\omega_t, \omega_v) \right\} &= \lim_{\phi \rightarrow 0} \left\{ \frac{j}{\omega_t} \left[\text{sinc}(a) e^{-jac} - \text{sinc}(b) e^{-jbc} \right] e^{jbc} \right. \\ &\quad \left. - \left(\frac{\omega_s f}{v_m \omega_t^2} \right) e^{jbc} \int_b^a \frac{\text{sinc}(\hat{\omega}_I)}{\sin(\phi)} e^{-jc\hat{\omega}_I} d\hat{\omega}_I \right\} \\ &\stackrel{(i)}{=} \frac{j}{\omega_t} \text{sinc}(b) \left[e^{-jT(\omega_t - \omega_s)} - 1 \right] \\ &\quad - \left(\frac{\omega_s f}{v_m \omega_t^2} \right) \lim_{\phi \rightarrow 0} \left\{ e^{jbc} \int_b^a \frac{\text{sinc}(\hat{\omega}_I)}{\sin(\phi)} e^{-jc\hat{\omega}_I} d\hat{\omega}_I \right\}, \quad (\text{B.2}) \end{aligned}$$

where (i) follows from the fact that a tends to b when ϕ tends to zero. Now the argument of the limit in (B.2) can be rearranged as follows

$$\begin{aligned} e^{jbc} \int_b^a \frac{\text{sinc}(\hat{\omega}_I)}{\sin(\phi)} e^{-j\hat{\omega}_I c} d\hat{\omega}_I &= \frac{1}{2j \sin(\phi)} \int_b^a \frac{e^{j\hat{\omega}_I} - e^{-j\hat{\omega}_I}}{\hat{\omega}_I} e^{-j\hat{\omega}_I c + jbc} d\hat{\omega}_I \\ &= \frac{1}{2j \sin(\phi)} \left[\int_b^a \frac{e^{-j\hat{\omega}_I(c-1)+jbc}}{\hat{\omega}_I} d\hat{\omega}_I - \int_b^a \frac{e^{-j\hat{\omega}_I(c+1)+jbc}}{\hat{\omega}_I} d\hat{\omega}_I \right] \\ &= \frac{1}{2j \sin(\phi)} \left[\int_{-jb}^{j(a-b)c-ja} \frac{e^{-\omega_B}}{\omega_B + jbc} d\omega_B - \int_{jb}^{j(a-b)c+ja} \frac{e^{-\omega_B}}{\omega_B + jbc} d\omega_B \right], \quad (\text{B.3}) \end{aligned}$$

where the last step follows from the substitution $\omega_B = j\hat{\omega}_I(c-1) - jbc$ and $\omega_B = j\hat{\omega}_I(c+1) - jbc$ in the second integral. If we expand the following quantities $(a-b)c = -T(\omega_s - \omega_t \cos(\phi))$ and

$$\left(\frac{1}{\sin(\phi)} \right) \left(\frac{1}{\omega_B + jbc} \right) = \frac{v_m \omega_t}{\omega_B \omega_t v_m \sin(\phi) + jb(\omega_s f - f \omega_t \cos(\phi))},$$

and then set $\phi = 0$, (B.3) becomes

$$\begin{aligned} &\left(\frac{v_m \omega_t}{2j} \right) \left[\int_{-jb}^{-j(\omega_s - \omega_t)T - jb} \frac{e^{-\omega_B}}{jbf(\omega_s - \omega_t)} d\omega_B - \int_{jb}^{-j(\omega_s - \omega_t)T + jb} \frac{e^{-\omega_B}}{jbf(\omega_s - \omega_t)} d\omega_B \right] \\ &= -\frac{v_m \omega_t}{2bf(\omega_s - \omega_t)} \left[e^{jb} - e^{-jT(\omega_t - \omega_s) + jb} - e^{-jb} + e^{-jT(\omega_t - \omega_s) - jb} \right] \\ &= -\frac{v_m \omega_t}{2bf(\omega_s - \omega_t)} \left(e^{-jT(\omega_t - \omega_s)} - 1 \right) \left(e^{-jb} - e^{jb} \right). \quad (\text{B.4}) \end{aligned}$$

Substituting (B.4) into (B.2), results in

$$\begin{aligned}
\lim_{\phi \rightarrow 0} \{P_S(\omega_t, \omega_v)\} &= \left[e^{-jT(\omega_t - \omega_s)} - 1 \right] \left(\text{sinc}(b) \frac{j}{\omega_t} + \frac{\omega_s}{2b(\omega_s - \omega_t)\omega_t} (e^{-jb} - e^{jb}) \right) \\
&= \text{sinc}(b) \left[e^{-jT(\omega_t - \omega_s)} - 1 \right] \left[\frac{j}{\omega_t} + \frac{j\omega_s}{\omega_t(\omega_t - \omega_s)} \right] \\
&= \frac{2T \text{sinc}(b)}{2j(\omega_t - \omega_s)T} \left[1 - e^{-jT(\omega_t - \omega_s)} \right] \\
&= T \text{sinc}(b) \text{sinc}\left(\frac{T(\omega_t - \omega_s)}{2}\right) e^{-j\frac{T}{2}(\omega_t - \omega_s)} \\
&= P_{FPP}(\omega_t, \omega_v).
\end{aligned} \tag{B.5}$$

B.2 The Symmetry Property

Consider two slanted planes that are almost identical except for the sign of ϕ . In other words they have the same minimum and maximum depths, z_{\min} and z_{\max} respectively, the same spatial positions, x_1 and x_2 , and the same texture $g(s) = e^{j\omega_s s}$, however one has an angle of slant $\phi = |\phi_1|$ and the other has an angle $\phi = -|\phi_1|$. Therefore the symmetry property states that the magnitude of the plenoptic spectra relating to both scenes is the equivalent, i.e.

$$\left| \underbrace{P_S(\omega_t, \omega_v)}_{\text{when } \phi = -|\phi_1|} \right| = \left| \underbrace{P_S(\omega_t, \omega_v)}_{\text{when } \phi = |\phi_1|} \right|. \tag{B.6}$$

To prove this statement let us examine the plenoptic spectra for both scenes. Starting with the plane that has a positive angle of slant, its plenoptic spectrum is

$$\begin{aligned}
\underbrace{P_S(\omega_t, \omega_v)}_{\text{when } \phi = |\phi_1|} &= \left(\frac{j2v_m}{\omega_t} \left[\text{sinc}(a_1) e^{-j(a_1 - b_1)c_1} - \text{sinc}(b_1) \right] + \frac{j\omega_s f}{\sin|\phi_1|\omega_t^2} \left[\zeta \{jb_1(c_1 - 1)\} \right. \right. \\
&\quad \left. \left. - \zeta \{ja_1(c_1 - 1)\} - \zeta \{jb_1(c_1 + 1)\} + \zeta \{ja_1(c_1 + 1)\} \right] e^{jb_1c_1} \right) e^{-j\omega_t x_1}, \tag{B.7}
\end{aligned}$$

where the parameters a_1 , b_1 and c_1 are defined as

$$a_1 = \omega_v v_m - \omega_t \frac{z_{\max} v_m}{f}, \quad b_1 = \omega_v v_m - \omega_t \frac{z_{\min} v_m}{f}, \quad \text{and} \quad c_1 = \frac{\omega_s f - f\omega_t \cos|\phi_1|}{\sin|\phi_1|\omega_t v_m}.$$

Moving to the second scene, its plenoptic spectrum is obtained by substituting $\phi = -|\phi_1|$ and switching a_1 with b_1 to give

$$\underbrace{P_S(\omega_t, \omega_v)}_{\text{when } \phi = -|\phi_1|} = \left(\frac{j2v_m}{\omega_t} \left[\text{sinc}(b_1) e^{j(b_1 - a_1)c_1} - \text{sinc}(a_1) \right] - \frac{j\omega_s f}{\sin|\phi_1| \omega_t^2} \left[\zeta \{-jb_1(c_1 - 1)\} \right. \right. \\ \left. \left. - \zeta \{-ja_1(c_1 - 1)\} - \zeta \{-jb_1(c_1 + 1)\} + \zeta \{-ja_1(c_1 + 1)\} \right] e^{-ja_1 c_1} \right) e^{-j\omega_t x_1}. \quad (\text{B.8})$$

Now taking the absolute value of (B.8) we obtain

$$\left| \underbrace{P_S(\omega_t, \omega_v)}_{\text{when } \phi = -|\phi_1|} \right| = \left| -\frac{j2v_m}{\omega_t} \left[\text{sinc}(a_1) - \text{sinc}(b_1) e^{j(b_1 - a_1)c_1} \right] - \frac{j\omega_s f}{\sin|\phi_1| \omega_t^2} \left[\zeta \{-jb_1(c_1 - 1)\} \right. \right. \\ \left. \left. - \zeta \{-ja_1(c_1 - 1)\} - \zeta \{-jb_1(c_1 + 1)\} + \zeta \{-ja_1(c_1 + 1)\} \right] e^{-ja_1 c_1} \right| \\ = \left| \frac{2v_m}{\omega_t} \left[\text{sinc}(a_1) e^{ja_1 c_1} - \text{sinc}(b_1) e^{jb_1 c_1} \right] + \frac{\omega_s f}{\sin|\phi_1| \omega_t^2} \left[\zeta \{-jb_1(c_1 - 1)\} \right. \right. \\ \left. \left. - \zeta \{-ja_1(c_1 - 1)\} - \zeta \{-jb_1(c_1 + 1)\} + \zeta \{-ja_1(c_1 + 1)\} \right] \right|. \quad (\text{B.9})$$

At this point, using [1], we note that

$$\zeta \{-jh\} = \zeta^* \{jh\}, \quad (\text{B.10})$$

which leads to the following relationships

$$\underbrace{|P_S(\omega_t, \omega_v)|}_{\text{when } \phi = -|\phi|} = \underbrace{|P_S^*(\omega_t, \omega_v)|}_{\text{when } \phi = |\phi|} = \underbrace{|P_S(\omega_t, \omega_v)|}_{\text{when } \phi = |\phi|}, \quad (\text{B.11})$$

hence the symmetry property of the plenoptic spectrum for a slanted plane.

