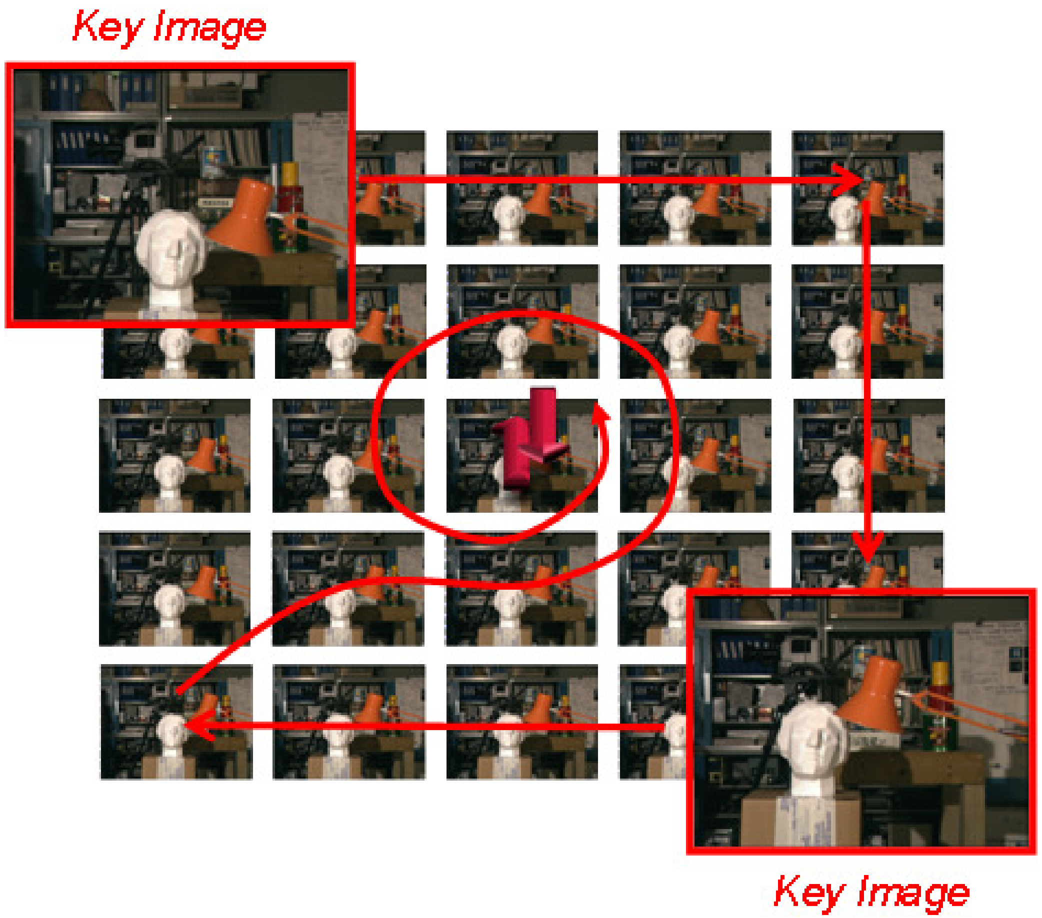
Image Based Rendering with Depth Cameras: How Many are Needed?

Christopher Gilliam, James Pearson, Mike Brookes and Pier Luigi Dragotti
Communications & Signal Processing Group
Imperial College London

30th March 2012

Motivation

Image Based Rendering (IBR) $\Longrightarrow$ Rendering new viewpoints of a scene from a multi-view image set



Courtesy of James Pearson [1]

$\looparrowright$ Trade-off between geometric information and image density

Motivation

Obtaining geometric information:

- Passive Depth Sensing

  $\Longrightarrow$ Large computational complexity and inconsistent

- Active Depth Sensing

  $\Longrightarrow$ Low-cost, fast and reliable depth cameras

Microsoft's Kinect

Mesa Imaging's SwissRanger^{TM}

Motivation

Obtaining geometric information:

- Passive Depth Sensing

  $\Longrightarrow$ Large computational complexity and inconsistent

- Active Depth Sensing

  $\Longrightarrow$ Low-cost, fast and reliable depth cameras

Mesa Imaging's SwissRanger^{TM}

Microsoft's Kinect

Therefore combine RGB and Depth:

$\looparrowright$ Multi-View Color and Depth Camera System

Motivation

Obtaining geometric information:

- Passive Depth Sensing

  $\Longrightarrow$ Large computational complexity and inconsistent

- Active Depth Sensing

  $\Longrightarrow$ Low-cost, fast and reliable depth cameras

Mesa Imaging's SwissRanger$^{\text{TM}}$

Microsoft's Kinect

Therefore combine RGB and Depth:

$\looparrowright$ Multi-View Color and Depth Camera System

Density of color cameras examined using the concept of the plenoptic function[2,3,4]

Motivation

Obtaining geometric information:

- Passive Depth Sensing

  $\Longrightarrow$ Large computational complexity and inconsistent

- Active Depth Sensing

  $\Longrightarrow$ Low-cost, fast and reliable depth cameras

Mesa Imaging's SwissRanger^{TM}

Microsoft's Kinect

Therefore combine RGB and Depth:

$\looparrowright$ Multi-View Color and Depth Camera System

$\looparrowright$ How many depth cameras are required?

$\looparrowright$ Interplay between: Number of Color Cameras, Number of Depth Cameras and Scene Geometry

Overview

- The Plenoptic Function and Plenoptic Sampling
  - Current Theory
  - Spectral Analysis of a Slanted Plane
  - Adaptive Sampling Algorithm
- Multi-view Depth Images
  - Concept $\longrightarrow$ Plenoptic style framework
  - Spectral Analysis for Multi-view Depth Images
  - Results for synthetic and real scenes
- Conclusions

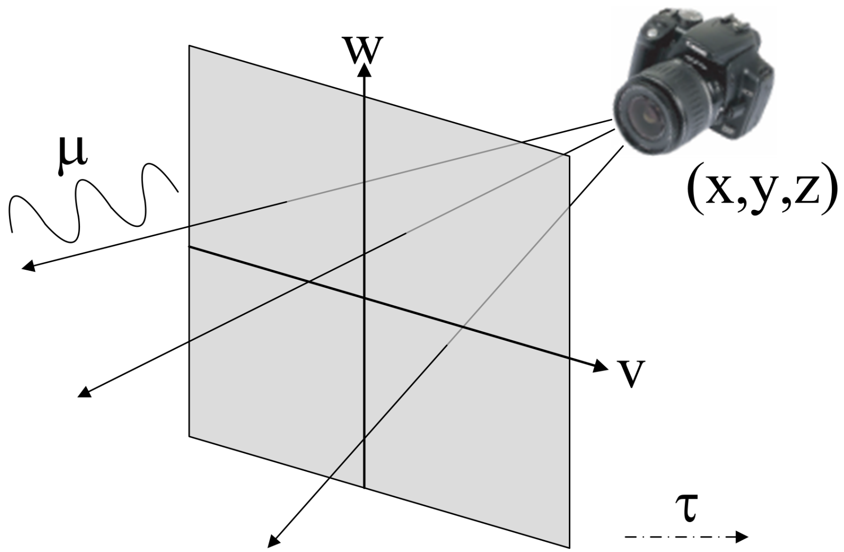
The Plenoptic Function

IBR in more detail:

- Images sample a set of lights rays from the scene to the camera
- New rendering interpolated from captured light rays
- Lights ray modelled using the 7D *Plenoptic Function* [2]

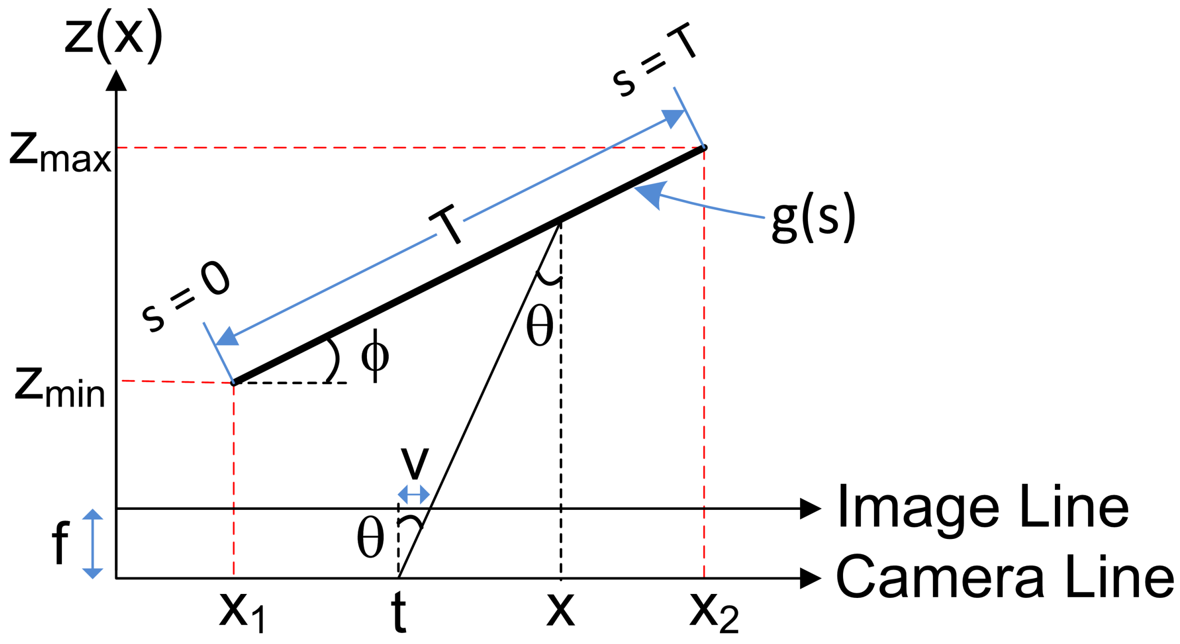
↬ IBR viewed as the Sampling and Reconstruction of the Plenoptic Function

↬ Camera density from spectral analysis of Plenoptic Function[3]



- Camera centre location (x, y, z),
- Pixel Coordinate (v, w),
- Wavelength μ,
- Time τ.

Slanted Plane Geometry



Functional Scene Model [4]:

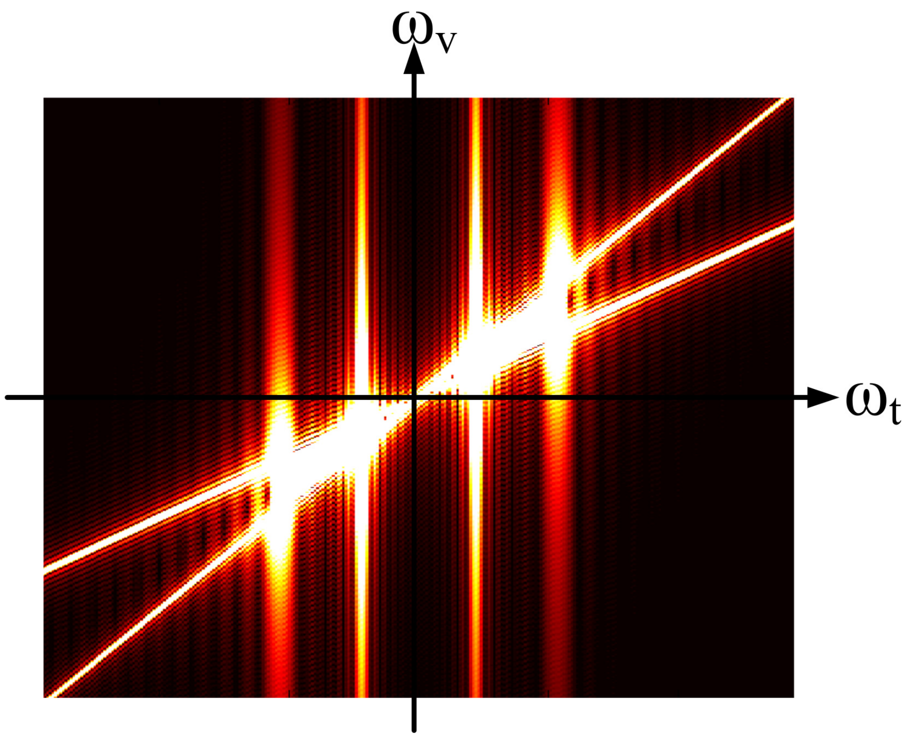
- s is the Curvilinear Coordinate
- x is the Projection of s onto t
- ϕ is the Slant of the Plane
- θ is the viewing angle
- f is the focal length

