Imperial College London

On the Sampling and Compression of the Plenoptic Function

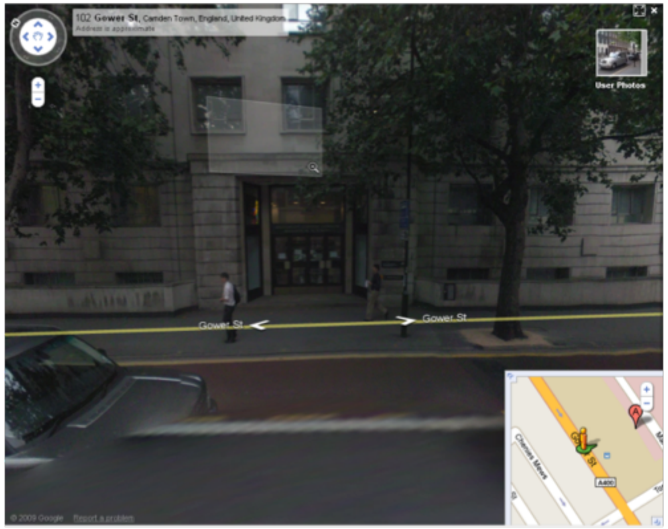
Pier Luigi Dragotti
Imperial College London

Joint work with J. Berent (Google), M. Brookes (ICL), A. Gelman (ICL), C. Gilliam (ICL), J. Pearson (ICL), V.Velisavljevic (Deutsche Telekom).

Imperial College London

The problem: Rendering Novel Views

iview BBC Research

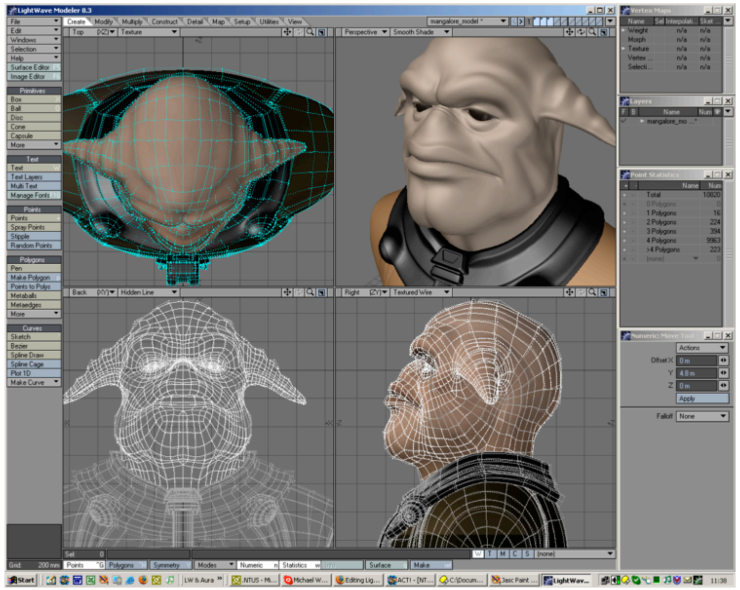


*There is a need for scalable, fast and unsupervised algorithms that can give the user a **photo-realistic** 'being there' experience*

Imperial College
London

Model-Based Rendering

- Detailed geometry and texture/ reflectance maps are available or estimated from the images
- New views are rendered by projecting the objects onto the virtual camera planes
- Source description



[Obtained from wikipedia]