Appendix C

The Plenoptic Spectrum with Rotated Camera Line

In this appendix we derive the plenoptic spectrum for a slanted plane when the camera line is rotated around a point x_r at an angle α . To determine this spectrum we require a relation similar to (2.7), i.e. a relationship between the point a light ray leaves the scene surface at $(x, z(x))$ and arrives at (t, v) . Figure C.1 illustrates a light ray fulfilling this relationship for a rotated camera line. The relationship is determined as follows: first we define the depth of the rotated camera line relative to the x -axis, which is

$$z_t(x) = (x - x_r) \tan(\alpha). \quad (\text{C.1})$$

Using this expression, the point the light ray intersects the camera line, t , is define as follows

$$\begin{aligned} t &= \frac{x}{\cos(\alpha)} - \left(z(x) - z_t(x) \right) \frac{\sin(\beta)}{\sin(\pi/2 - \theta)} \\ &= \frac{x}{\cos(\alpha)} - \left(z(x) - (x - x_r) \tan(\alpha) \right) \frac{\sin(\beta)}{\cos(\theta)}, \end{aligned} \quad (\text{C.2})$$

where β is the angle the light ray leaves the surface relative to the z -axis, see Figure C.1. The expression in (C.2), however, does not include the pixel coordinate v . To

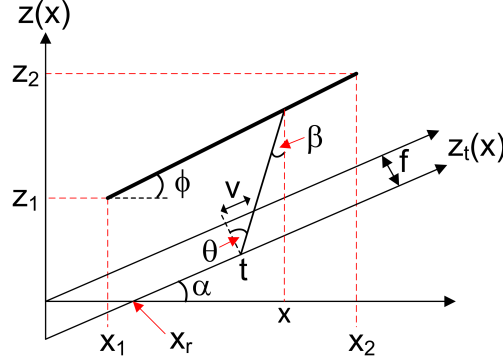


Figure C.1: Diagram illustrating a light ray travelling from the scene to a camera at position t . The camera line is rotated around a point x_r at an angle α . Note that θ is the viewing angle and v is the relating pixel position.

introduce v , we notice that $\beta = \theta - \alpha$, thus the following is true

$$\frac{\sin(\beta)}{\cos(\theta)} = \tan(\theta) \cos(\alpha) - \sin(\alpha) = \frac{v \cos(\alpha)}{f} - \sin(\alpha), \quad (\text{C.3})$$

where $\tan(\theta) = v/f$. As a result (C.2) becomes

$$\begin{aligned} t &= \frac{x}{\cos(\alpha)} - \left(z(x) - (x - x_p) \tan(\alpha) \right) \left(\frac{v}{f} \cos(\alpha) - \sin(\alpha) \right) \\ &= \frac{x}{\cos(\alpha)} + z(x) \sin(\alpha) - (x - x_p) \frac{\sin^2(\alpha)}{\cos(\alpha)} - \frac{v}{f} \left(z(x) \sin(\alpha) - (x - x_p) \sin(\alpha) \right). \end{aligned} \quad (\text{C.4})$$

Notice that if $\alpha = 0$ then (C.4) reverts back to the geometric relationship stated in (2.7).

Having determined the relationship in (C.4), the plenoptic spectrum for a slanted plane when the camera line is rotated is derived as follows. Starting from the definition of the plenoptic spectrum

$$\begin{aligned} P(\omega_t, \omega_v) &= \mathcal{F}_{t,v} \{p(t, v)\} \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(t, v) e^{-j(\omega_t t + \omega_v v)} dt dv, \end{aligned} \quad (\text{C.5})$$

we use the geometric relationship in (C.4) to change the variable of integration from t

to x . The Jacobian generated from this change of variable is

$$dt = \left(\frac{1 - \sin^2(\alpha)}{\cos(\alpha)} + z'(x) \sin(\alpha) \right) dx + \frac{v}{f} \left(\sin(\alpha) - z'(x) \cos(\alpha) \right) dx. \quad (\text{C.6})$$

At the same time we also switch from v to \bar{v} using $\bar{v} = v/f$. As a result we are able to relate $p(t, v)$ to $l_x(x, \bar{v})$. Consequently (C.5) becomes

$$\begin{aligned} P(\omega_t, \omega_v) = & \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} l_x(x, \bar{v}) e^{-j\omega_t \left[\frac{x}{\cos(\alpha)} + z(x) \sin(\alpha) - (x - x_r) \frac{\sin^2(\alpha)}{\cos(\alpha)} - \bar{v} \left(z(x) \cos(\alpha) - (x - x_r) \sin(\alpha) \right) \right]} \\ & \cdot e^{-j\omega_v \bar{v} f} \left[\left(\frac{1 - \sin^2(\alpha)}{\cos(\alpha)} + z'(x) \sin(\alpha) \right) + \bar{v} \left(\sin(\alpha) - z'(x) \cos(\alpha) \right) \right] f dx d\bar{v}. \end{aligned} \quad (\text{C.7})$$

Now, assuming a Lambertian scene and applying the finite constraints, then $l_x(x, \bar{v}) = l_x(x)$, $x \in [x_1, x_2]$ and $\bar{v} \in [-\bar{v}_m, \bar{v}_m]$, which results in the following

$$\begin{aligned} P(\omega_t, \omega_v) = & f \int_{x=x_1}^{x=x_2} l_x(x) e^{-j \left(\frac{x}{\cos(\alpha)} + z(x) \sin(\alpha) - (x - x_r) \frac{\sin^2(\alpha)}{\cos(\alpha)} \right) \omega_t} \\ & \cdot \int_{\bar{v}=-\bar{v}_m}^{\bar{v}=\bar{v}_m} \left[\frac{1 - \sin^2(\alpha)}{\cos(\alpha)} + z'(x) \sin(\alpha) + \bar{v} \left(\sin(\alpha) - z'(x) \cos(\alpha) \right) \right] \\ & \cdot e^{-j \left(\omega_v f - \omega_t \left(z(x) \cos(\alpha) - (x - x_p) \sin(\alpha) \right) \right) \bar{v}} d\bar{v} dx. \end{aligned} \quad (\text{C.8})$$

In order to solve (C.8), we start by defining the FFoV constraint as the following rectangular windowing function

$$\text{rect} \left(\frac{f}{2v_m} \bar{v} \right) = \begin{cases} 1 & , \text{ if } |\bar{v}| \leq \frac{v_m}{f} \\ 0 & , \text{ else} \end{cases} \quad (\text{C.9})$$

whose corresponding Fourier transform is

$$\mathcal{F}_{\bar{v}} \left\{ \text{rect} \left(\frac{f}{2v_m} \bar{v} \right) \right\} = \frac{2v_m}{f} \text{sinc} \left(\frac{v_m}{f} \omega_v \right).$$

Using this definition, the integral in \bar{v} can be evaluated as follows

$$\begin{aligned}
I_3 &= f \int_{-\infty}^{\infty} \text{rect} \left(\frac{f}{2v_m} \bar{v} \right) \left[\bar{v} \left(\sin(\alpha) - z'(x) \cos(\alpha) \right) + \frac{1 - \sin^2(\alpha)}{\cos(\alpha)} \right. \\
&\quad \left. + z'(x) \sin(\alpha) \right] e^{-j \left(\omega_v f - \omega_t \left(z(x) \cos(\alpha) - (x - x_p) \sin(\alpha) \right) \right) \bar{v}} d\bar{v} \\
&\stackrel{(i)}{=} f \left[\frac{1 - \sin^2(\alpha)}{\cos(\alpha)} + z'(x) \sin(\alpha) \right] \mathcal{F}_{\bar{v}} \left\{ \text{rect} \left(\frac{f}{2v_m} \bar{v} \right) \right\} \\
&\quad + j f \left[\sin(\alpha) - z'(x) \cos(\alpha) \right] \frac{\partial}{\partial \omega_v} \left\{ \mathcal{F}_{\bar{v}} \left\{ \text{rect} \left(\frac{f}{2v_m} \bar{v} \right) \right\} \right\} \\
&= 2v_m \left(\frac{1 - \sin^2(\alpha)}{\cos(\alpha)} + z'(x) \sin(\alpha) \right) \text{sinc}(\omega_r) + j \frac{v_m^2}{f} \left(\sin(\alpha) - z'(x) \cos(\alpha) \right) \text{sinc}'(\omega_r),
\end{aligned} \tag{C.10}$$

where step (i) follows from properties of the Fourier transform, $\text{sinc}'(\omega_r)$ is the first derivative of the sinc function with respect to ω_r , and

$$\omega_r = \omega_v v_m - \omega_t v_m \left(\frac{z(x) \cos(\alpha) + (x - x_r) \sin(\alpha)}{f} \right).$$