Bandlimited Texture Signal Pasted to Scene Surface, $g(s)$

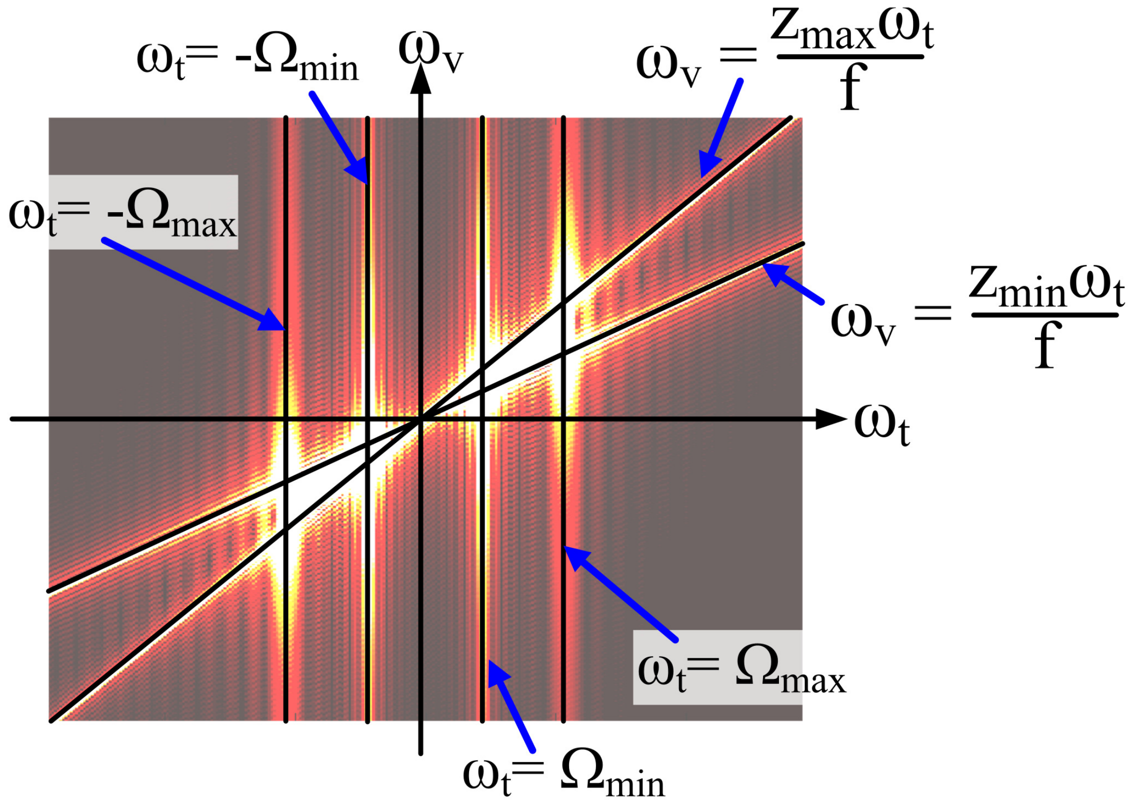
Constraints:

- 2D Plenoptic Function $\implies$ $p(t, v)$
- Finite Field of View (FoV) for the Cameras $\implies v \in [-v_m, v_m]$
- Finite Plane Width $\implies s \in [0, T]$

Slanted Plane Geometry



(a) Example Spectrum



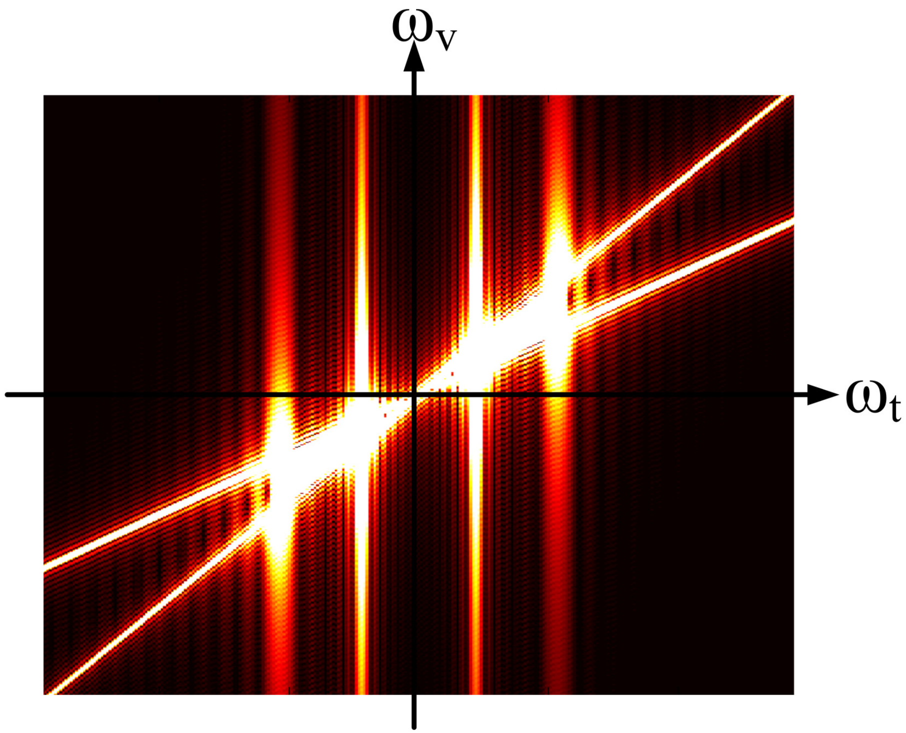
(b) Characteristics

Spectral Analysis:

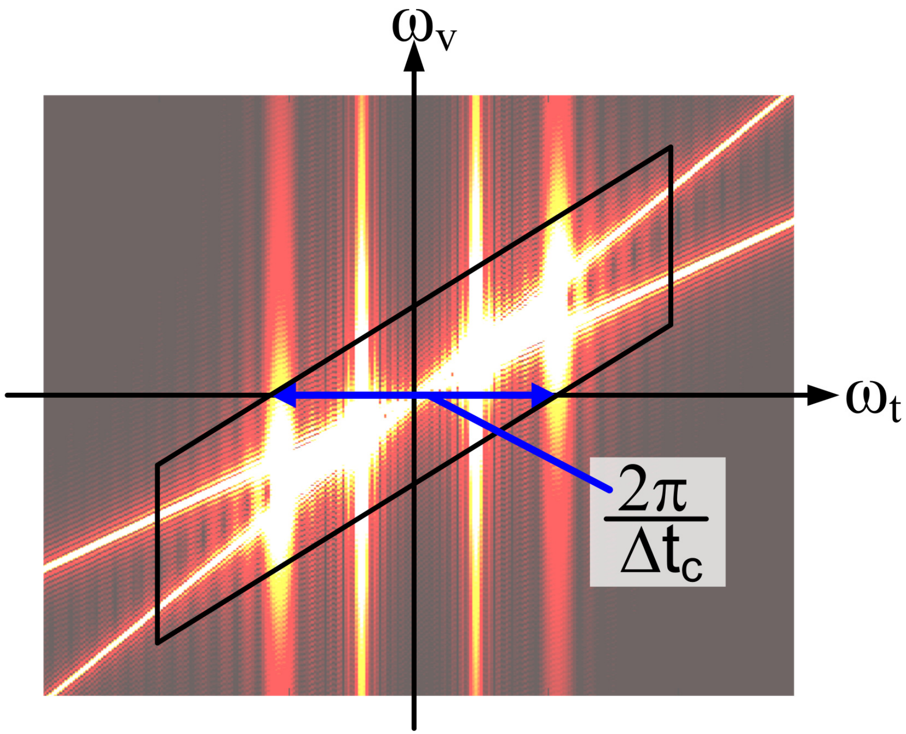
Close-form expression for the plenoptic spectrum of a slanted plane[5]

↬ Characterised by lines relating to the depth and warped texture signal

Slanted Plane Geometry



(a) Example Spectrum



(b) Its Essential Bandwidth

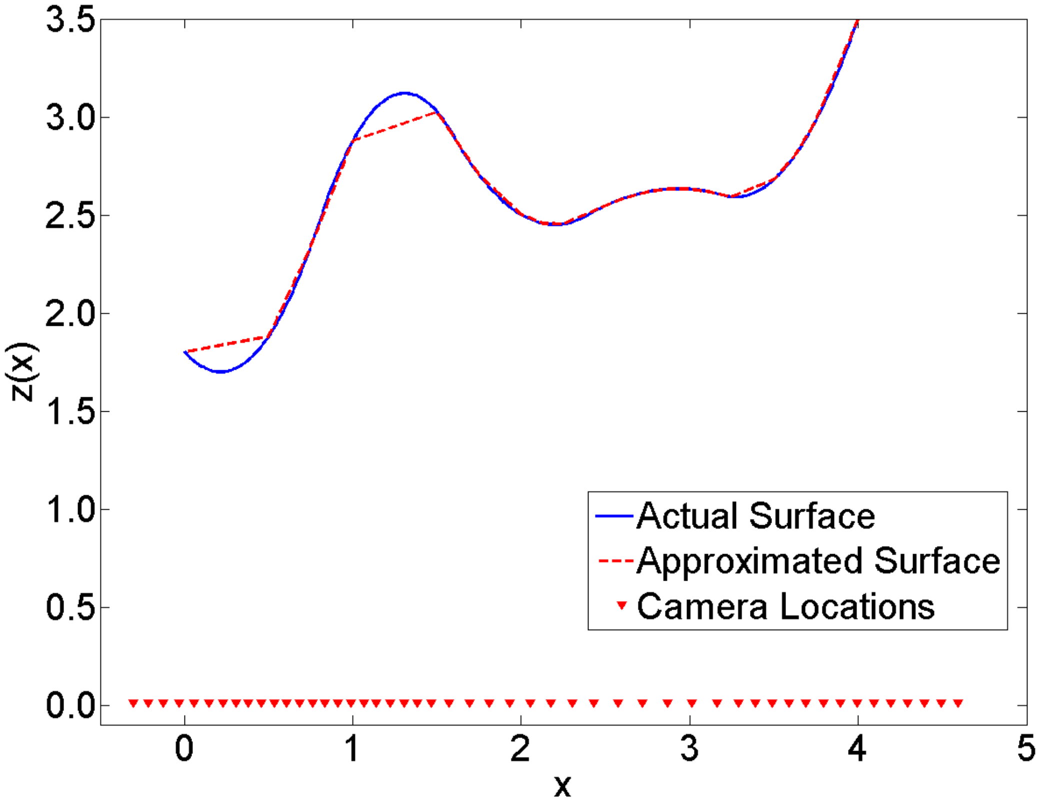
Spectral Analysis:

Close-form expression for the plenoptic spectrum of a slanted plane[5]

Spectrum's support $\implies$ Maximum camera spacing, Δt_c

Sampling Realistic Scenes

Smoothly varying scene surface with bandlimited texture:

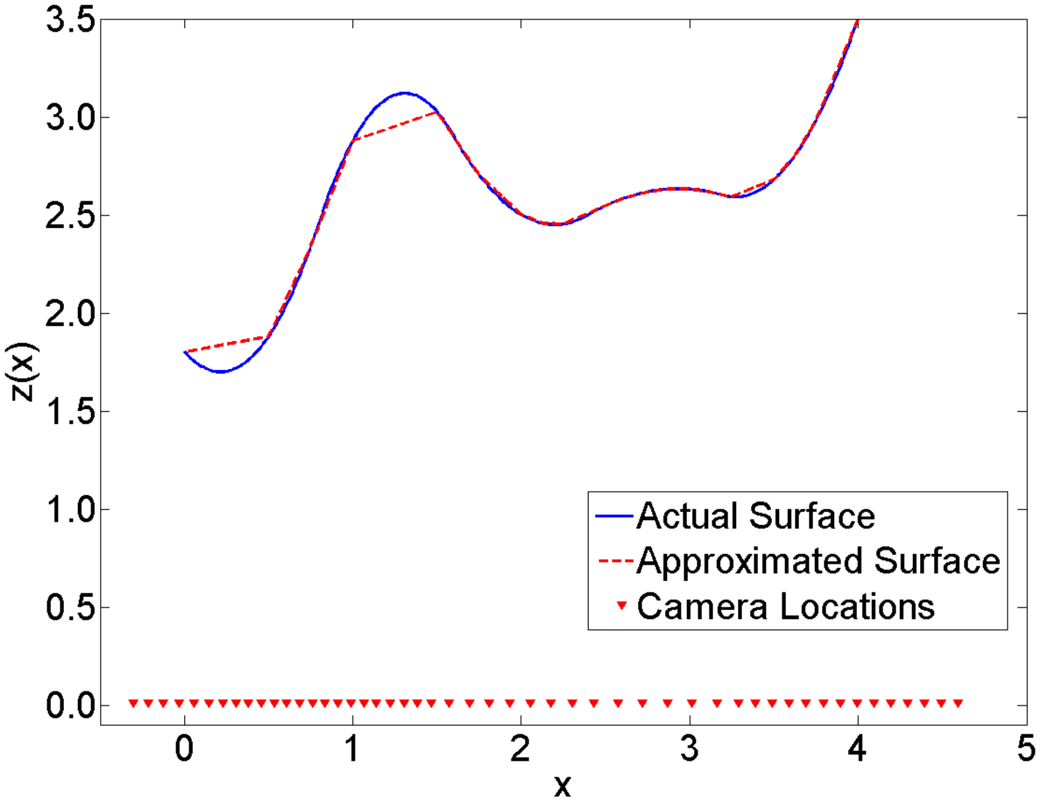


Adaptive Sampling Algorithm[6]:

- Slanted plane as a basis element $\Longrightarrow$ Construct more complicated scenes.
- Approximate the scene surface with a set of slanted planes.
- Non-uniformly sample the plenoptic function based on the approximation.