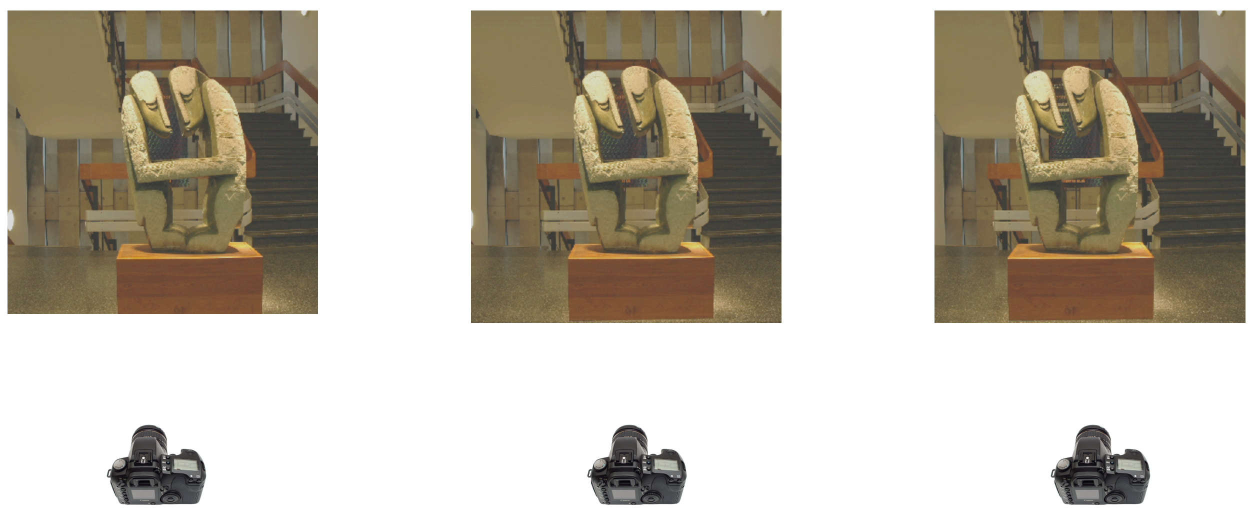
[Obtained from Middlebury stereo vision]

The models can be very difficult to obtain from real images

Imperial College
London

Image Based Rendering

- IBR uses many images of the scene (100-1000)
- New views are synthesized by interpolating intensities from nearby available images. Little or no geometry and can be very fast
- Photo-realistic rendering of complicated environments
- Appearance description

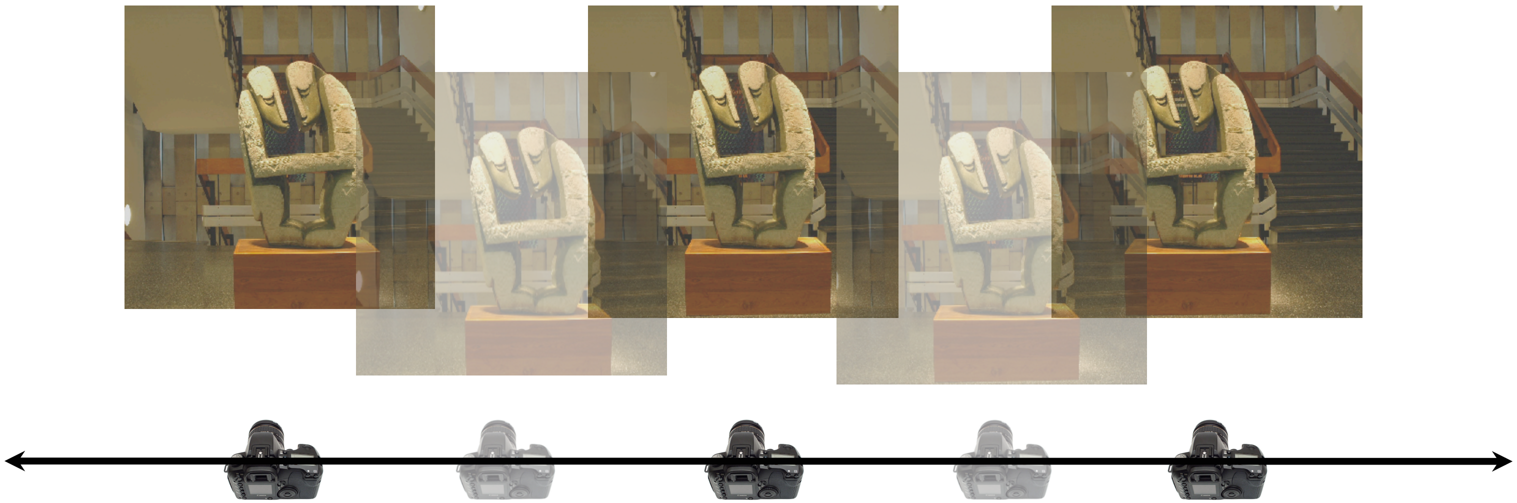


Available images

Imperial College London

Image Based Rendering

- IBR uses many images of the scene (100-1000)
- New views are synthesized by interpolating intensities from nearby available images. Little or no geometry and can be very fast
- Photo-realistic rendering of complicated environments
- Appearance description



Fill in the gaps and create a walkthrough environment

Imperial College London

Why Image Based Rendering?