Substituting the solution for I_3 into (C.8), we obtain

$$\begin{aligned}
P(\omega_t, \omega_v) &= M_4 \int_{x_1}^{x_2} l_x(x) \left[\left(\frac{1}{\cos(\alpha)} + \sin(\alpha) \left(z'(x) - \tan(\alpha) \right) \right) \text{sinc}(\omega_r) \right. \\
&\quad \left. + \frac{j v_m}{f} \left(\sin(\alpha) - z'(x) \cos(\alpha) \right) \text{sinc}'(\omega_r) \right] e^{-j \left(\frac{x}{\cos(\alpha)} + z(x) \sin(\alpha) - x \frac{\sin^2(\alpha)}{\cos(\alpha)} \right) \omega_t} dx,
\end{aligned} \tag{C.11}$$

where

$$M_4 = 2v_m e^{-j x_r \omega_t \sin(\alpha) \tan(\alpha)}.$$

Now, using the scene geometry equations (3.1), we change the variable of integration in (C.11) from x to the curvilinear coordinate s and substitute $z'(x) = \tan(\phi)$, resulting

in

$$\begin{aligned}
P_{Sr}(\omega_t, \omega_v) = M_4 \int_0^T g(s) \cos(\phi) & \left[\left(\frac{1}{\cos(\alpha)} + \sin(\alpha) (\tan(\phi) - \tan(\alpha)) \right) \text{sinc}(\omega_r) \right. \\
& + \frac{jv_m}{f} \left(\sin(\alpha) - \tan(\phi) \cos(\alpha) \right) \text{sinc}'(\omega_r) \Big] \\
& \cdot e^{-j \left[s \left(\frac{\cos(\phi)}{\cos(\alpha)} + \sin(\phi) \sin(\alpha) - \frac{\cos(\phi) \sin^2(\alpha)}{\cos(\alpha)} \right) + \frac{x_1}{\cos(\alpha)} + z_1 \sin(\alpha) - x_1 \frac{\sin^2(\alpha)}{\cos(\alpha)} \right] \omega_t} ds,
\end{aligned} \tag{C.12}$$

where $g(s) = l(s) = l_x(x)$ and

$$\omega_r = \omega_v v_m - \frac{\omega_t v_m}{f} \left(s \sin(\phi - \alpha) + z_1 \cos(\alpha) - x_1 \sin(\alpha) + x_r \sin(\alpha) \right).$$

This integral can be simplified using the following trigonometric identities

$$\begin{cases} \sin(\alpha) - \tan(\phi) \cos(\alpha) & = \frac{\sin(\alpha - \phi)}{\cos(\phi)}, \\ \sec(\alpha) - \sin(\alpha) \tan(\phi) - \frac{\sin^2(\alpha)}{\cos(\alpha)} & = \frac{\cos(\alpha - \phi)}{\cos(\phi)}, \\ \frac{\cos(\phi)}{\cos(\alpha)} + \sin(\alpha) \sin(\phi) - \frac{\cos(\phi) \sin^2(\alpha)}{\cos(\alpha)} & = \cos(\alpha - \phi), \\ \sec(\alpha) - \frac{\sin^2(\alpha)}{\cos(\alpha)} & = \cos(\alpha). \end{cases}$$

Therefore, using the above and substituting $g(s) = e^{j\omega_s s}$, (C.12) becomes

$$\begin{aligned}
P_{Sr}(\omega_t, \omega_v) = M_4 \int_0^T & \left[\cos(\alpha - \phi) \text{sinc}(\omega_r) + \frac{jv_m}{f} \sin(\alpha - \phi) \text{sinc}'(\omega_r) \right] \\
& \cdot e^{j\omega_s s} e^{-j(s \cos(\alpha - \phi) + x_1 \cos(\phi) + z_1 \sin(\alpha)) \omega_t} ds.
\end{aligned} \tag{C.13}$$

In order to solve (C.13), we change the variable of integration from s to ω_r . As such we have the following substitution and Jacobian

$$s = \frac{\omega_v f - z_1 \cos(\alpha) \omega_t + x_1 \sin(\alpha) \omega_t - x_r \sin(\alpha) \omega_t}{\omega_t \sin(\phi - \alpha)} - \omega_r \frac{f}{\omega_t v_m \sin(\phi - \alpha)},$$

and

$$ds = \frac{-f}{\omega_t v_m \sin(\phi - \alpha)} d\omega_r.$$

Under this change of variable, the limits of integration become

$$\begin{aligned} s = 0 &\longleftrightarrow \omega_r = \omega_v v_m - \frac{\omega_t v_m}{f} \left(z_1 \cos(\alpha) - x_1 \sin(\alpha) + x_r \sin(\alpha) \right) \\ &= b_r, \\ s = T &\longleftrightarrow \omega_r = \omega_v v_m - \frac{\omega_t v_m}{f} \left(z_1 \cos(\alpha) - x_1 \sin(\alpha) + x_r \sin(\alpha) + T \sin(\phi - \alpha) \right) \\ &= a_r. \end{aligned}$$

Final, to accompany this substitution, we define the following parameter

$$c_r = \frac{\omega_s f - f \cos(\alpha - \phi) \omega_t}{\omega_t v_m \sin(\phi - \alpha)},$$

and point out the following relationship

$$b_r c_r = -(\cos(\alpha - \phi) - \omega_s) \left(\frac{\omega_v f - z_1 \omega_t \cos(\alpha) + x_1 \omega_t \sin(\alpha) - x_r \sin(\alpha) \omega_t}{\omega_t \sin(\phi - \alpha)} \right).$$

Therefore, applying this change of variable to (C.13), we obtain

$$\begin{aligned} P_{Sr}(\omega_t, \omega_v) &= \int_{b_r}^{a_r} \left(\frac{-M_5 f}{\omega_t v_m \sin(\phi - \alpha)} \right) \left[\cos(\alpha - \phi) \text{sinc}(\omega_r) + \frac{j v_m}{f} \sin(\alpha - \phi) \text{sinc}'(\omega_r) \right] \\ &\quad \cdot e^{-j(\cos(\alpha - \phi) \omega_t - \omega_s) \left[\frac{\omega_v f - z_1 \cos(\alpha) \omega_t + x_1 \sin(\alpha) \omega_t - x_r \sin(\alpha) \omega_t}{\omega_t \sin(\phi - \alpha)} - \omega_r \frac{f}{\omega_t v_m \sin(\phi - \alpha)} \right]} d\omega_r \\ &= M_5 \int_{b_r}^{a_r} \left[\left(\frac{-f}{\omega_t v_m \tan(\phi - \alpha)} \right) \text{sinc}(\omega_r) + \frac{j}{\omega_t} \text{sinc}'(\omega_r) \right] e^{j b_r c_r} e^{-j c_r \omega_r} d\omega_r \\ &\stackrel{(i)}{=} M_5 \left(\frac{j}{\omega_t} \left[\text{sinc}(a_r) e^{-j a_r c_r} - \text{sinc}(b_r) e^{-j b_r c_r} \right] e^{j b_r c_r} \right. \\ &\quad \left. - \left(\frac{f}{v_m \omega_t \tan(\phi - \alpha)} + \frac{c_r}{\omega_t} \right) e^{j b_r c_r} \int_{b_r}^{a_r} \text{sinc}(\omega_r) e^{-j \omega_r c_r} d\omega_r \right). \end{aligned} \tag{C.14}$$

where (i) follows from integration by parts and

$$M_5 = M_4 e^{-j(x_{min} \cos(\alpha) + z_{min} \sin(\alpha)) \omega_t} = 2v_m e^{-j \omega_t (x_1 \cos(\alpha) + z_1 \sin(\alpha) - x_1 \sin(\alpha) \tan(\alpha))}.$$

Finally, using (A.8), the plenoptic spectrum for a slanted plane observed from a camera line rotated around a point x_r at an angle α is

$$P_{Sr}(\omega_t, \omega_v) = e^{-j\omega_t t_1} \left(\frac{j2v_m}{\omega_t} \left[\text{sinc}(a_r) e^{-jT(\omega_t \cos(\phi-\alpha) - \omega_s)} - \text{sinc}(b_r) \right] + \frac{j\omega_s f e^{jb_r c_r}}{\sin(\phi - \alpha) \omega_t^2} \right. \\ \left. \cdot \left[\zeta \{jb_r(c_r - 1)\} - \zeta \{ja_r(c_r - 1)\} - \zeta \{jb_r(c_r + 1)\} + \zeta \{ja_r(c_r + 1)\} \right] \right), \quad (\text{C.15})$$

if $\omega_t \neq 0$, else

$$P_{Sr}(0, \omega_v) = 2v_m T \text{sinc}\left(\frac{\omega_s T}{2}\right) \left[\cos(\phi - \alpha) \text{sinc}(\omega_v v_m) - j \frac{\sin(\phi - \alpha) v_m}{f} \text{sinc}'(\omega_v v_m) \right] e^{j\omega_s \frac{T}{2}},$$

where $t_1 = x_1 \cos(\alpha) + z_1 \sin(\alpha) + x_r \sin(\alpha) \tan(\alpha)$.

Appendix D

Determining the Essential Bandwidth for $\text{sinc}'(\omega)$

In this appendix we determine an estimate of the essential bandwidth for the first derivative of a sinc function. This estimate is calculated numerically as follows: given $\text{sinc}'(\omega)$, we first assume the essential bandwidth is symmetric around origin. The problem, therefore, can be reduced to determining a parameter W that satisfies

$$\arg_W \left\{ \int_{-W}^W |\text{sinc}'(\omega)|^2 d\omega = 0.9 \cdot \int_{-\infty}^{\infty} |\text{sinc}'(\omega)|^2 d\omega \right\}. \quad (\text{D.1})$$

To solve the above, we calculate the overall energy of $\text{sinc}'(\omega)$ as follows

$$\begin{aligned} \int_{-\infty}^{\infty} |\text{sinc}'(\omega)|^2 d\omega &= \lim_{W \rightarrow \infty} \left\{ \int_{-W}^W |\text{sinc}'(\omega)|^2 d\omega \right\} \\ &= \lim_{W \rightarrow \infty} \left\{ \frac{2}{3} \text{Si}(2W) + \frac{\cos(2W)(W^2 + 1) - 3W^2 - 2W \sin(2W)}{3W^3} \right\} \\ &= \frac{\pi}{3}, \end{aligned} \quad (\text{D.2})$$

where $\text{Si}(W)$ is the Sine integral, see [1]. Using this value, we numerically compute the left-hand integral in (D.1) and determine the percentage of the overall energy as a function of W . This is illustrated in Figure D.1. The figure shows the ESD of $\text{sinc}'(\omega)$ in Figure D.1(a) and the percentage of energy as a function of W in Figure D.1(b).