Sampling Realistic Scenes

Smoothly varying scene surface with bandlimited texture:



Limitation:

Approximation of the surface requires knowledge of the full scene geometry

$\looparrowright$ Need to determine the scene geometry...

Multi-View Depth Images

Use a similar concept to the 2D plenoptic function:

$\looparrowright$ Treat multi-view depth images as samples of a 2D function $q(t, v)$

Multi-View Depth Images

Use a similar concept to the 2D plenoptic function:

$\looparrowright$ Treat multi-view depth images as samples of a 2D function $q(t, v)$

$$q(t, v) \implies \text{inverse depth of the scene at camera location } t \text{ and pixel position } v$$

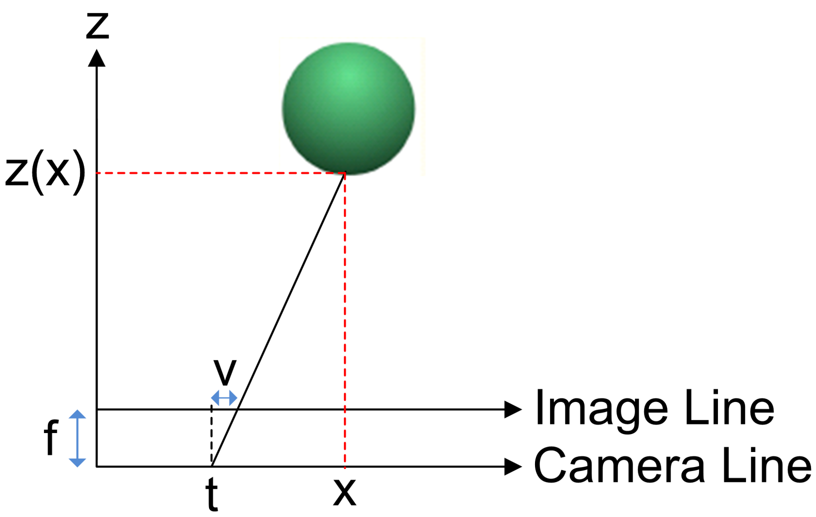
Multi-View Depth Images

Use a similar concept to the 2D plenoptic function:

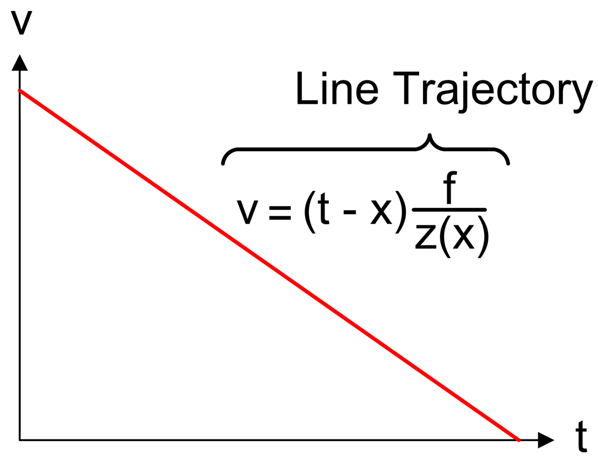
↬ Treat multi-view depth images as samples of a 2D function $q(t, v)$

$$q(t, v) \implies \text{inverse depth of the scene at camera location } t \text{ and pixel position } v$$

Similar to the Epipolar Plane Depth Image (EPDI)[7]:



(a) Scene



(b) $q(t, v)$

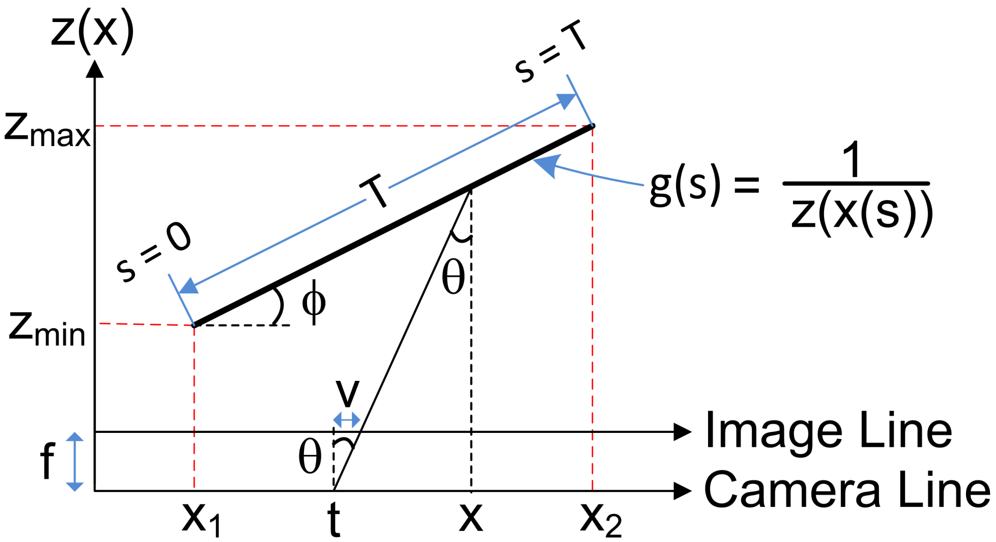
↬ Point in the scene, $(x, z(x)) \implies$ Line in $q(t, v)$

↬ Intensity along the line = $1/z(x)$

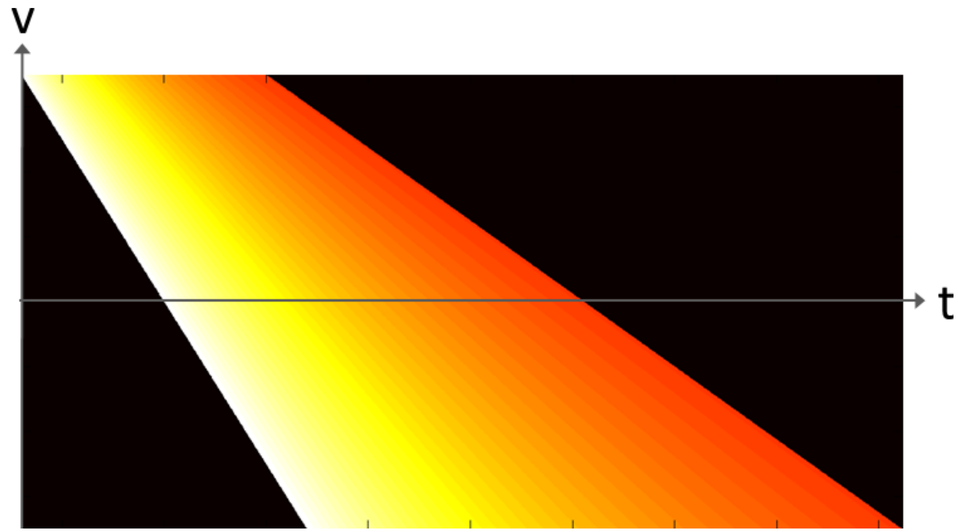
The Pantelic Function for a Slanted Plane

We term $q(t, v)$ the Pantelic Function:

Pantelic $\implies$ A slight abuse of the Greek $\pi\alpha\nu$ meaning *all* and $\tau\eta\lambda\epsilon$ meaning *at distance*



(a) The Scene



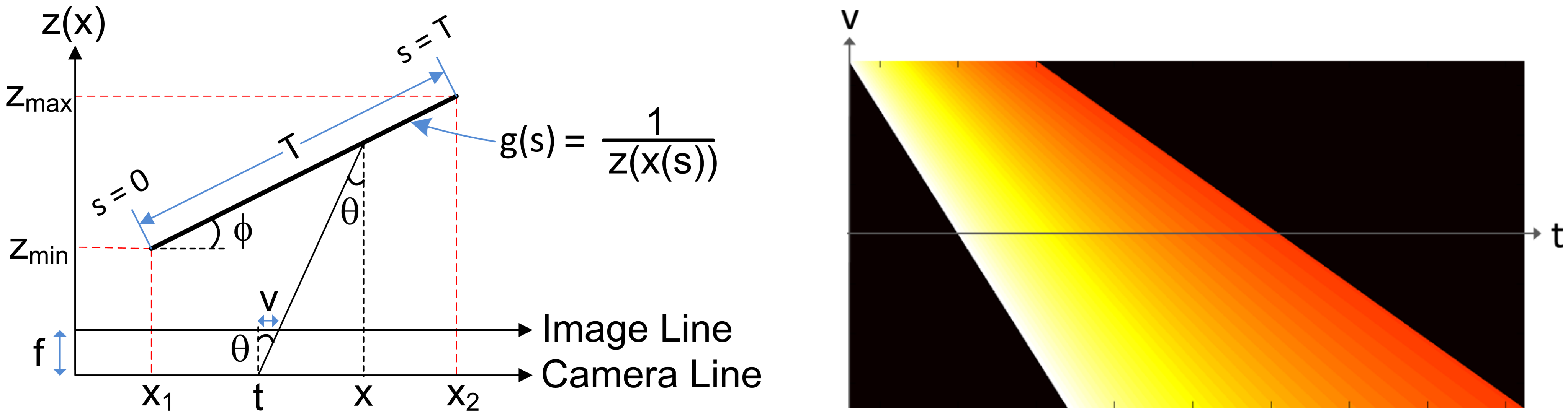
(b) Its Pantelic Function

$\looparrowright$ $g(s)$ is a false texture equivalent to the inverse depth of the scene

The Pantelic Function for a Slanted Plane

We term $q(t, v)$ the Pantelic Function:

Pantelic $\Longrightarrow$ A slight abuse of the Greek $\pi\alpha\nu$ meaning *all* and $\tau\eta\lambda\epsilon$ meaning *at distance*



(a) The Scene

(b) Its Pantelic Function

Therefore like the plenoptic function:

↬ Determine number of depth cameras needed through spectral analysis

Evaluating the Spectrum for a Slanted Plane

Fourier Transform of Pantelic Function:

$$\int_{t=-\infty}^{t=\infty} \int_{v=-\infty}^{v=\infty} q(t, v)\, e^{-j(\omega_t t + \omega_v v)}\, dv dt$$

Evaluating the Spectrum for a Slanted Plane

Fourier Transform of Pantelic Function:

$$\int_{t=-\infty}^{t=\infty} \int_{v=-\infty}^{v=\infty} q(t, v)\, e^{-j(\omega_t t + \omega_v v)}\, dv dt$$

Changing from t to x:

$$\implies \int_{x=-\infty}^{x=\infty} \int_{v=-\infty}^{v=\infty} \hat{q}(x) \left[1 - v\frac{z'(x)}{f}\right] e^{-j(\omega_v - z(x)\frac{\omega_t}{f})v} e^{-j\omega_t x}\, dv dx$$

Evaluating the Spectrum for a Slanted Plane

Fourier Transform of Pantelic Function:

$$\int_{t=-\infty}^{t=\infty} \int_{v=-\infty}^{v=\infty} q(t, v)\, e^{-j(\omega_t t + \omega_v v)}\, dv dt$$

Changing from t to x:

$$\Longrightarrow \int_{x=-\infty}^{x=\infty} \int_{v=-\infty}^{v=\infty} \hat{q}(x) \left[1 - v\frac{z'(x)}{f}\right] e^{-j(\omega_v - z(x)\frac{\omega_t}{f})v} e^{-j\omega_t x}\, dv dx$$

Finite FoV Constraint:

$$\Longrightarrow \int_{x=-\infty}^{x=\infty} \hat{q}(x) \int_{v=-v_m}^{v=v_m} \left[1 - v\frac{z'(x)}{f}\right] e^{-j(\omega_v - z(x)\frac{\omega_t}{f})v} e^{-j\omega_t x}\, dv dx$$