- Data driven society: Cameras are everywhere!
- Mobile phone subscription to hit 5 billion this year. Most phones are equipped with cameras.
- IBR is a classical sampling and interpolation problem.

- Open Questions: How many cameras and where to place them? What is the optimal trade-off between geometry information and number of cameras given a certain scene?

Notice: Image courtesy of Chris Jordan

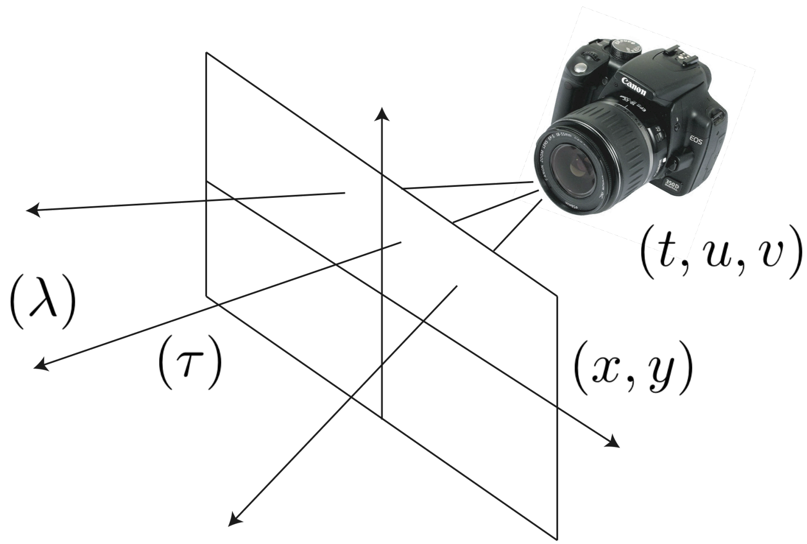
Talk Outline

1. Structure of the data: The plenoptic function, the EPI and the lightfield
2. Spectral Analysis of the plenoptic function
3. Essential Plenoptic Bandwidth and Adaptive Sampling
4. Layer Based Sampling and Interpolation:
   1. Plenoptic layers: Extraction and interpolation algorithms
   2. Adaptive methods
5. Distributed and centralized compression of the plenoptic function
6. Conclusions and Outlook

The Plenoptic Function

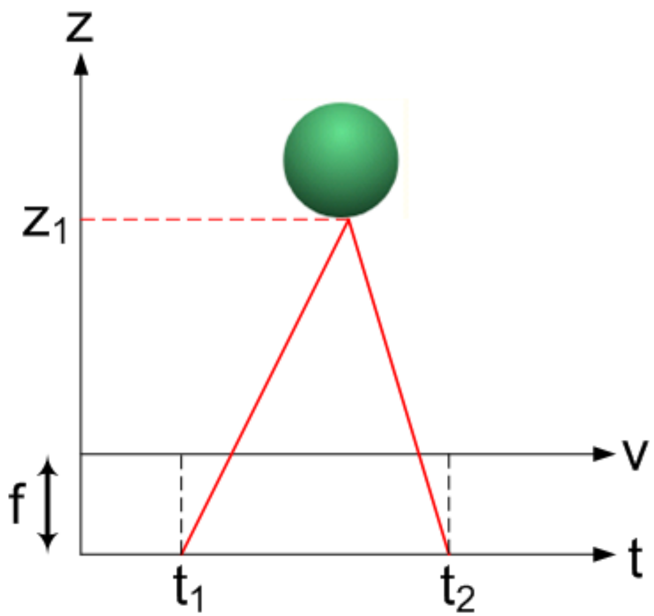
- "The sole communication link between physical objects and their corresponding images" – Adelson/Bergen
- 7D function that describes the intensity of each light ray that reaches a point in space [AdelsonB:91]
- Assumption can be made to reduce the number of dimensions:
    - Intensity remains constant unless occluded
    - 3 channels for RGB
    - Static scenes
    - Viewing position constraints