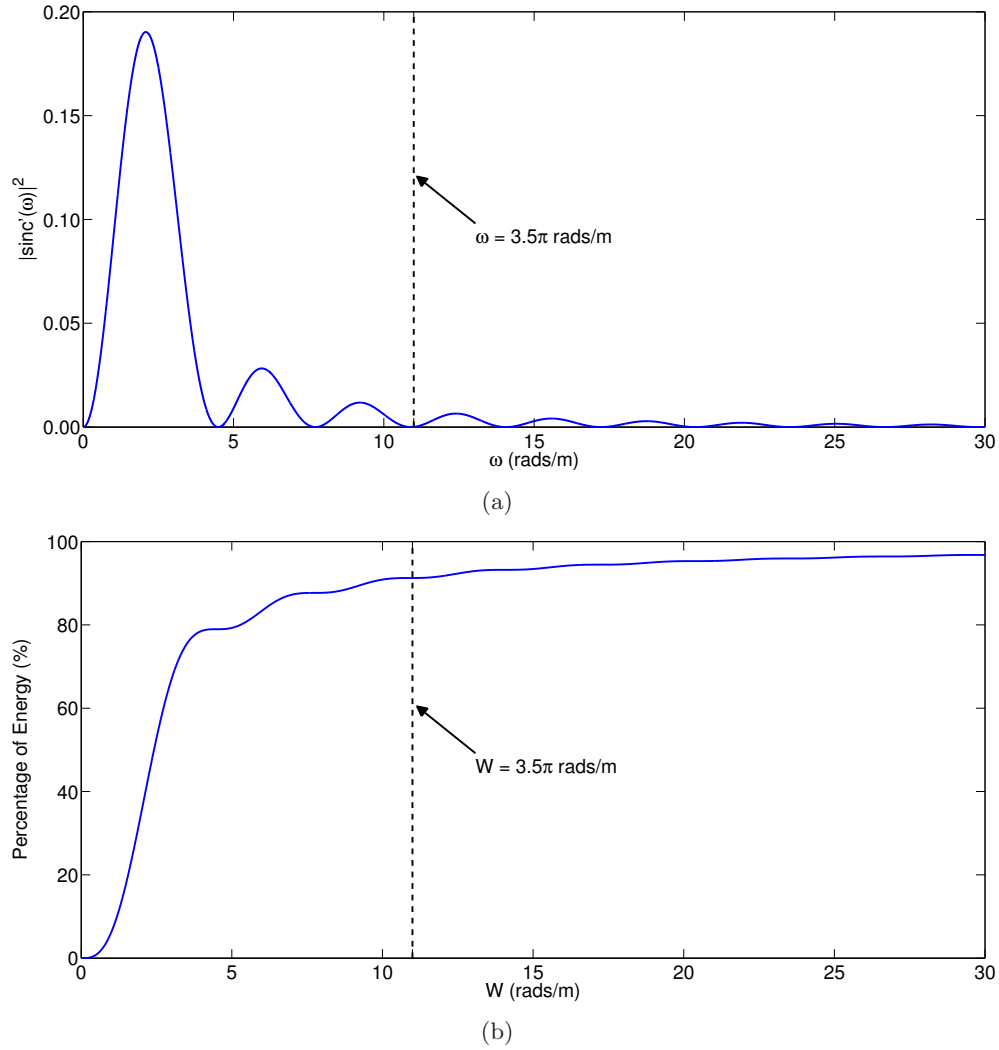


Figure D.1: Graph (a) illustrates the Energy Spectral Density (ESD) for $\text{sinc}'(\omega)$. The percentage of the overall energy within a bandwidth W as a function of W is shown in (b).

From this figure, we choose $W = 3.5\pi$ as an approximate solution to (D.1). Using this estimate, the essential bandwidth for $\text{sinc}'(\omega)$ is

$$\{\omega : |\omega| \leq 3.5\pi\}. \quad (\text{D.3})$$

Although this essential bandwidth contains slightly more energy than necessary, 91.3%, it also corresponds to a zero crossing in $\text{sinc}'(\omega)$. Therefore it has a nice consistency when compared with the essential bandwidth for a sinc function.

Appendix E

The Cardinal Operator

Given an arbitrary MOMS interpolator, the cardinal operator \mathcal{C} constructs a cardinal basis. This cardinal basis allows the interpolation of an arbitrary function from its samples rather than a set of coefficients relating to the MOMS interpolator. The basis is known as the cardinal version of the MOMS. In this appendix we define the cardinal operator and show how the cardinal MOMS is constructed. This definition is based on the construction of the cardinal B-spline presented in [61].

Starting with a MOMS interpolator of order M , we sample at integer values to obtain its discrete version:

$$d_{\text{MOMS}}^{\{M\}}(i) = \varphi_{\text{MOMS}}^{\{M\}}(x) \Big|_{x=i}, \quad (\text{E.1})$$

where $i \in \mathbb{Z}$. The corresponding Z-transform of $d_{\text{MOMS}}^{\{M\}}(i)$ is

$$\begin{aligned} D_{\text{MOMS}}^{\{M\}}(z) &= \mathcal{Z}_i \left\{ d_{\text{MOMS}}^{\{M\}}(i) \right\} \\ &= \sum_{i \in \mathbb{Z}} d_{\text{MOMS}}^{\{M\}}(i) z^{-i}. \end{aligned} \quad (\text{E.2})$$

From this Z-transform, we define the following function

$$\xi_{\text{MOMS}}^{\{M\}}(i) = \mathcal{Z}_z^{-1} \left\{ \frac{1}{D_{\text{MOMS}}^{\{M\}}(z)} \right\}. \quad (\text{E.3})$$

Finally, using this function, the cardinal MOMS of order M is

$$\begin{aligned}\eta_{\text{MOMS}}^{\{M\}}(x) &= \mathcal{C} \left\{ \varphi_{\text{MOMS}}^{\{M\}}(x) \right\} \\ &= \sum_{i \in \mathbb{Z}} \xi_{\text{MOMS}}^{\{M\}}(i) \varphi_{\text{MOMS}}^{\{M\}}(x - i).\end{aligned}\tag{E.4}$$

In the Fourier domain this cardinal MOMS has the following frequency response

$$\mathcal{F}_x \left\{ \eta_{\text{MOMS}}^{\{M\}}(x) \right\} = \frac{\Phi_{\text{MOMS}}^{\{M\}}(\omega)}{D_{\text{MOMS}}^{\{M\}}(e^{j\omega})},\tag{E.5}$$

where $\Phi_{\text{MOMS}}^{\{M\}}(\omega)$ is the Fourier transform of $\varphi_{\text{MOMS}}^{\{M\}}(x)$.

Appendix F

Rendered Images for the Planar Scene

In this appendix, we present six examples of rendered images from the EPI-volume of a planar scene. The EPI-volume is described in Section 5.5.1, and the planar scene is shown in Figure 5.11. For each example, we render the target image using three different sampling and reconstruction methods: uniform sampling and reconstruction using a 3rd order I-MOMS; uniform sampling and reconstruction using the baseline algorithm; and adaptive sampling and reconstruction using our adaptive sampling algorithm. The six target images and their renderings are shown in Figures F.1, F.2, F.3, F.4, F.5 and F.6. Similar to Section 5.5.1, each figure shows the original (target) image in part (a), its rendering using uniform sampling with I-MOMS in (b), its rendering using the baseline algorithm in (c) and its rendering using adaptive sampling in (d). The figures also show the corresponding PSNR for each rendered image. Note that, in all of the figures, the rendered images are generated using 37 original images.

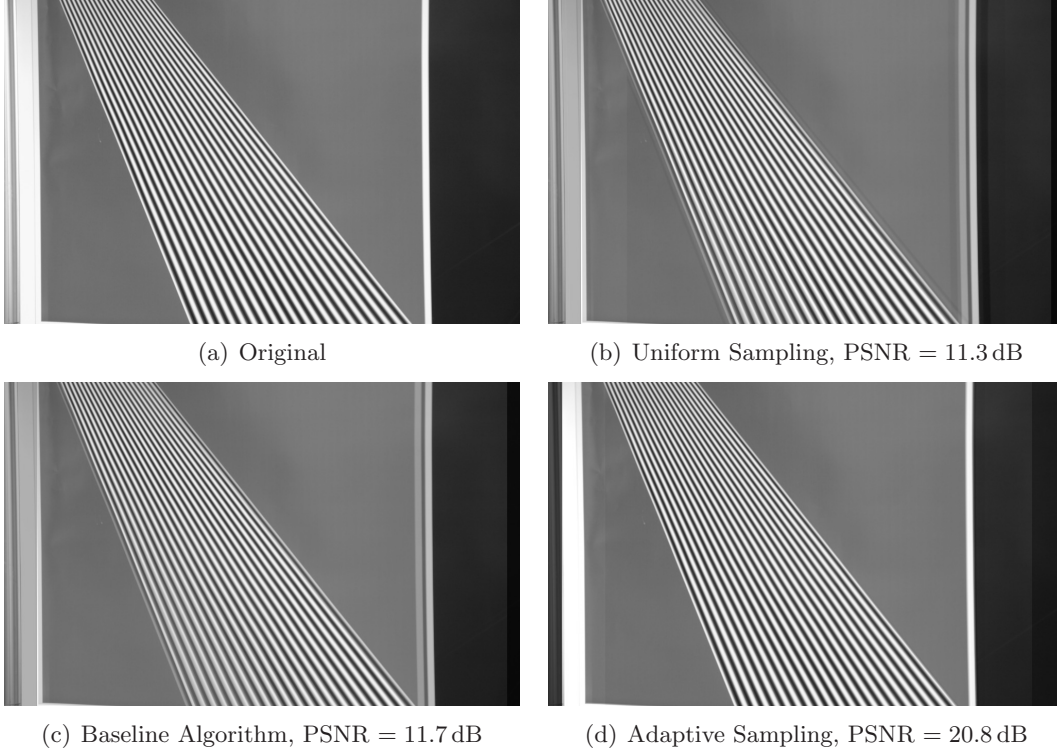


Figure F.1: Rendered image 1.