Evaluating the Spectrum for a Slanted Plane

Fourier Transform of Pantelic Function:

$$\int_{t=-\infty}^{t=\infty} \int_{v=-\infty}^{v=\infty} q(t,v)\, e^{-j(\omega_t t + \omega_v v)}\, dv dt$$

Changing from t to x:

$$\Longrightarrow \quad \int_{x=-\infty}^{x=\infty} \int_{v=-\infty}^{v=\infty} \hat{q}(x) \left[1 - v \frac{z'(x)}{f} \right] e^{-j(\omega_v - z(x)\frac{\omega_t}{f})v} e^{-j\omega_t x}\, dv dx$$

Finite FoV Constraint:

$$\Longrightarrow \quad \int_{x=-\infty}^{x=\infty} \hat{q}(x) \int_{v=-v_m}^{v=v_m} \left[1 - v \frac{z'(x)}{f} \right] e^{-j(\omega_v - z(x)\frac{\omega_t}{f})v} e^{-j\omega_t x}\, dv dx$$

Finite Plane Width Constraint:

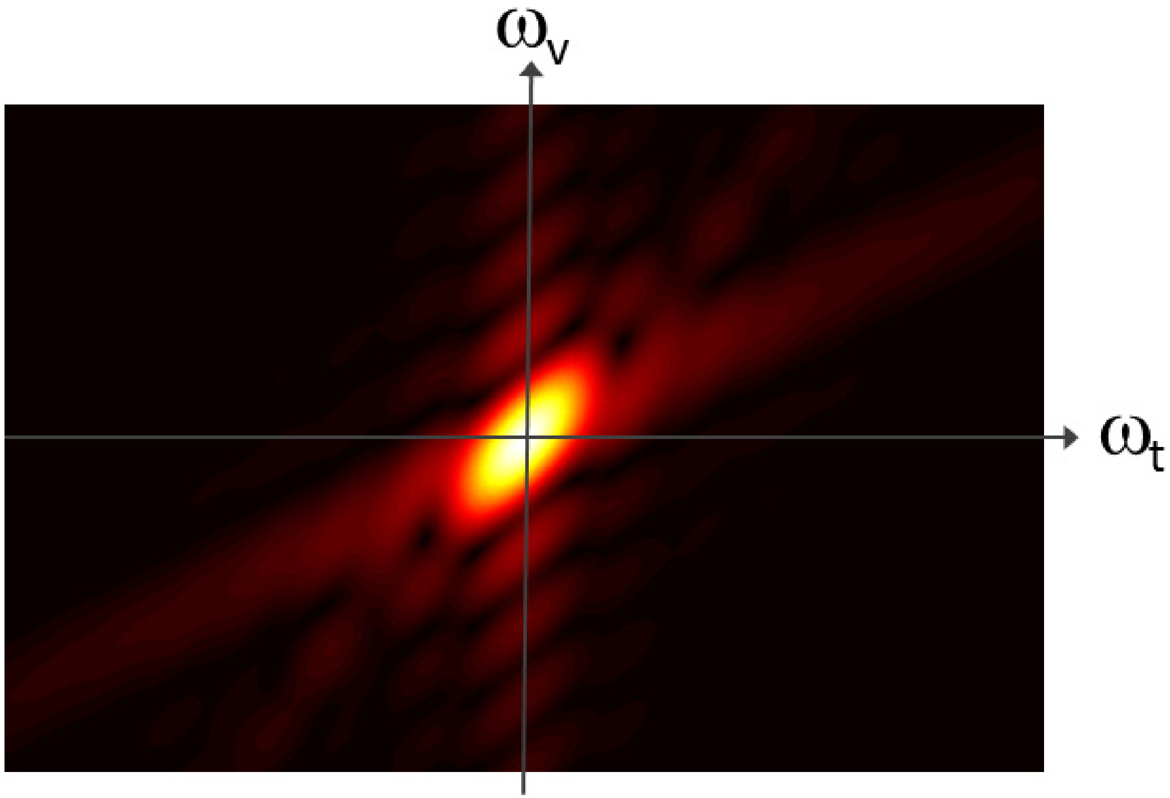
$$\Longrightarrow \quad \int_{s=0}^{s=T} g(s) \cos(\phi) \int_{v=-v_m}^{v=v_m} \left[1 - v \frac{\tan(\phi)}{f} \right] e^{-j(\omega_v - s\frac{\sin(\phi)\,\omega_t}{f})v} e^{-j\omega_t \cos(\phi) s}\, dv ds$$

The Result...

The Spectrum

$$Q_S = j\,2v_m \left(\frac{\text{sinc}(a)}{z_{max}\omega_t} e^{-j(a-b)c} - \frac{\text{sinc}(b)}{z_{min}\omega_t} \right) - j\frac{2v_m^2}{f} e^{jbc} \int_b^a \frac{\text{sinc}(\Omega)}{(\omega_v v_m - \Omega)^2} e^{-j\Omega c}\, d\Omega$$

Where $a = \omega_v v_m - \omega_t \frac{z_{max} v_m}{f}$, $b = \omega_v v_m - \omega_t \frac{z_{min} v_m}{f}$, $c = \frac{-f}{\tan(\phi) v_m}$



Structure:

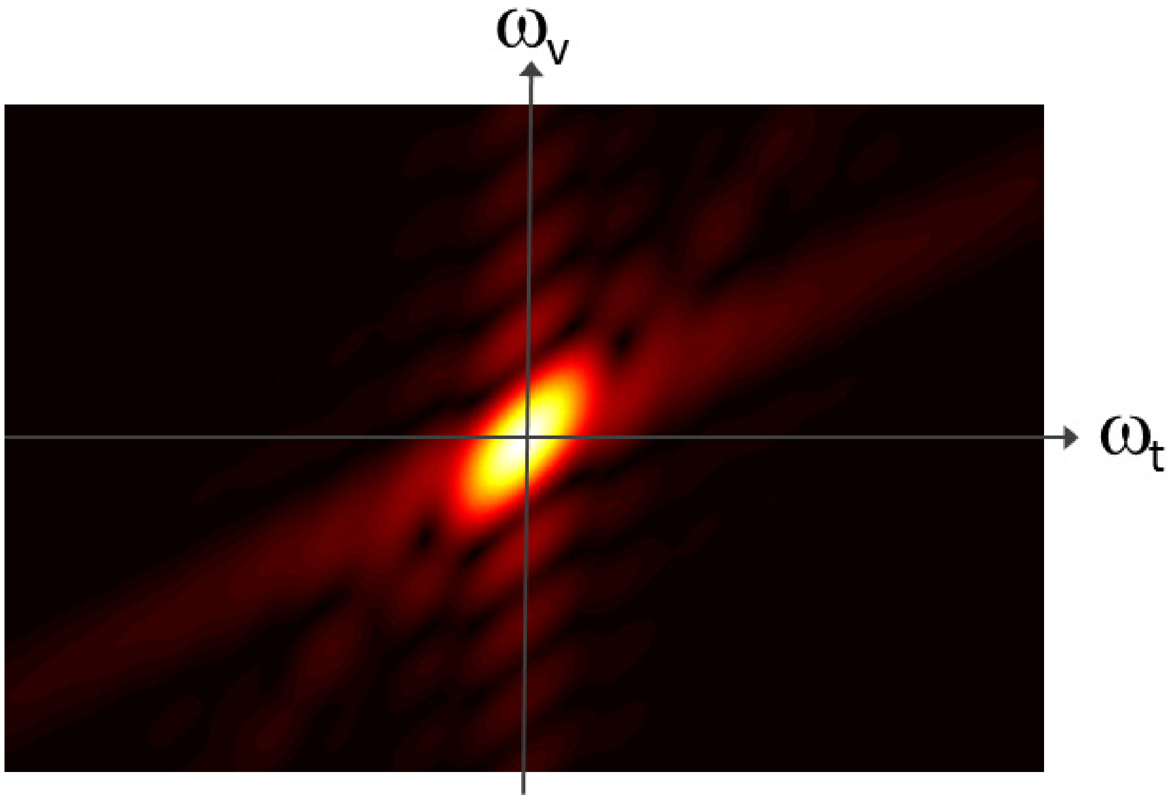
- Lines relating to z_{min} and z_{max}
- Spectrum concentrated around the origin

The Result...

The Spectrum

$$Q_S = j\,2v_m \left(\frac{\text{sinc}(a)}{z_{max}\omega_t} e^{-j(a-b)c} - \frac{\text{sinc}(b)}{z_{min}\omega_t} \right) - j\frac{2v_m^2}{f} e^{jbc} \int_b^a \frac{\text{sinc}(\Omega)}{(\omega_v v_m - \Omega)^2} e^{-j\Omega c}\, d\Omega$$

Where $\boxed{a = \omega_v v_m - \omega_t \frac{z_{max} v_m}{f}}$, $b = \omega_v v_m - \omega_t \frac{z_{min} v_m}{f}$, $c = \frac{-f}{\tan(\phi) v_m}$



Structure:

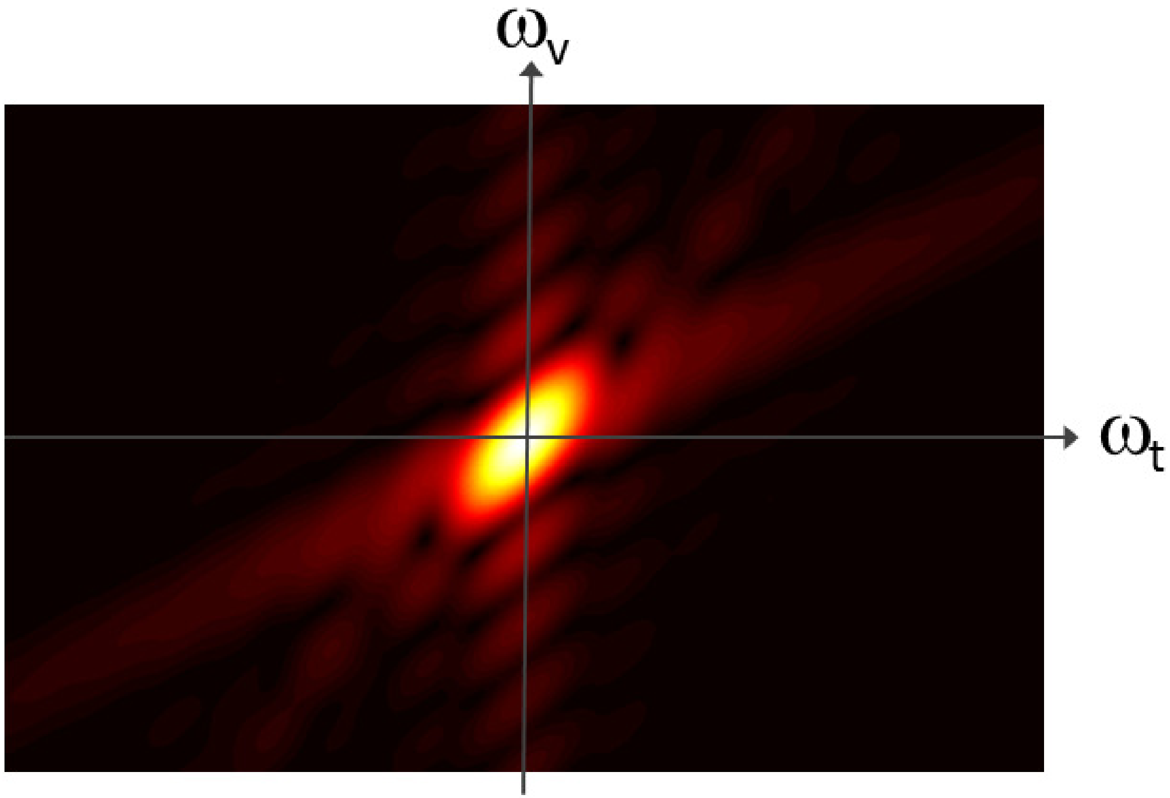
- Lines relating to z_{min} and z_{max}
- Spectrum concentrated around the origin

The Result...