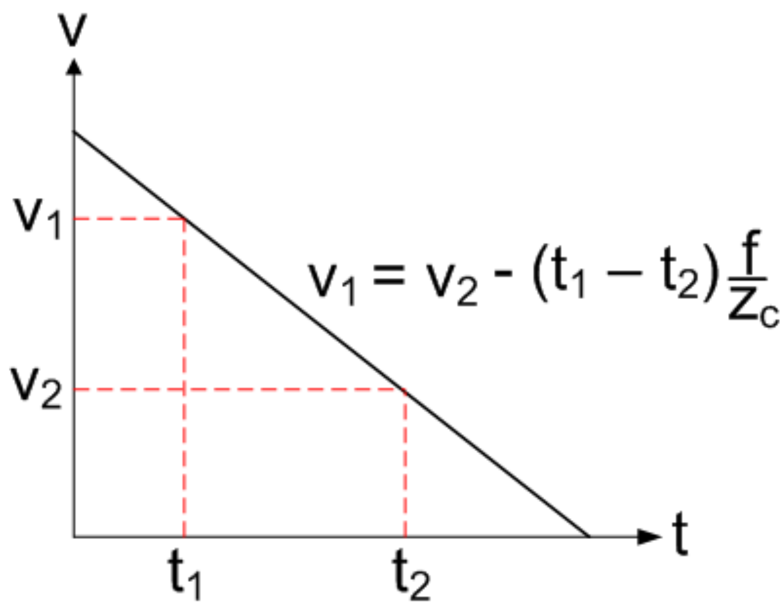
$$I = I(x, y, \lambda, \tau, t, u, v)$$



Epipolar Plane Image (EPI)

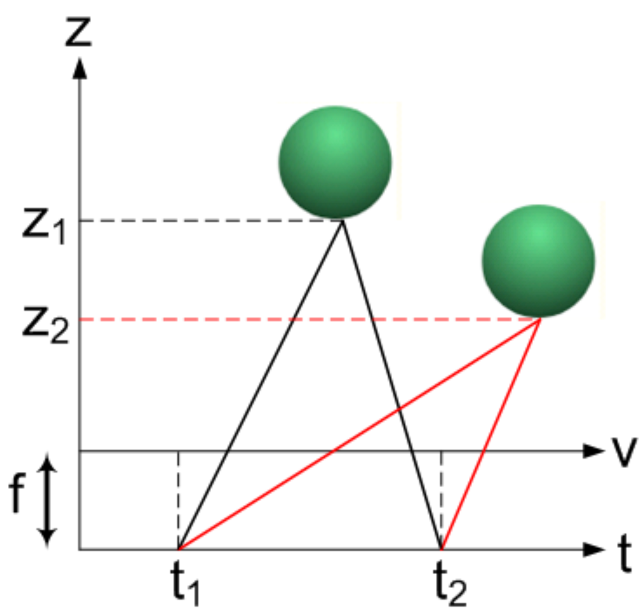
- Pinhole camera model
- Points are mapped onto lines in the (EPI)
- Slope of lines are inversely proportional to the depth
- Lines with larger slopes occlude lines with smaller slopes

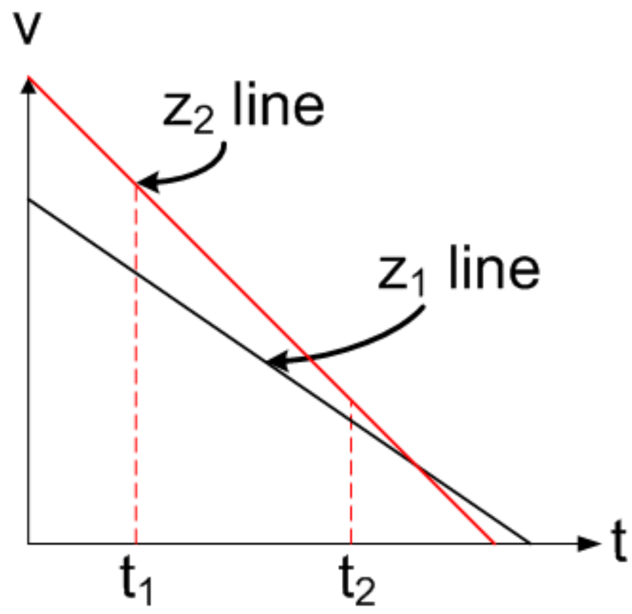




Epipolar Plane Image (EPI)

- Pinhole camera model
- Points are mapped onto lines in the (EPI)
- Slope of lines are inversely proportional to the depth
- Lines with larger slopes occlude lines with smaller slopes