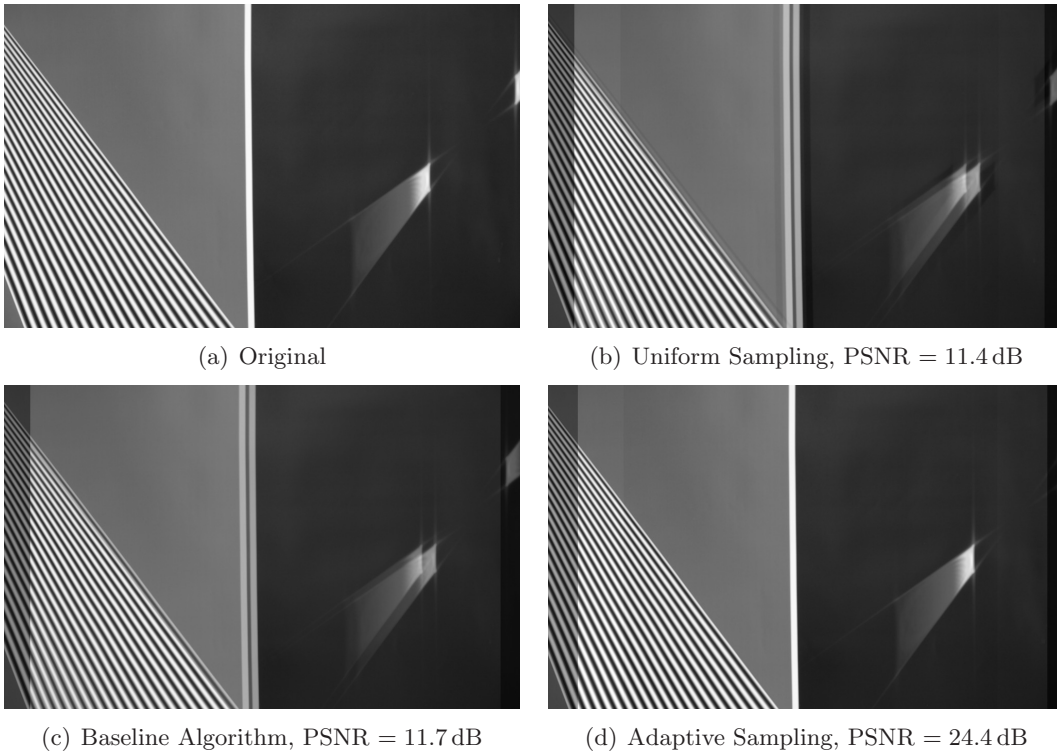


Figure F.2: Rendered image 2.

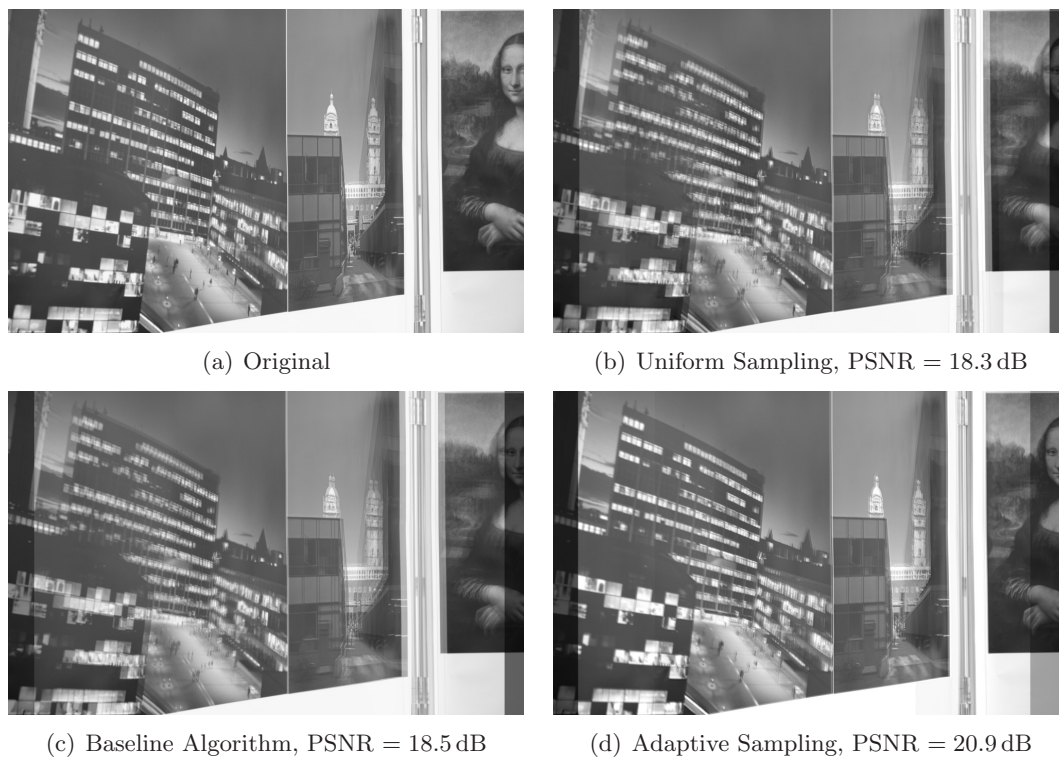


Figure F.3: Rendered image 3.

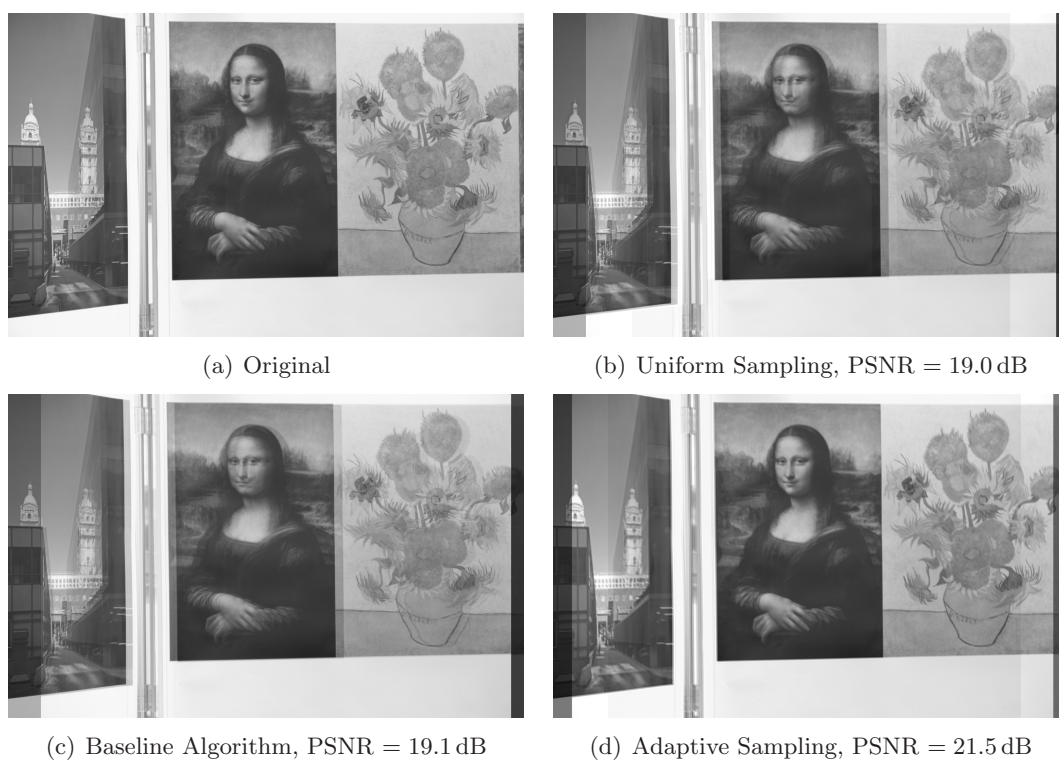


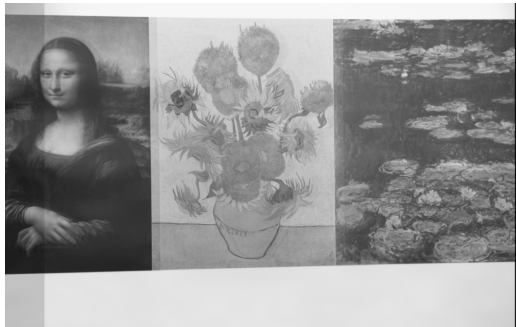
Figure F.4: Rendered image 4.



(a) Original



(b) Uniform Sampling, PSNR = 22.6 dB



(c) Baseline Algorithm, PSNR = 22.8 dB



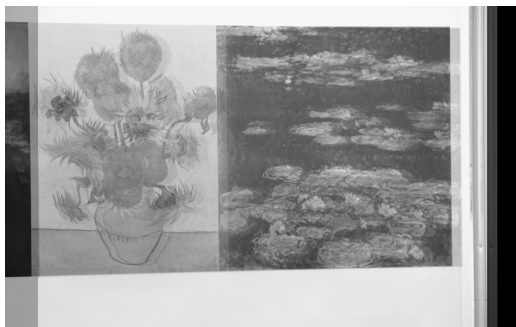
(d) Adaptive Sampling, PSNR = 25.5 dB

Figure F.5: Rendered image 5.