The Spectrum

$$Q_S = j\,2v_m\left(\frac{\text{sinc}(a)}{z_{max}\omega_t}e^{-j(a-b)c} - \frac{\text{sinc}(b)}{z_{min}\omega_t}\right) - j\frac{2v_m^2}{f}e^{jbc}\int_b^a \frac{\text{sinc}(\Omega)}{(\omega_v v_m - \Omega)^2}e^{-j\Omega c}\,d\Omega$$

Where $\boxed{a = \omega_v v_m - \omega_t \frac{z_{max}v_m}{f}}$, $\boxed{b = \omega_v v_m - \omega_t \frac{z_{min}v_m}{f}}$, $c = \frac{-f}{\tan(\phi)v_m}$

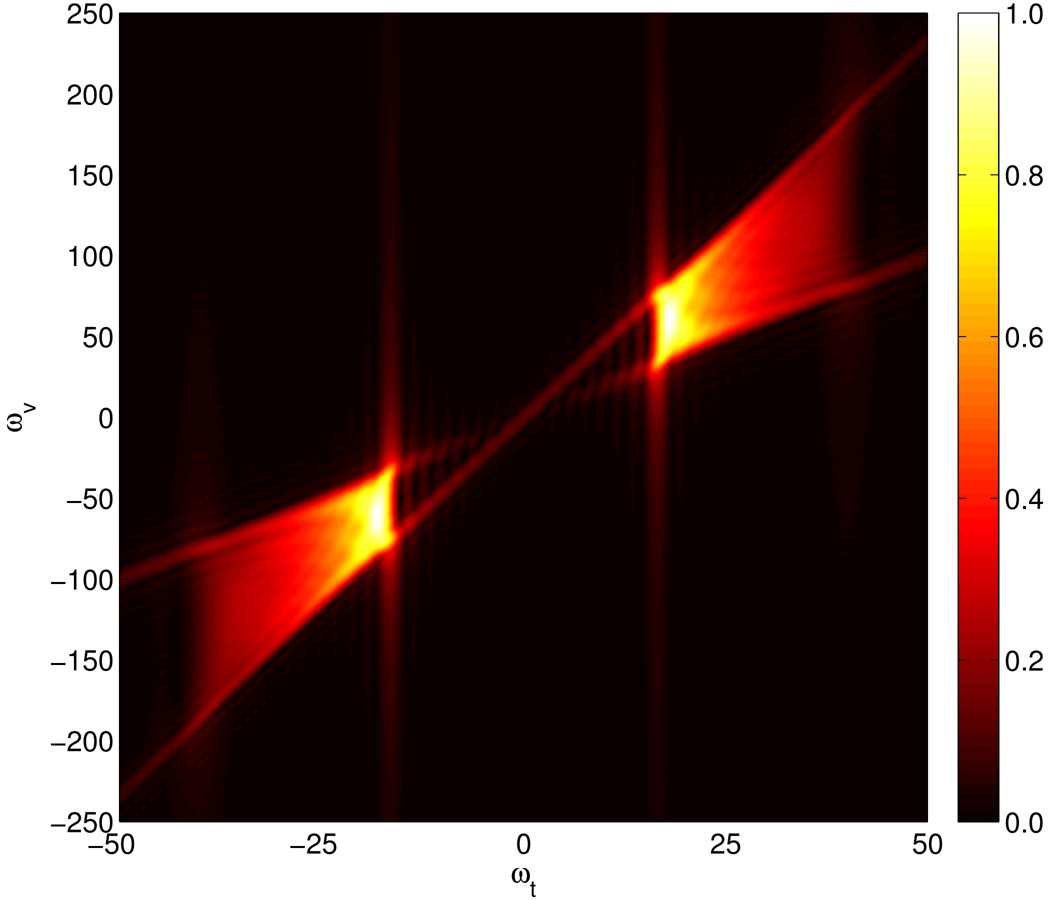


Structure:

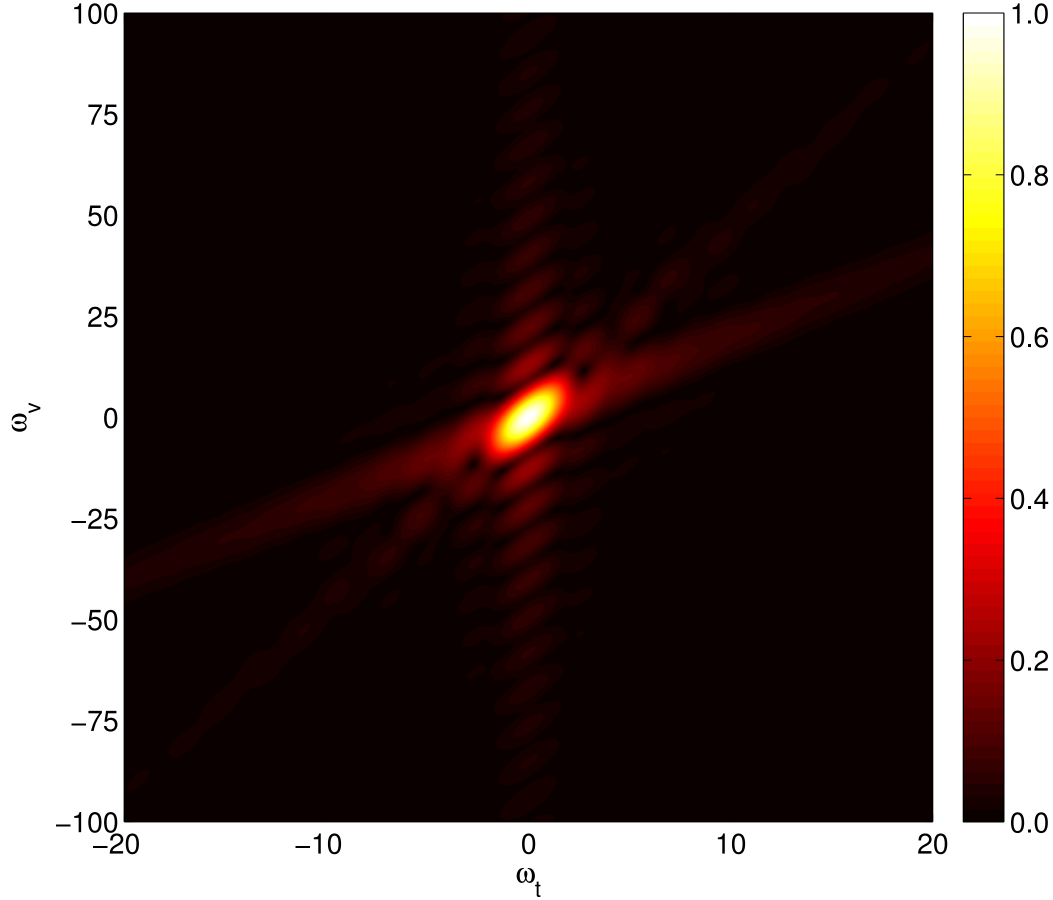
- Lines relating to z_{min} and z_{max}
- Spectrum concentrated around the origin

The Result...

Comparison between the Plenoptic and Pantelic Spectrum:

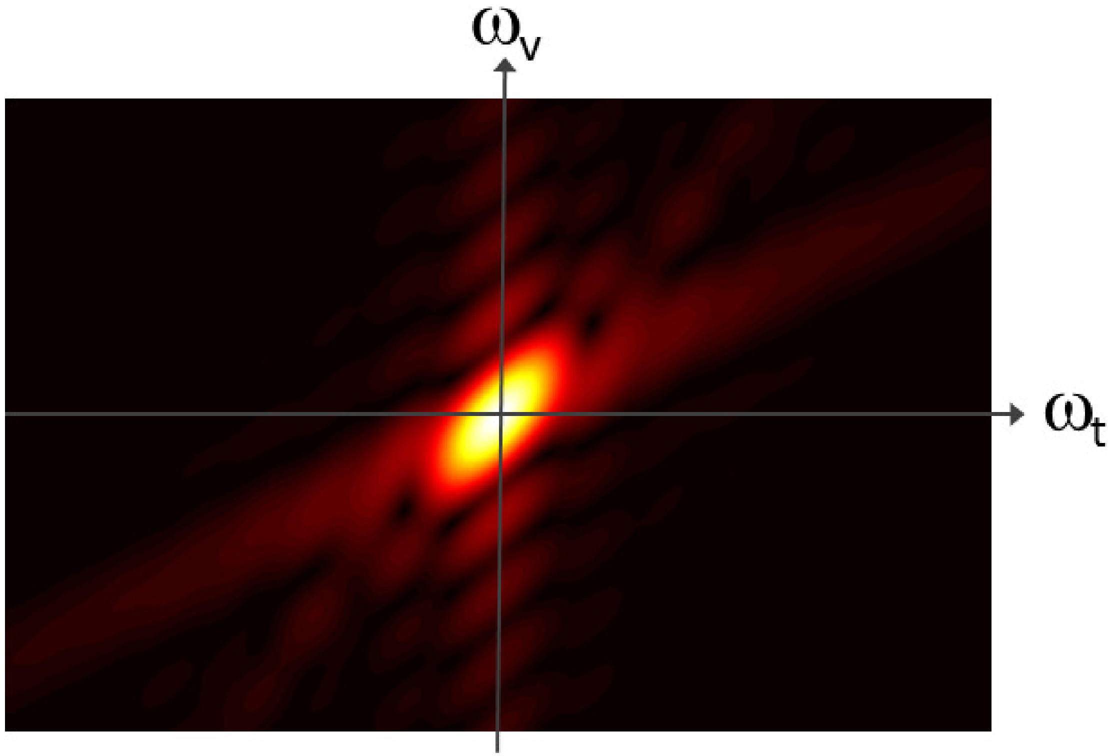


(a) Plenoptic Spectrum



(b) Pantelic Spectrum

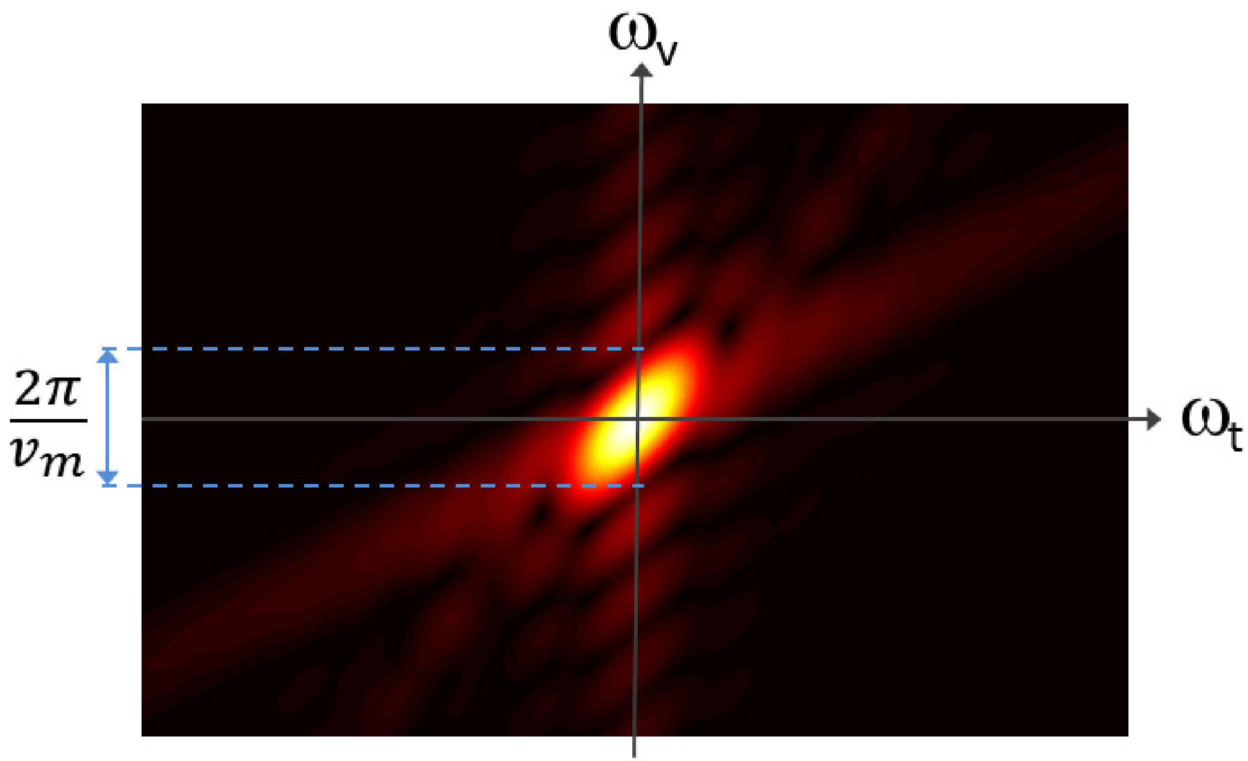
Estimating the Essential Bandwidth



The spectrum along $\omega_t = 0$:

$$Q_S(0, \omega_v) = \frac{2v_m}{\tan(\phi)} \ln\left(\frac{z_{max}}{z_{min}}\right)\left[\text{sinc}\left(\omega_v v_m\right) - j\frac{v_m \tan(\phi)}{f}\text{sinc}'\left(\omega_v v_m\right)\right]$$