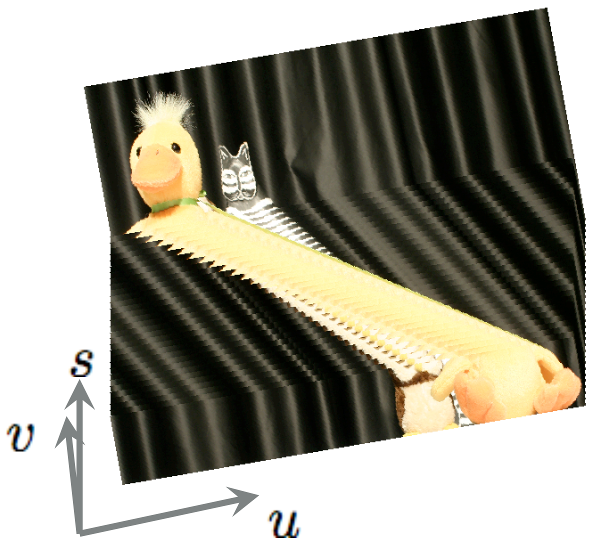


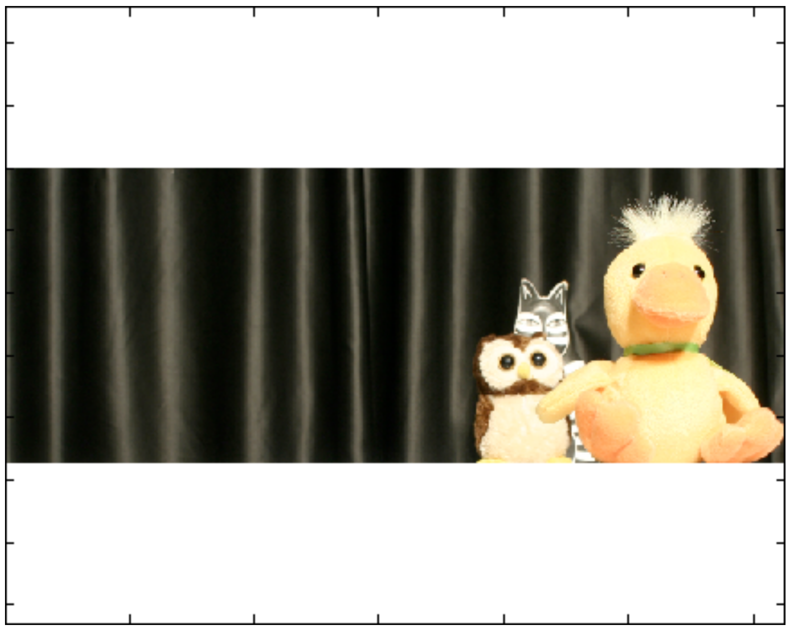


Imperial College London

EPI Structure

- Pinhole camera model
- Points are mapped onto lines in the (EPI)
- Slope of lines are inversely proportional to the depth
- Lines with larger slopes occlude lines with smaller slopes