(a) Original



(b) Uniform Sampling, PSNR = 19.8 dB



(c) Baseline Algorithm, PSNR = 19.9 dB



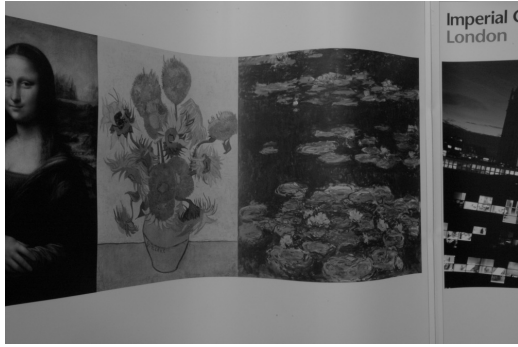
(d) Adaptive Sampling, PSNR = 25.9 dB

Figure F.6: Rendered image 6.

Appendix G

Rendered Images for the Smoothly Varying Scene

In this appendix, we present six examples of rendered images from the EPI-volume of a smoothly varying scene. The EPI-volume is described in Section 5.5.2, and the smoothly varying scene is shown in Figure 5.14. For each example, we render the target image using three different sampling and reconstruction methods: uniform sampling and reconstruction using a 3rd order I-MOMS; uniform sampling and reconstruction using the baseline algorithm; and adaptive sampling and reconstruction using our adaptive sampling algorithm. The six target images and their renderings are shown in Figures G.1(d), G.2, G.3, G.4, G.5 and G.6. Similar to Section 5.5.2, each figure shows the original (target) image in part (a), its rendering using uniform sampling with I-MOMS in (b), its rendering using the baseline algorithm in (c) and its rendering using adaptive sampling in (d). The figures also show the corresponding PSNR for each rendered image. Note that, in all of the figures, the rendered images are generated using 10 original images.



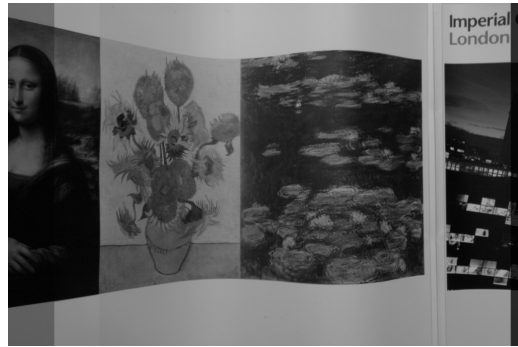
(a) Original



(b) Uniform Sampling, PSNR = 20.3 dB



(c) Baseline Algorithm, PSNR = 20.5 dB



(d) Adaptive Sampling, PSNR = 21.7 dB

Figure G.1: Rendered image 1.



(a) Original



(b) Uniform Sampling, PSNR = 21.5 dB

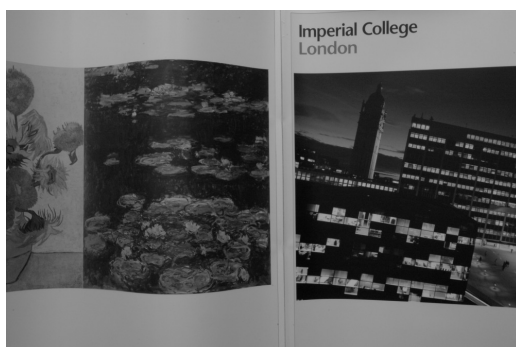


(c) Baseline Algorithm, PSNR = 21.6 dB

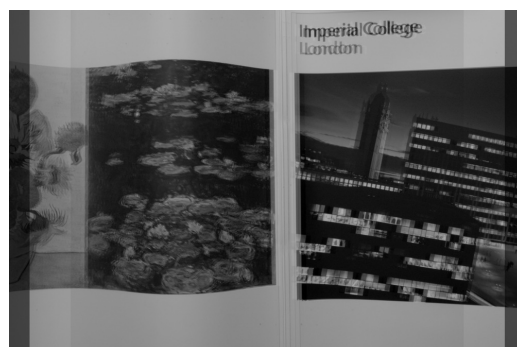


(d) Adaptive Sampling, PSNR = 25.2 dB

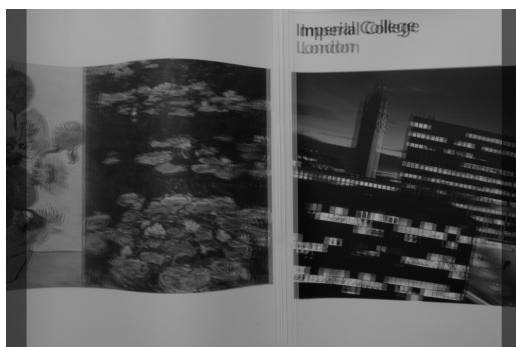
Figure G.2: Rendered image 2.



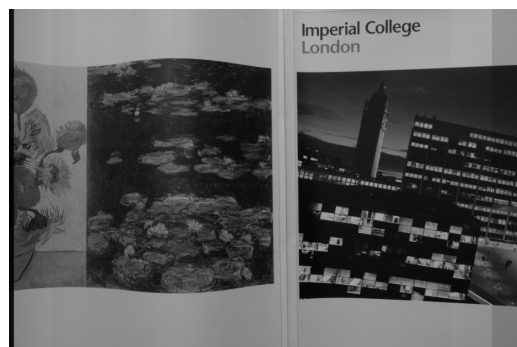
(a) Original



(b) Uniform Sampling, PSNR = 19.4 dB

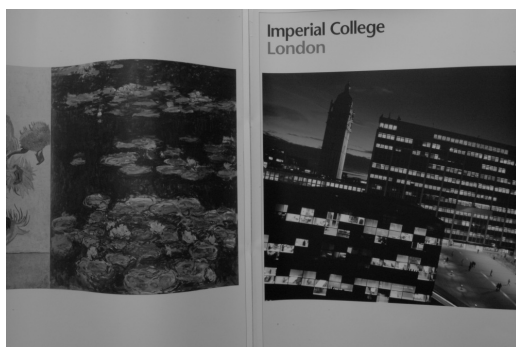


(c) Baseline Algorithm, PSNR = 19.6 dB

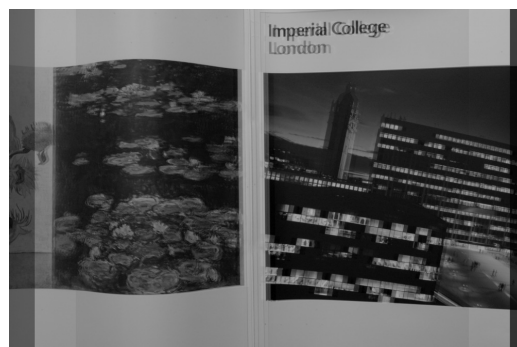


(d) Adaptive Sampling, PSNR = 25.5 dB

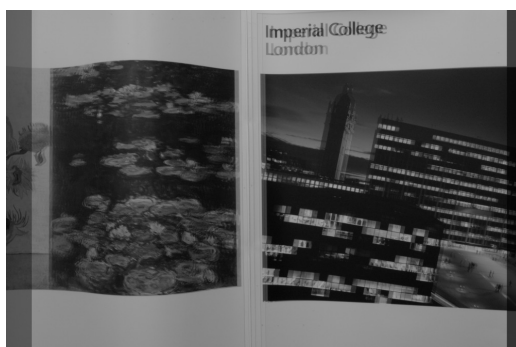
Figure G.3: Rendered image 3.



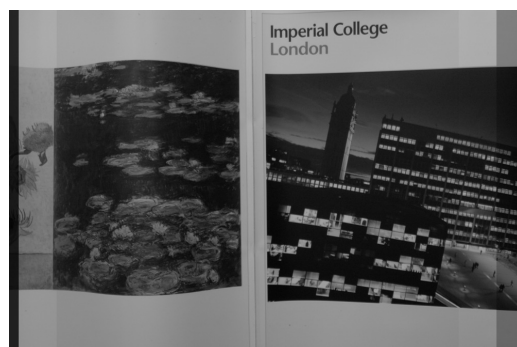
(a) Original



(b) Uniform Sampling, PSNR = 19.4 dB

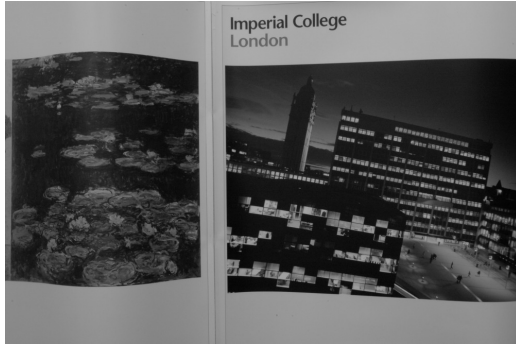


(c) Baseline Algorithm, PSNR = 19.5 dB

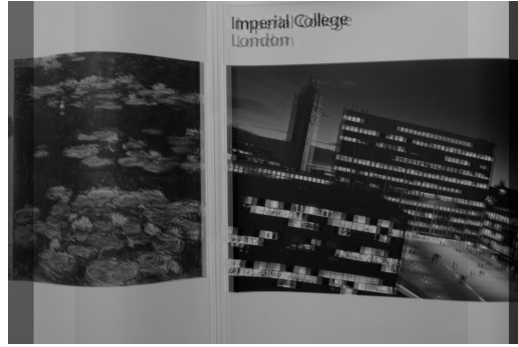


(d) Adaptive Sampling, PSNR = 22.1 dB

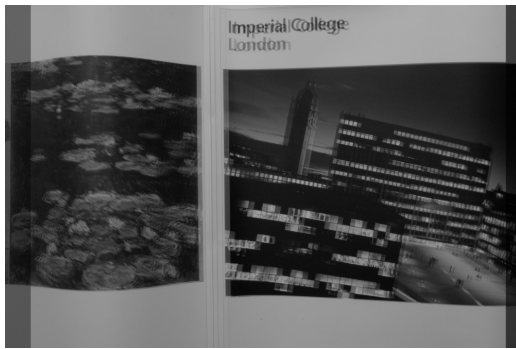
Figure G.4: Rendered image 4.