Estimating the Essential Bandwidth

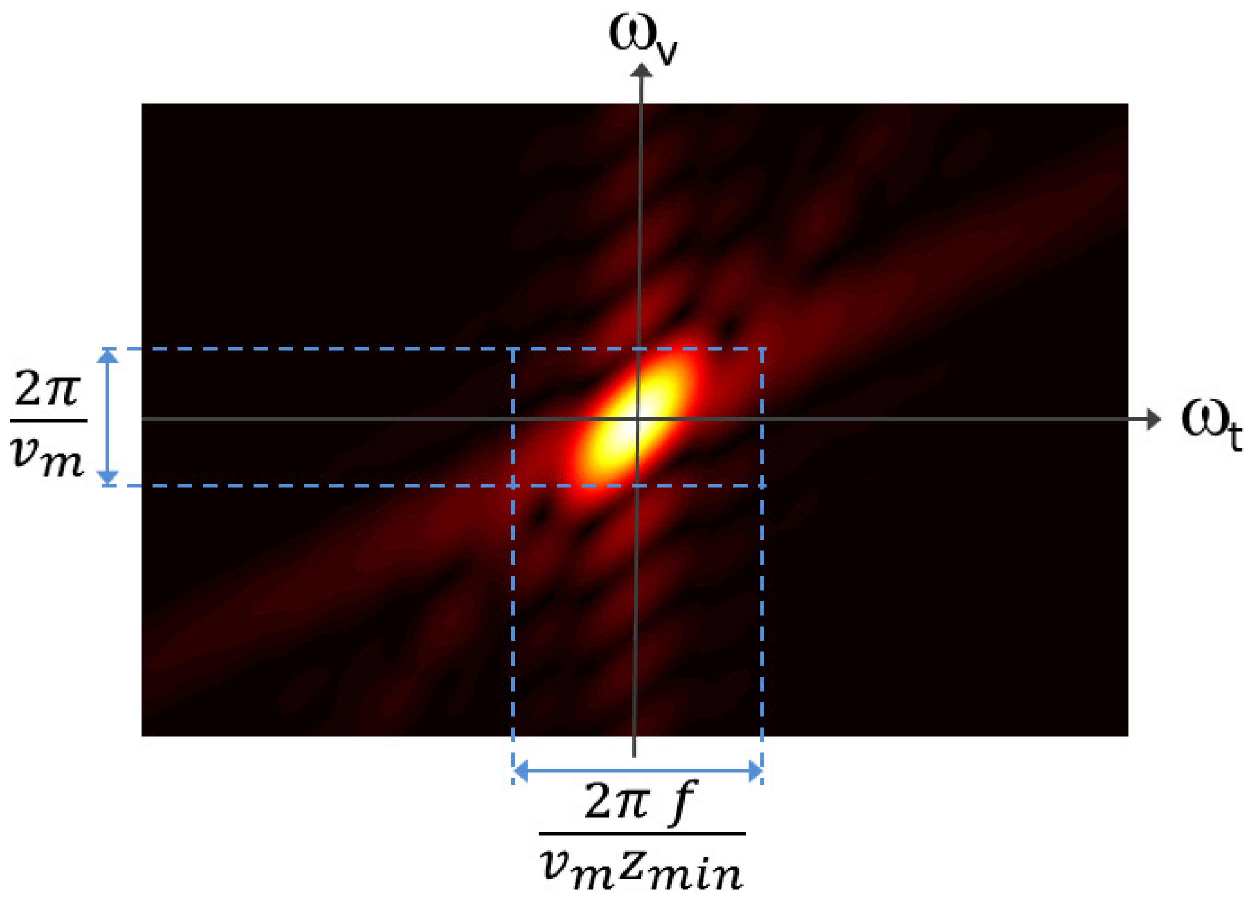


Estimate the essential bandwidth along $\omega_t = 0$ as:

$$Q_S(0, \omega_v) \sim \text{sinc}\left(\omega_v v_m\right)$$

↬ EB of sinc function = width of main lobe, $\frac{2\pi}{v_m}$

Estimating the Essential Bandwidth



Project onto ω_t-axis using:

$$\omega_v = \omega_t \frac{z_{min}}{f}$$

Depth Cameras vs Colour Cameras

Maximum spacing between depth cameras for a slanted plane:

$$\Delta t_d = \frac{z_{min} v_m}{f}$$

Maximum spacing between colour cameras for a slanted plane:

$$\Delta t_c = \frac{2\pi z_{opt} v_m}{v_m \Omega_t T \left|\sin(\phi)\right| + 2\pi\, f}, \quad \text{where} \quad \Omega_t = \frac{\omega_s f}{f\cos(\phi) - v_m \left|\sin(\phi)\right|} + \frac{2\pi}{T},$$

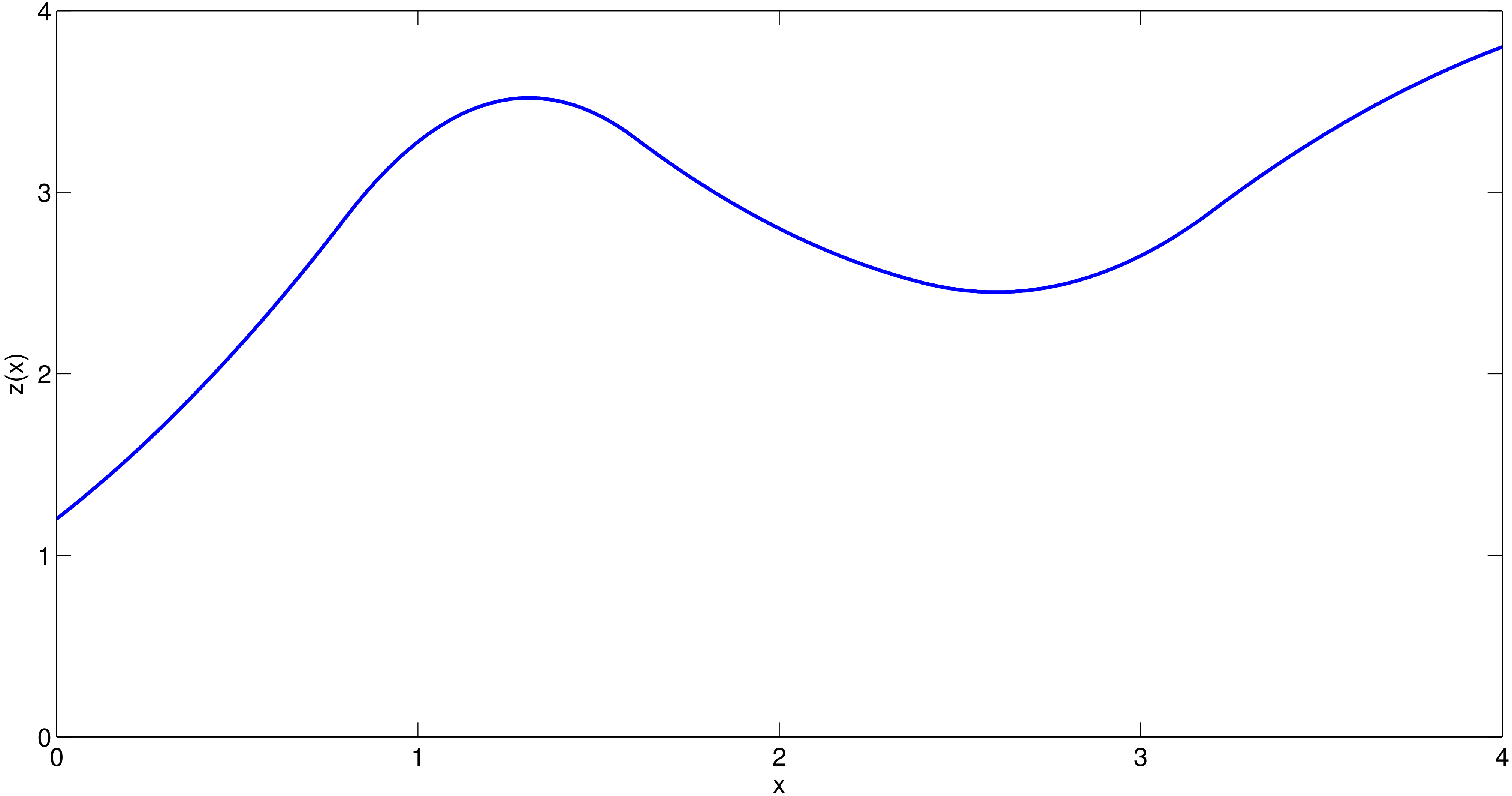
and z_{opt} is the average depth of the scene.

Observations:

- A point in the scene will be seen by at least two depth cameras
- Assuming $\omega_s > 1\, rad/m \implies$ Fewer depth cameras required than color cameras
- If plane fronto-parallel ($\phi = 0$) $\implies \Delta t_c = \Delta t_d$

Results: Synthetic Scene

Sample and reconstruct plenoptic function $\Longrightarrow$ Adaptive sampling algorithm[6]