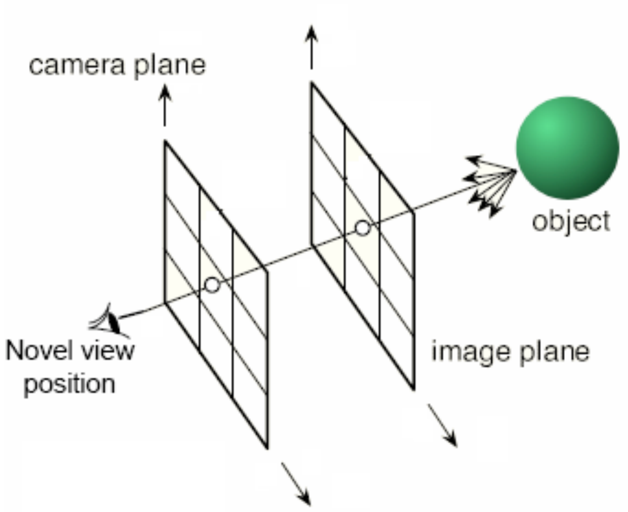


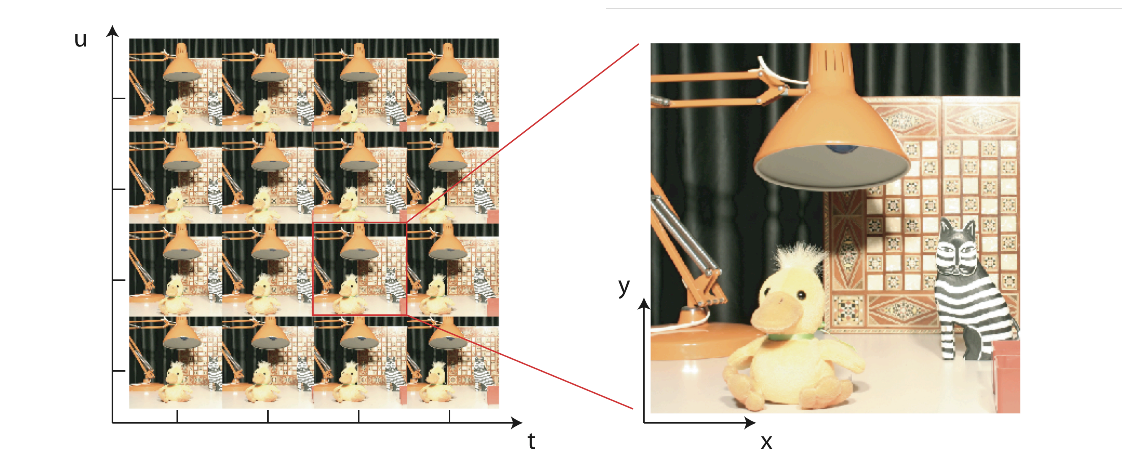


Imperial College London

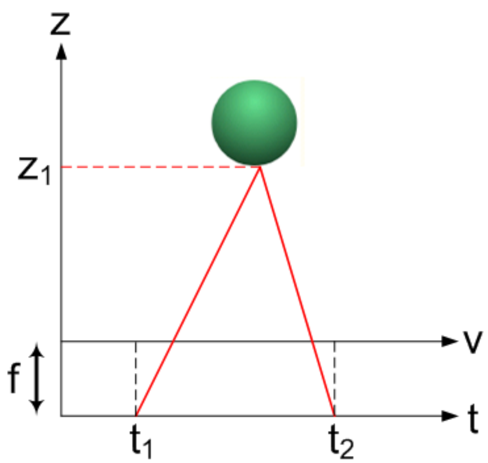
The Light Field

- First introduced in [LevoyH96]
- Light rays are characterized by their intersection with the camera plane and the image plane
- 4D parameterization of the plenoptic function