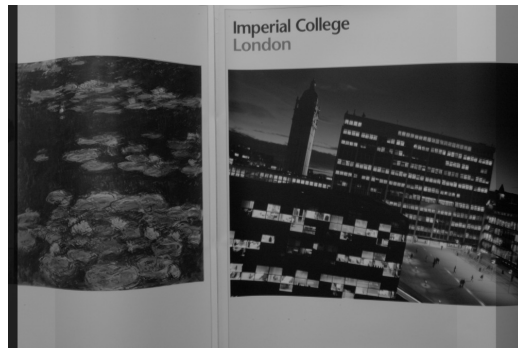
(a) Original



(b) Uniform Sampling, PSNR = 19.5 dB

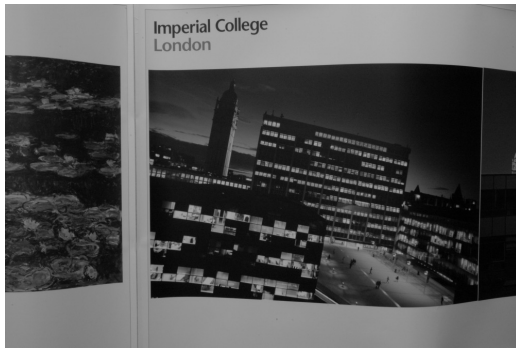


(c) Baseline Algorithm, PSNR = 19.6 dB

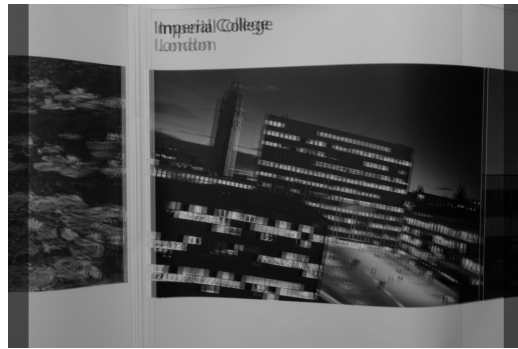


(d) Adaptive Sampling, PSNR = 21.9 dB

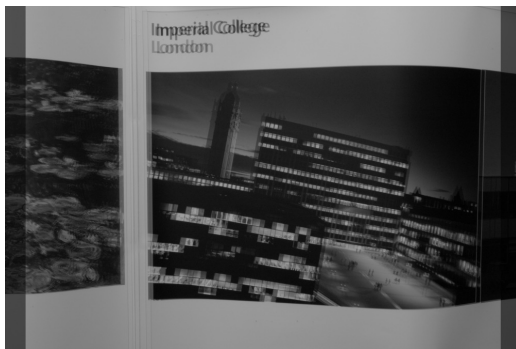
Figure G.5: Rendered image 5.



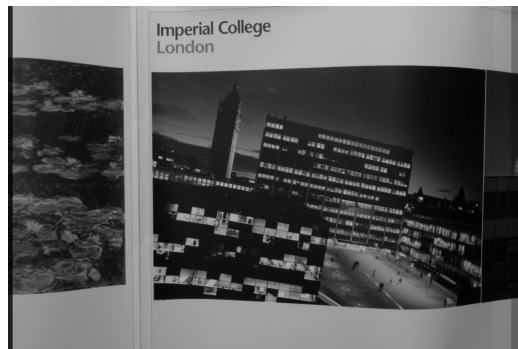
(a) Original



(b) Uniform Sampling, PSNR = 19.1 dB



(c) Baseline Algorithm, PSNR = 19.7 dB



(d) Adaptive Sampling, PSNR = 23.9 dB

Figure G.6: Rendered image 6.

Bibliography

- [1] M. Abramowitz and I. A. Stegun, Eds., *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover, 1964.
- [2] E. H. Adelson and J. R. Bergen, “The plenoptic function and the elements of early vision,” in *Computational Models of Visual Processing*. MIT Press, Cambridge, MA, 1991, pp. 3–20.
- [3] J. G. Apostolopoulos, P. A. Chou, B. Culbertson, T. Kalker, M. D. Trott, and S. Wee, “The road to immersive communication,” *Proc. IEEE*, vol. 100, no. 4, pp. 974–990, 2012.
- [4] J. Berent, “Coherent multi-dimensional segmentation of multiview images using a variational framework and applications to image based rendering,” Ph.D. dissertation, Imperial College London, 2008.
- [5] J. Berent and P. L. Dragotti, “Plenoptic manifolds,” *IEEE Signal Processing Mag.*, vol. 24, no. 6, pp. 34–44, 2007.
- [6] T. Blu, P. Thévenaz, and M. Unser, “MOMS: Maximal-order interpolation of minimal support,” *IEEE Trans. Image Processing*, vol. 10, no. 7, pp. 1069–1080, 2001.
- [7] R. Bolles, H. Baker, and D. Marimont, “Epipolar-plane image analysis: An approach to determining structure from motion,” *Int. J. Computer Vision*, vol. 1, no. 1, pp. 7–55, 1987.

- [8] Y. Bresler, “Bounds on the aliasing error in multidimensional shannon sampling,” *IEEE Trans. Information Theory*, vol. 42, no. 6, pp. 2238–2241, 1996.
- [9] C. Buehler, M. Bosse, L. McMillan, S. Gortler, and M. Cohen, “Unstructured lumigraph rendering,” in *Proc. SIGGRAPH*, Los Angeles, CA, 2001, pp. 425–432.
- [10] J. X. Chai, S. C. Chan, H. Y. Shum, and X. Tong, “Plenoptic sampling,” in *Proc. SIGGRAPH*, New Orleans, LA, 2000, pp. 307–318.
- [11] S. C. Chan, H. Y. Shum, and K. T. Ng, “Image-based rendering and synthesis,” *IEEE Signal Processing Mag.*, vol. 24, no. 6, pp. 22–33, 2007.
- [12] C. F. Chang, G. Bishop, and A. Lastra, “LDI tree: A hierarchical representation for image-based rendering,” in *Proc. SIGGRAPH*, Los Angeles, CA, 1999, pp. 291–298.
- [13] C. Chen and D. Schonfeld, “Geometrical plenoptic sampling,” in *Proc. IEEE Int. Conf. Image Processing*, Cairo, Egypt, 2009, pp. 3769–3772.
- [14] S. E. Chen and L. Williams, “View interpolation for image synthesis,” in *Proc. SIGGRAPH*, Anaheim, CA, 1993, pp. 279–288.
- [15] A. Criminisi, S. B. Kang, R. Swaminathan, R. Szeliski, and P. Anandan, “Extracting layers and analyzing their specular properties using epipolar-plane-image analysis.” *Computer Vision and Image Understanding*, vol. 97, no. 1, pp. 51–85, 2005.
- [16] F. de Sorbier, Y. Uematsu, and H. Saito, “Depth camera based system for auto-stereoscopic displays,” in *IEEE Int. Workshop Multimedia Signal Processing*, 2010, pp. 361–366.
- [17] M. N. Do, D. Marchand-Maillet, and M. Vetterli, “On the bandwidth of the plenoptic function,” *IEEE Trans. Image Processing*, vol. 21, no. 2, pp. 708–717, 2012.

- [18] M. N. Do, Q. H. Nguyen, H. T. Nguyen, D. Kubacki, and S. J. Patel, "Immersive visual communication," *IEEE Signal Processing Mag.*, vol. 28, no. 1, pp. 58–66, 2011.
- [19] C. Fehn, "Depth-image-based rendering (DIBR), compression, and transmission for a new approach on 3D-TV," *Proc. SPIE*, vol. 5291, pp. 93–104, 2004.
- [20] C. Fehn and R. S. Pastoor, "Interactive 3-DTV-concepts and key technologies," *Proc. IEEE*, vol. 94, no. 3, pp. 524–538, 2006.
- [21] I. Feldmann, P. Eisert, and P. Kauff, "Extension of epipolar image analysis to circular camera movements," in *Proc. IEEE Int. Conf. Image Processing*, Barcelona, Spain, 2003, pp. 697–700.
- [22] S. Fleishman, D. Cohen-Or, and D. Lischinski, "Automatic camera placement for image-based modeling," *Computer Graphics Forum*, vol. 19, no. 2, pp. 101–110, 2000.
- [23] S. Foix, G. Alenya, and C. Torras, "Lock-in Time-of-Flight (ToF) cameras: A survey," *IEEE Sensors J.*, vol. 11, no. 9, pp. 1917–1926, 2011.
- [24] T. Fujii, T. Kimoto, and M. Tanimoto, "Ray space coding for 3D visual communication," in *Proc. Picture Coding Symp.*, 1996, pp. 447–451.
- [25] Z. F. Gan, S. C. Chan, K. T. Ng, and H. Y. Shum, "An object-based approach to plenoptic videos," in *IEEE Int. Symp. Circuits and Systems*, 2005, pp. 3435–3438.
- [26] C. Gilliam, J. Pearson, M. Brookes, and P. L. Dragotti, "Image based rendering with depth cameras: How many are needed?" in *Proc. IEEE Int. Conf. Acoustic, Speech, and Signal Processing*, Kyoto, Japan, 2012.
- [27] S. Gortler, R. Grzeszczuk, R. Szeliski, and M. Cohen, "The lumigraph," in *Proc. SIGGRAPH*, New Orleans, LA, 1996, pp. 43–54.
- [28] I. Ihm, S. Park, and R. K. Lee, "Rendering of spherical light fields," in *Pacific Conf. Computer Graphics and Applications*, 1997, pp. 59–68.