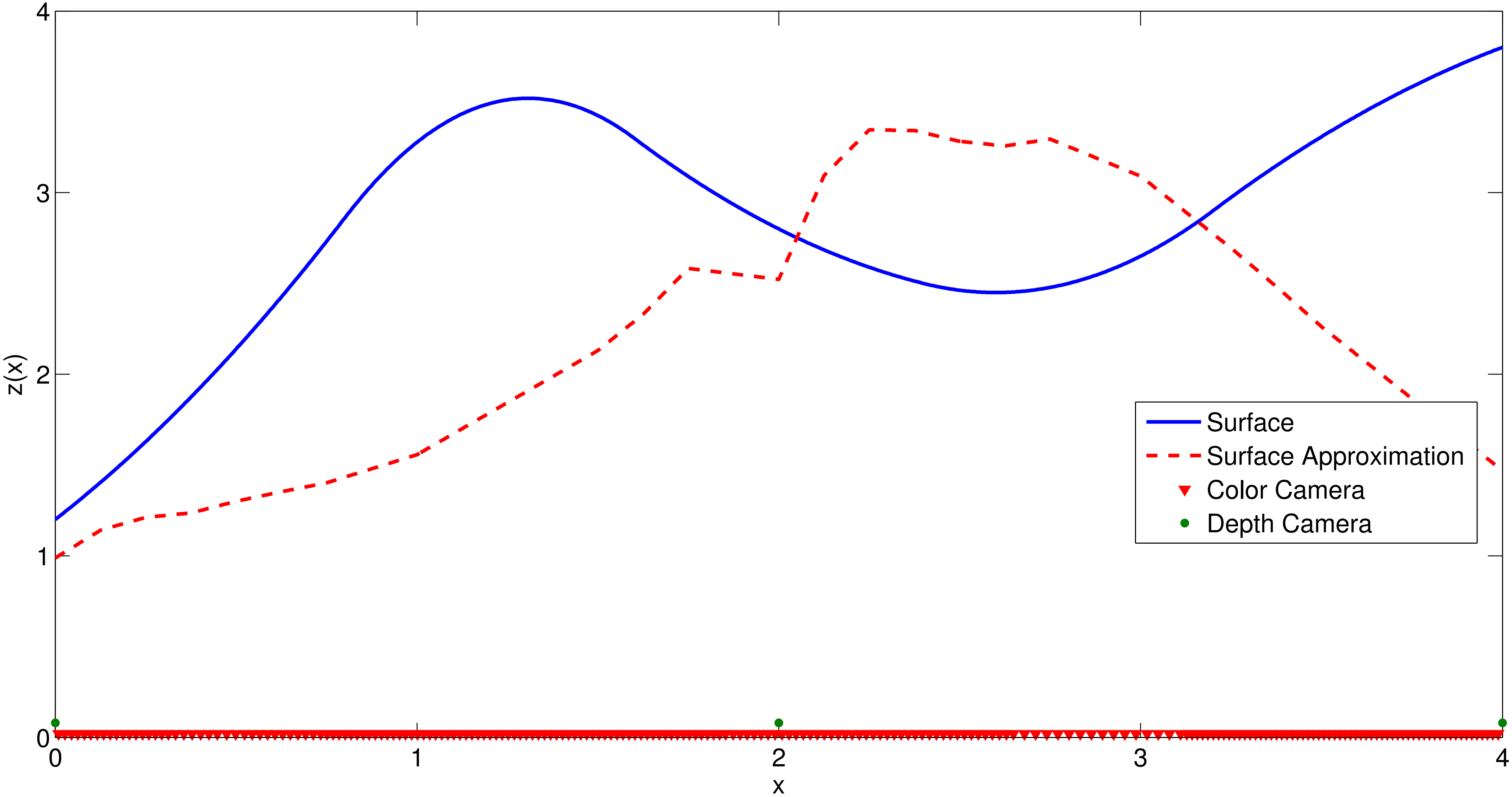


Piecewise quadratic surface with sinusoidal texture

$\looparrowright$ Estimate scene geometry from the Pantelic function

Results: Synthetic Scene

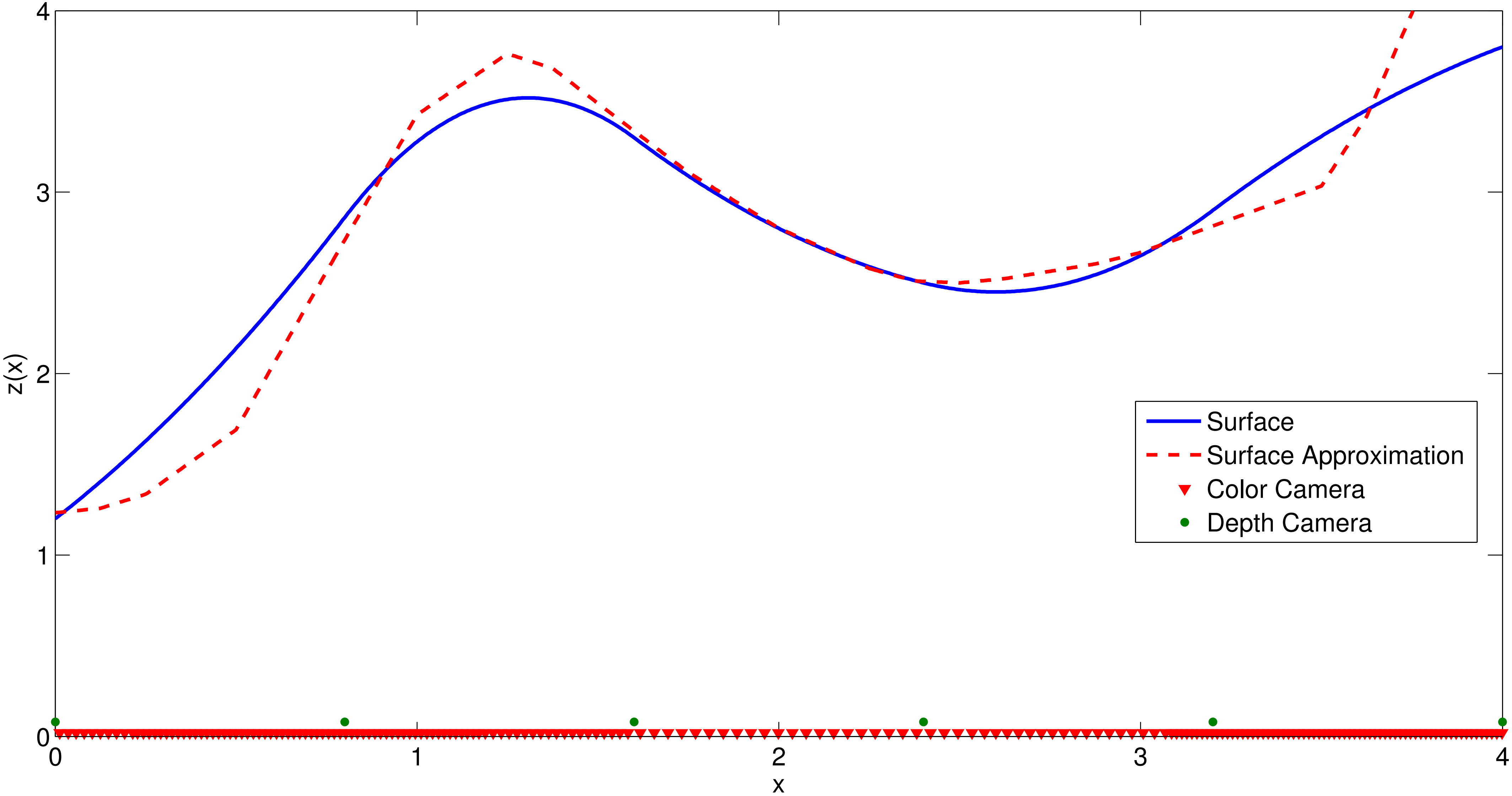
Analysing the position of the 250 color cameras:



3 Depth Cameras $\Longrightarrow$ Plenoptic PSNR $= 20.4$dB

Results: Synthetic Scene

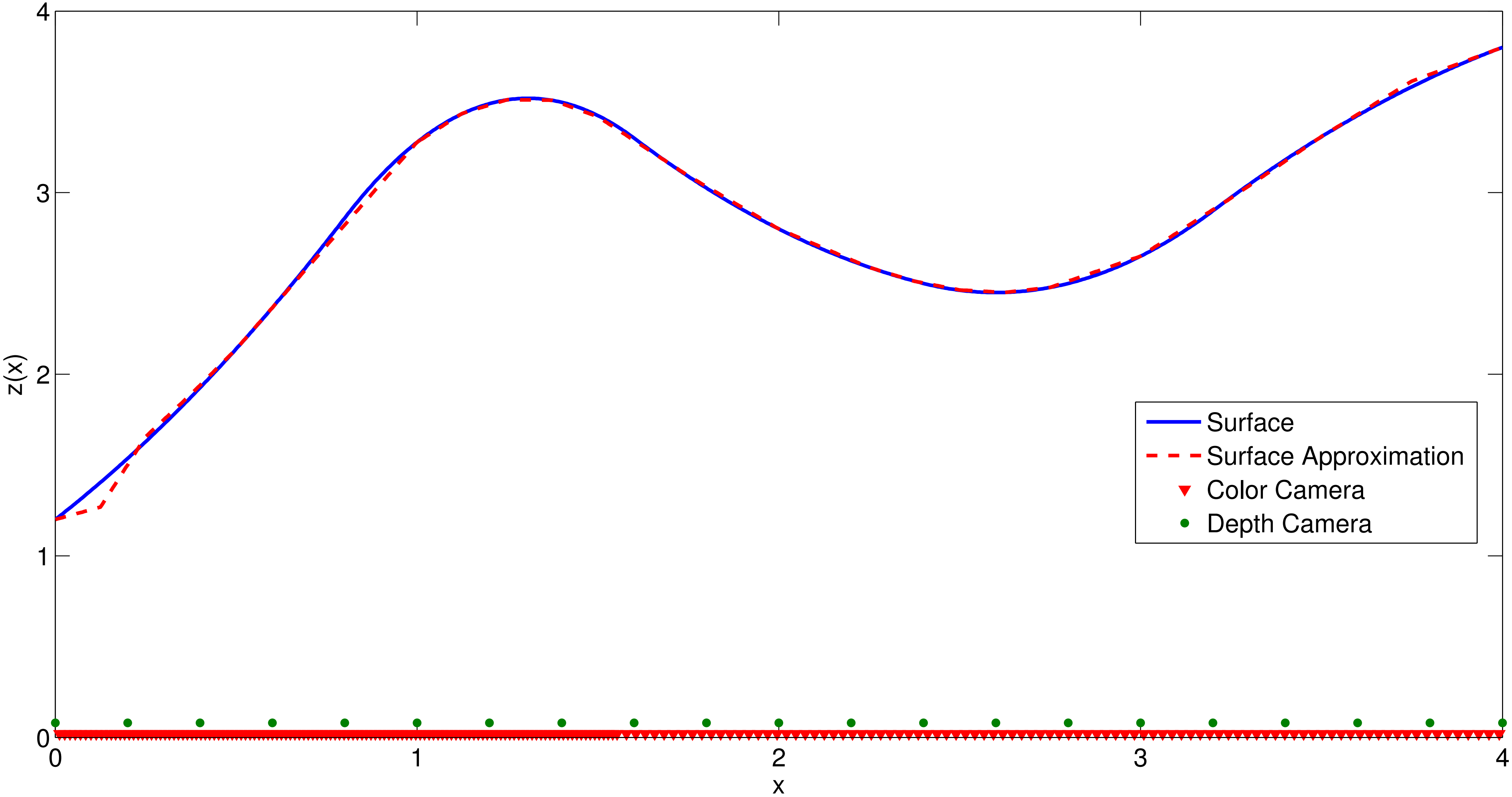
Analysing the position of the 250 color cameras:



6 Depth Cameras $\Longrightarrow$ Plenoptic PSNR $= 26.8$dB

Results: Synthetic Scene

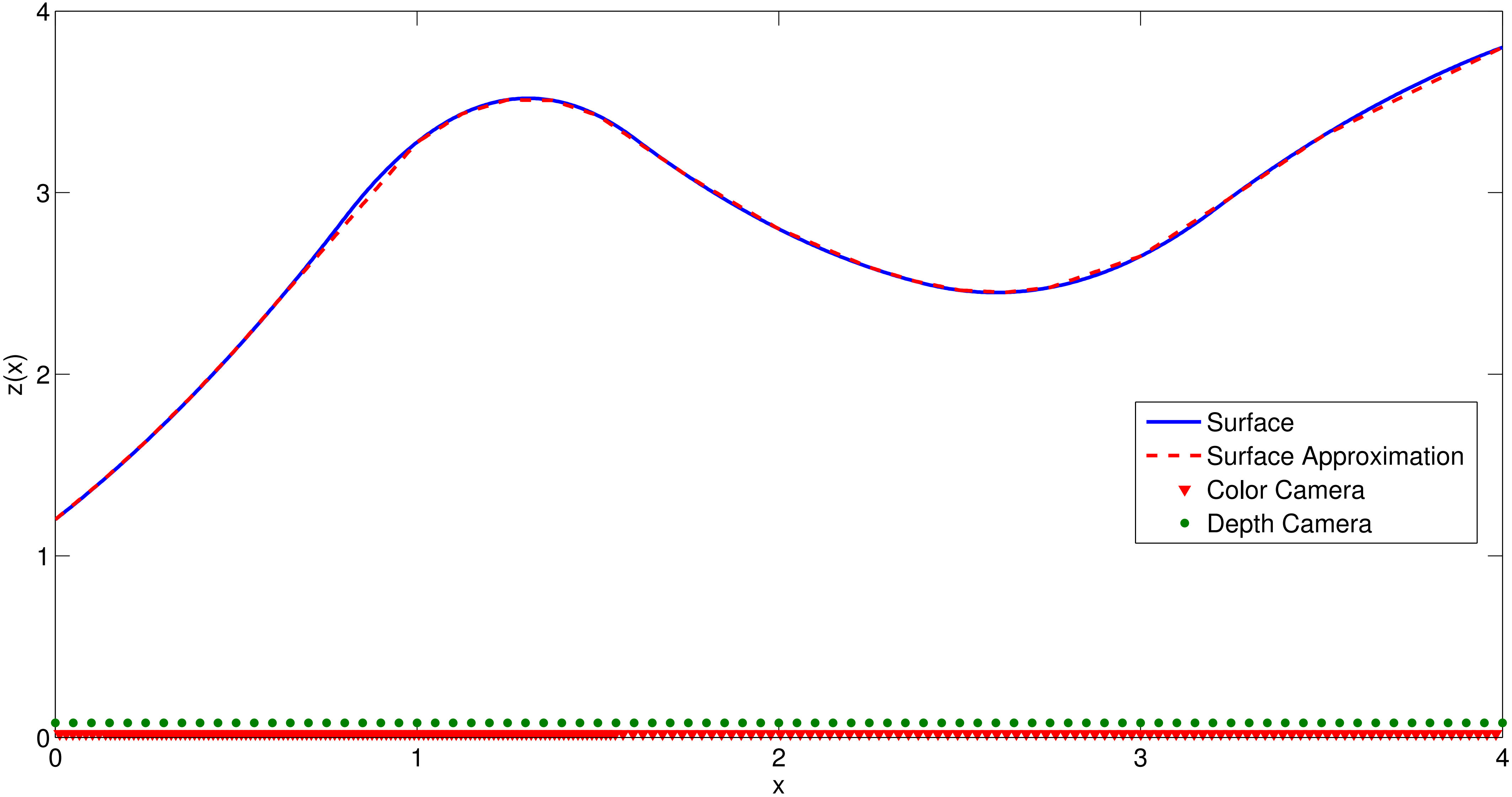
Analysing the position of the 250 color cameras:



21 Depth Cameras $\Longrightarrow$ Plenoptic PSNR $= 27.6$dB

Results: Synthetic Scene

Analysing the position of the 250 color cameras:

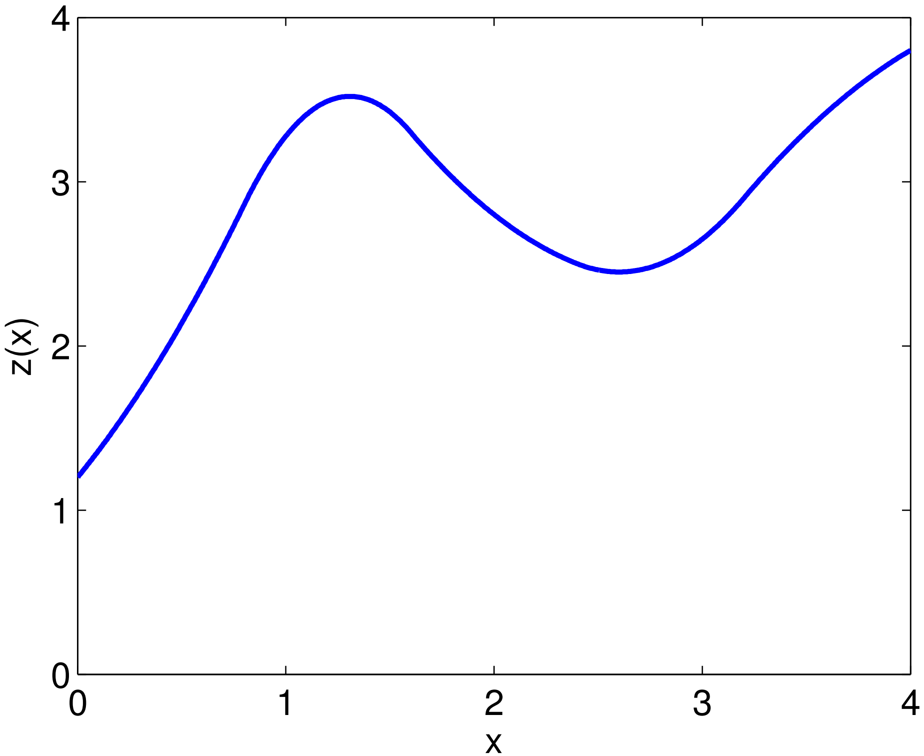


81 Depth Cameras $\Longrightarrow$ Plenoptic PSNR $= 27.6$dB

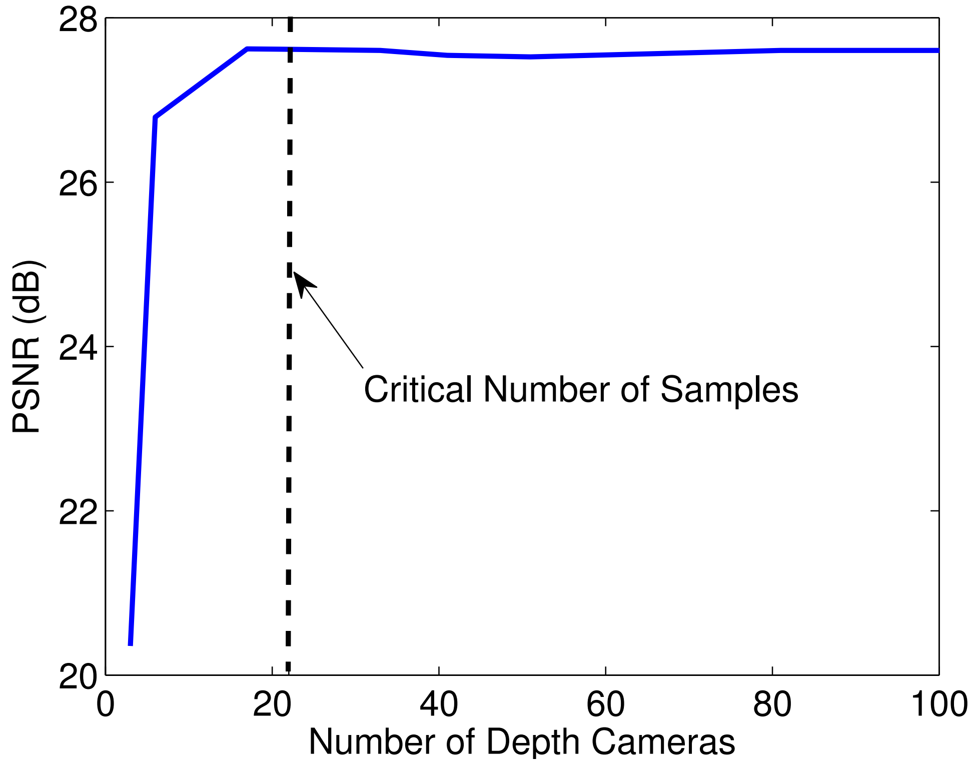
Results: Synthetic Scene

Approximating the number of depth cameras with

$$\Delta t_d = \frac{z_{min} v_m}{f}$$

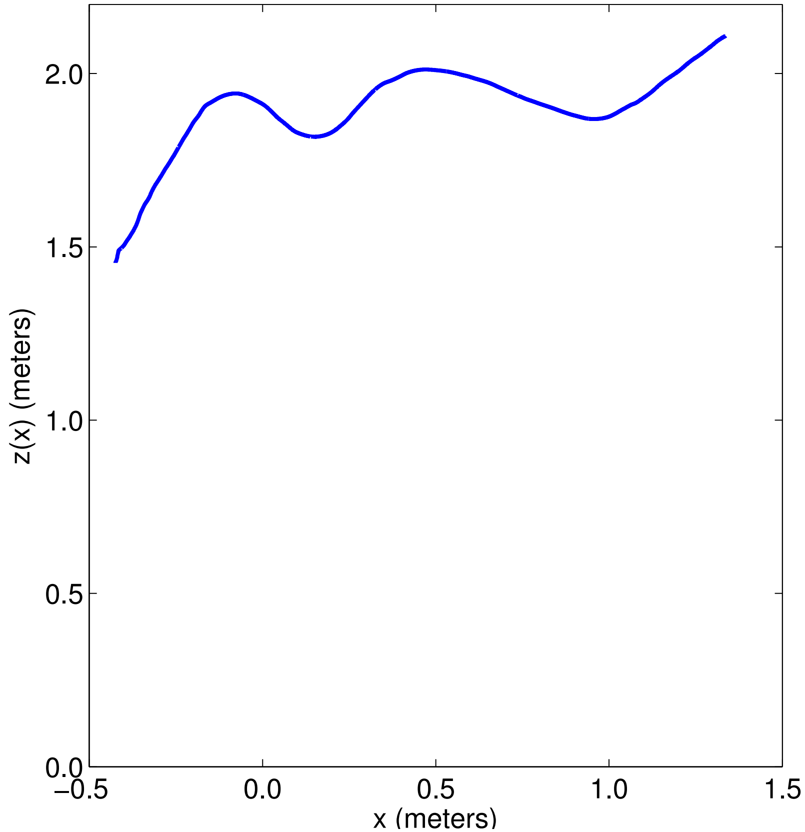


(a) Piecewise quadratic surface with sinusoidal texture



(b) PSNR of the reconstructed Plenoptic Function

Results: Real Scene



(a) Scene Geometry

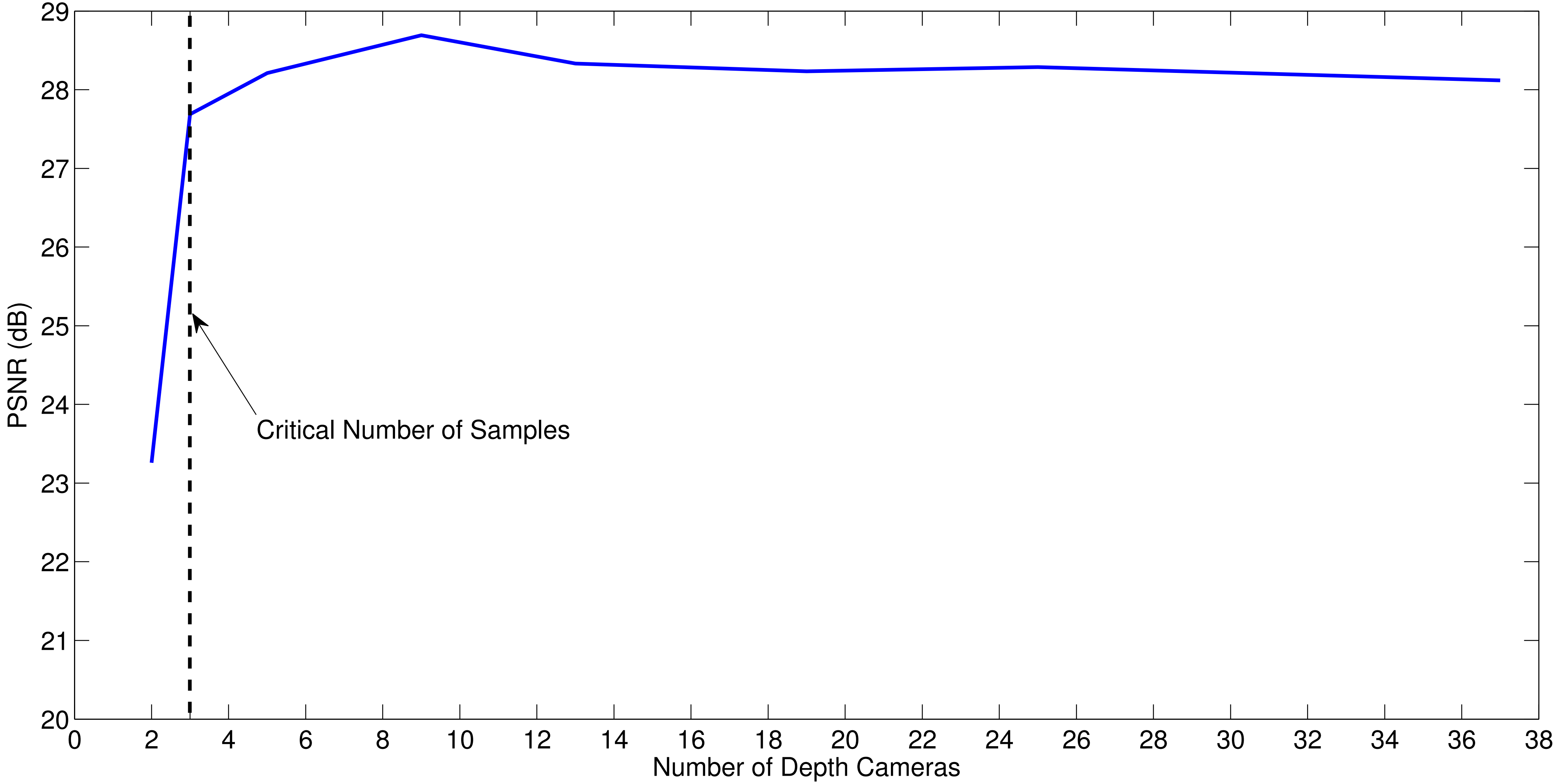
(b) Acquiring the Data

$\looparrowright$ Initial Image Set = 73 Color and Depth Images (1cm apart)

$\Longrightarrow$ Color Images 3008 by 2000 pixels $\Longrightarrow$ Depth Images 374 by 248 pixels

Results: Real Scene

PSNR of Plenoptic Function as the Number of Depth Cameras Increases:



↪ Number of Color Images = 37

Conclusions

- Presented a framework, similar to the plenoptic function, for multi-view depth images.
- Extended the spectral analysis of a slanted plane to multi-view depth images, assuming finite field of view and scene width.
- Maximum depth camera spacing for a slanted plane depends on the minimum depth and the camera characteristics.
- Empirical results, for synthetic and real scenes, show that fewer depth images are required than colour images for IBR view synthesis.

References

1. J. Pearson, P.L. Dragotti and M. Brookes. Accurate non-iterative depth layer extraction algorithm for image based rendering. IEEE ICASSP 2011, pages 901-904.
2. E.H. Adelson and J.R. Bergen. The plenoptic function and the elements of early vision. In Computational Models of Visual Processing, pages 3-20. MIT Press, Cambridge, MA, 1991.
3. J.X. Chai, S.C. Chan, H.Y. Shum, and X. Tong. Plenoptic sampling. In Computer graphics (SIGGRAPH'00), pages 307-318, 2000.
4. M.N. Do, D Marchand-Maillet, and M. Vetterli. On the bandwidth of the plenoptic function. IEEE Transactions on Image Processing, vol. 28, pages 708-717, 2012.
5. 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. IEEE ICIP 2010, pages 3965-3968.
6. C. Gilliam, P.L. Dragotti, and M. Brookes. Adaptive Plenoptic Sampling. IEEE ICIP 2011, pages 2581-2584.
7. T. Ishibashi, T. Yendo, M.P. Tehrani, T. Fujii and M. Tanimoto. 3D Space representation using epipolar plane depth image. Picture Coding Symp., pages 22-25, 2010