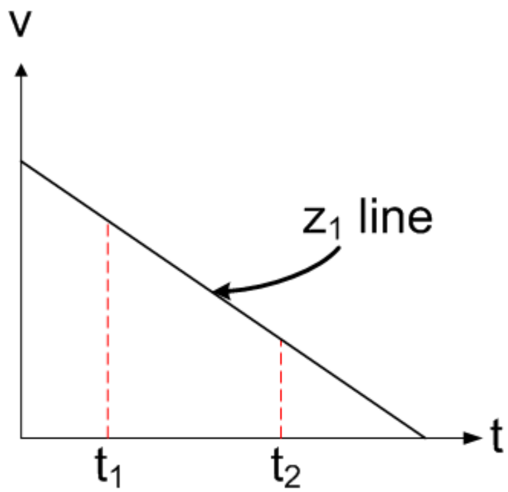




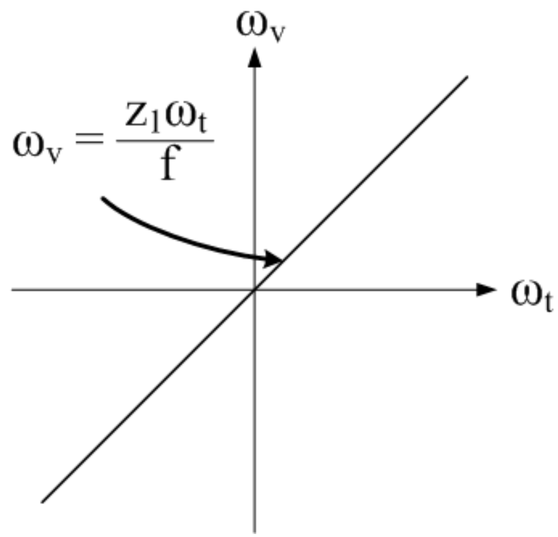
Plenoptic Spectral Analysis



(a) Scene

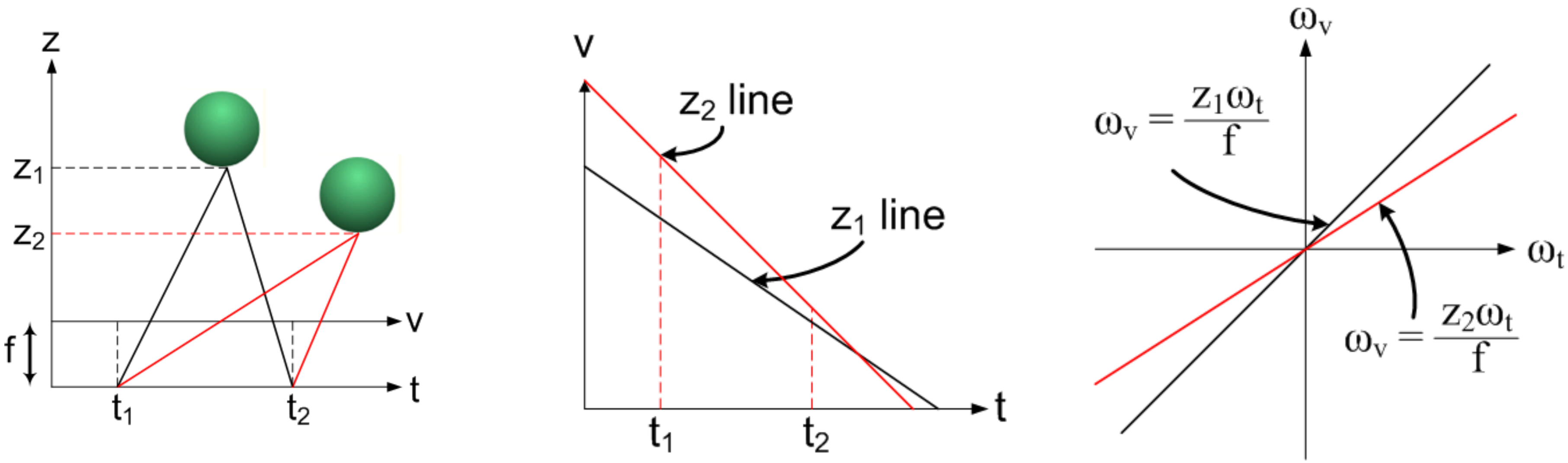


(b) EPI



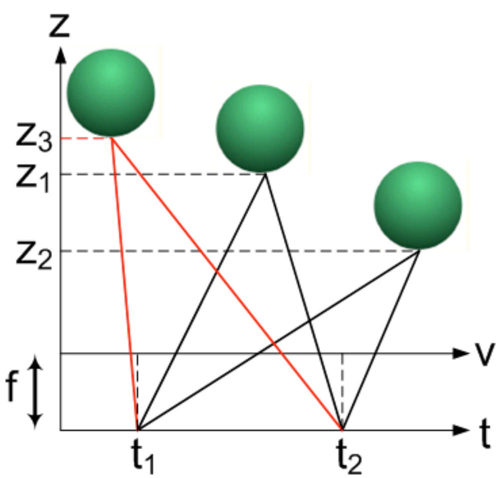
(c) Plenoptic Spectrum

Plenoptic Spectral Analysis

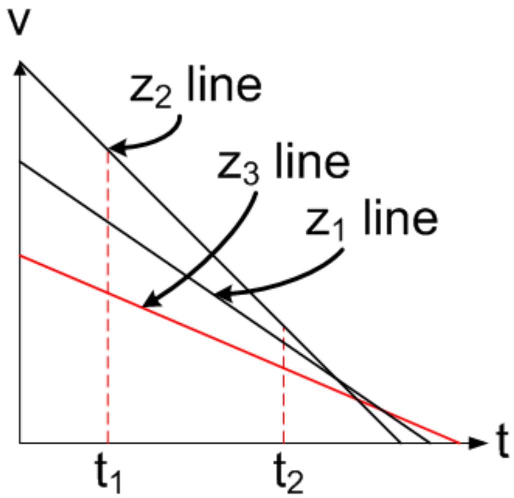


(a) Scene (b) EPI (c) Plenoptic Spectrum