- [29] Y. S. Kang and Y. S. Ho, "High-quality multi-view depth generation using multiple color and depth cameras," in *IEEE Int. Conf. Multimedia and Expo*, 2010, pp. 1405–1410.
- [30] A. Kirmani, A. Colaço, F. N. C. Wong, and V. K. Goyal, "Exploiting sparsity in time-of-flight range acquisition using a single time-resolved sensor," *Opt. Express*, vol. 19, no. 22, pp. 21 485–21 507, 2011.
- [31] —, "CODAC: A compressive depth acquisition camera framework," in *Proc. IEEE Int. Conf. Acoustics., Speech, and Signal Processing*, Kyoto, Japan, 2012.
- [32] A. Kolb, E. Barth, R. Koch, and R. Larsen, "Time-of-flight sensors in computer graphics," *Computer Graphics Forum*, vol. 29, no. 1, pp. 141–159, 2010.
- [33] J. Konrad and M. Halle, "3-D displays and signal processing," *IEEE Signal Processing Magazine*, vol. 24, no. 6, pp. 97 –111, 2007.
- [34] A. Kubota, A. Smolic, M. Magnor, M. Tanimoto, T. Chen, and C. Zhang, "Multi-view imaging and 3DTV," *IEEE Signal Processing Mag.*, vol. 24, no. 6, pp. 10–21, 2007.
- [35] B. P. Lathi, "Signal energy and energy spectral density," in *Modern Digital and Analog Communication Systems*. Oxford University Press, 1998, ch. 3.7, pp. 115–123.
- [36] E. K. Lee and Y. S. Ho, "Generation of high-quality depth maps using hybrid camera system for 3D video," *J. of Visual Communication and Image Representation*, vol. 22, no. 1, pp. 73–84, 2011.
- [37] M. Levoy and P. Hanrahan, "Light field rendering," in *Proc. SIGGRAPH*, New Orleans, LA, 1996, pp. 31–40.
- [38] Z. Lin and H.-Y. Shum, "A geometric analysis of light field rendering," *Int. J. Computer Vision*, vol. 58, pp. 121–138, 2004.

- [39] A. Maimone and H. Fuchs, “Encumbrance-free telepresence system with real-time 3D capture and display using commodity depth cameras,” in *IEEE Symp. Mixed and Augmented Reality*, 2011.
- [40] L. McMillan and G. Bishop, “Plenoptic modelling: An image-based rendering system,” in *Proc. SIGGRAPH*, Los Angeles, CA, 1995, pp. 39–46.
- [41] G. Miller, S. Rubin, and D. Ponceleon, “Lazy decompression of surface light fields for precomputed global illumination,” in *Proc. Rendering Techniques*, 1998, pp. 281–292.
- [42] Y. Mori, N. Fukushima, T. Yendo, T. Fujii, and M. Tanimoto, “View generation with 3D warping using depth information for FTV,” *Signal Processing: Image Communication*, vol. 24, no. 1, pp. 65–72, 2009.
- [43] R. Namboori, H. C. Teh, and Z. Huang, “An adaptive sampling method for layered depth image,” in *Proc. Computer Graphics Int.*, Crete, Greece, 2004, pp. 206–213.
- [44] H. T. Nguyen and M. N. Do, “Error analysis for image-based rendering with depth information,” *IEEE Trans. Image Process.*, vol. 18, no. 4, pp. 703–716, 2009.
- [45] J. Pearson, P.-L. Dragotti, and M. Brookes, “Accurate non-iterative depth layer extraction algorithm for image based rendering,” in *IEEE International Conference on Acoustics, Speech and Signal Processing*, Prague, Czech Republic, 2011, pp. 901–904.
- [46] H. Saito, “Computer vision for 3DTV and augmented reality,” in *Int. Symp. on Ubiquitous Virtual Reality*, 2011, pp. 5–8.
- [47] H. Schirmacher, W. Heidrich, and H. P. Seidel, “Adaptive acquisition of lumigraphs from synthetic scenes,” in *Proc. Eurographics*, vol. 18, no. 3, Milan, Italy, 1999, pp. 151–160.
- [48] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski, “A comparison and evaluation of multi-view stereo reconstruction algorithms,” in *Proc. IEEE*

- Conf. Computer Vision and Pattern Recognition*, vol. 1, New York, NY, 2006, pp. 519–528.
- [49] J. Shade, S. Gortler, L. W. He, and R. Szeliski, “Layered depth images,” in *Proc. SIGGRAPH*, New York, NY, 1998, pp. 231–242.
- [50] H. Y. Shum, S. C. Chan, and S. B. Kang, *Image-Based Rendering*. Springer, 2007.
- [51] H. Y. Shum and L. W. He, “Rendering with concentric mosaics,” in *Proc. SIGGRAPH*, Los Angeles, CA, 1999, pp. 299–306.
- [52] H. Y. Shum, S. B. Kang, and S. C. Chan, “Survey of image-based representations and compression techniques,” *IEEE Trans. Circuits and Systems for Video Technology*, vol. 13, no. 11, pp. 1020–1037, 2003.
- [53] H. Y. Shum, J. Sun, S. Yamazaki, Y. Li, and C. K. Tang, “Pop-up light field: An interactive image-based modeling and rendering system,” *ACM Trans. Graphics*, vol. 23, pp. 143–162, 2004.
- [54] S. Smirnov, A. Gotchev, and K. Egiazarian, “Methods for depth-map filtering in view-plus-depth 3D video representation,” *EURASIP J. on Advances in Signal Processing*, vol. 2012, no. 1, p. 25, 2012.
- [55] A. Smolic, K. Muller, K. Dix, P. Merkle, P. Kauff, and T. Wiegand, “Intermediate view interpolation based on multiview video plus depth for advanced 3D video systems,” in *Proc. IEEE Int. Conf. Image Processing*, San Diego, CA, 2008, pp. 2448–2451.
- [56] B. Stancil, C. Zhang, and T. Chen, “Active multicamera networks: From rendering to surveillance,” *IEEE J. Sel. Topics Signal Process.*, vol. 2, no. 4, pp. 597–605, 2008.
- [57] E. Stoykova, A. A. Alatan, P. Benzie, N. Grammalidis, S. Malassiotis, J. Ostermann, S. Piekh, V. Sainov, C. Theobalt, T. Thevar, and X. Zabulis, “3-D time-

- varying scene capture technologies - a survey,” *IEEE Trans. Circ. and Syst. for Video Tech.*, vol. 17, no. 11, pp. 1568–1586, 2007.
- [58] K. Takahashi, “Theoretical analysis of view interpolation with inaccurate depth information,” *IEEE Trans. Image Processing*, vol. 21, no. 2, pp. 718–732, 2012.
- [59] K. Takahashi, A. Kubota, and T. Naemura, “A focus measure for light field rendering,” in *Proc. IEEE Int. Conf. Image Processing*, vol. 4, 2004, pp. 2475–2478.
- [60] M. Tanimoto, M. P. Tehrani, T. Fujii, and T. Yendo, “Free-viewpoint TV,” *IEEE Signal Processing Mag.*, vol. 28, no. 1, pp. 67–76, 2011.
- [61] M. Unser, “Splines: A perfect fit for signal and image processing,” *IEEE Signal Processing Mag.*, vol. 16, no. 6, pp. 22–38, 1999.
- [62] B. Wilburn, N. Joshi, V. Vaish, E. V. Talvala, E. Antunez, A. Barth, A. Adams, M. Horowitz, and M. Levoy, “High performance imaging using large camera arrays,” in *Proc. SIGGRAPH*, Los Angeles, CA, 2005, pp. 765–776.
- [63] B. Wilburn, M. Smulski, K. Lee, and M. Horowitz, “The light field video camera,” in *Proc. Media Processors*, 2002, pp. 29–36.
- [64] A. D. Wilson and H. Benko, “Combining multiple depth cameras and projectors for interactions on, above and between surfaces,” in *Proc. ACM symp. User interface software and technology*, 2010, pp. 273–282.
- [65] D. Wood, D. Azuma, K. Aldinger, B. Curless, T. Duchamp, D. Salesin, and W. Stuetzle, “Surface light fields for 3D photography,” in *Proc. SIGGRAPH*, New Orleans, LA, 2000, pp. 287–296.
- [66] S. Würmlin, E. Lamboray, and M. Gross, “3D video fragments: Dynamic point samples for real-time free-viewpoint video,” *Computer Graphics*, vol. 28, no. 1, pp. 3–14, 2004.

- [67] C. Zhang and T. Chen, “Non-uniform sampling of image-based rendering data with the position-interval error (PIE) function,” in *Visual Communications and Image Processing*, 2003, pp. 1347–1358.
- [68] —, “Spectral analysis for sampling image-based rendering data,” *IEEE Trans. Circuits and Systems for Video Technology*, vol. 13, no. 11, pp. 1038–1050, 2003.
- [69] —, “A survey on image-based rendering - representation, sampling and compression,” *Signal Processing: Image Communication*, vol. 19, no. 1, pp. 1–28, 2004.
- [70] —, *Light Field Sampling*, ser. Synthesis Lectures on Image, Video, and Multimedia Processing. Morgan & Claypool Publishers, 2006.
- [71] —, “Active rearranged capturing of image-based rendering scenes - theory and practice,” *IEEE Trans. Multimedia*, vol. 9, no. 3, pp. 520–531, 2007.
- [72] J. Zhu, L. Wang, R. Yang, and J. Davis, “Fusion of time-of-flight depth and stereo for high accuracy depth maps,” in *IEEE Conf. Computer Vision and Pattern Recognition*, 2008, pp. 1–8.
- [73] S. Zinger, L. Do, and P. H. N. de With, “Free-viewpoint depth image based rendering,” *J. Visual Communication and Image Representation*, vol. 21, pp. 533–541, 2010.
- [74] C. L. Zitnick, S. B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski, “High-quality video view interpolation using a layered representation,” in *Proc. SIGGRAPH*, Los Angeles, CA, 2004, pp. 600–608.