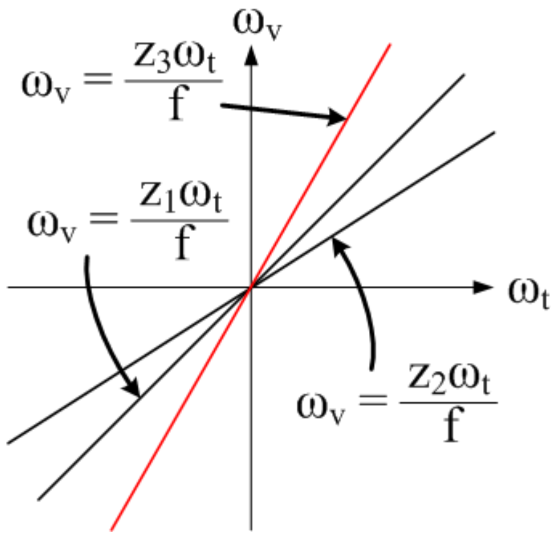
Plenoptic Spectral Analysis



(a) Scene



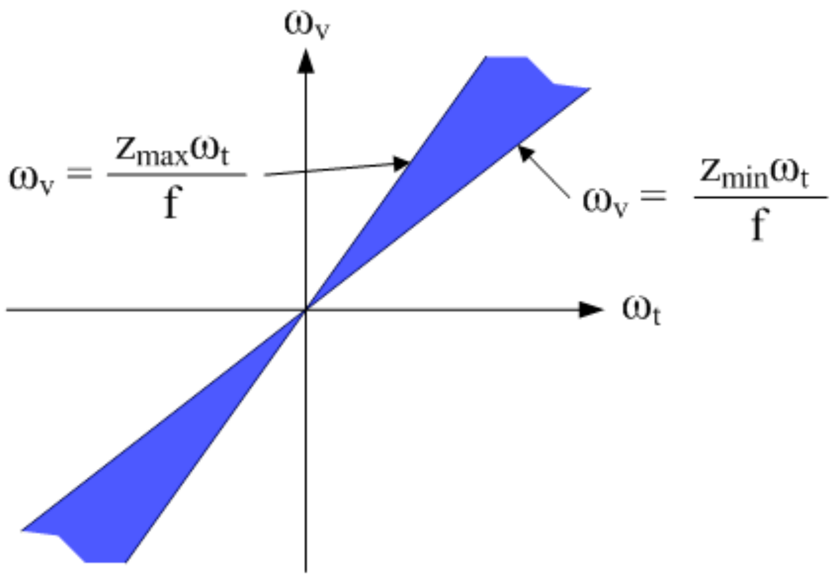
(b) EPI



(c) Plenoptic Spectrum

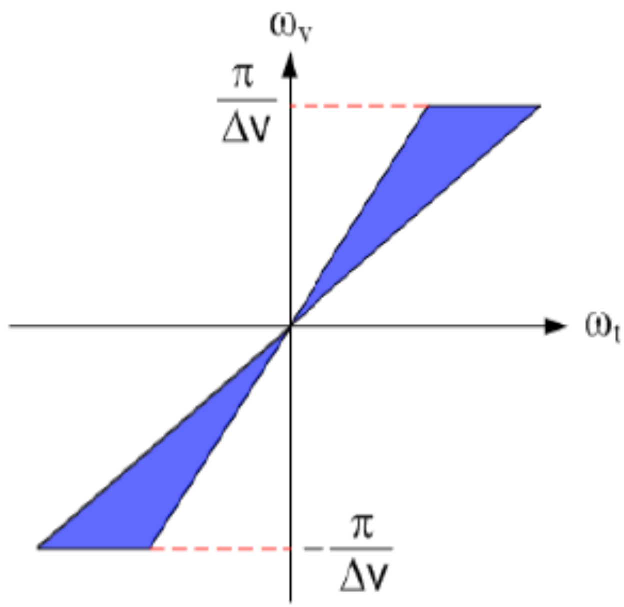
Plenoptic Spectral Analysis

Plenoptic spectrum exactly bound within two lines relating to the minimum and maximum depths of the scene [ChaiCST:00]

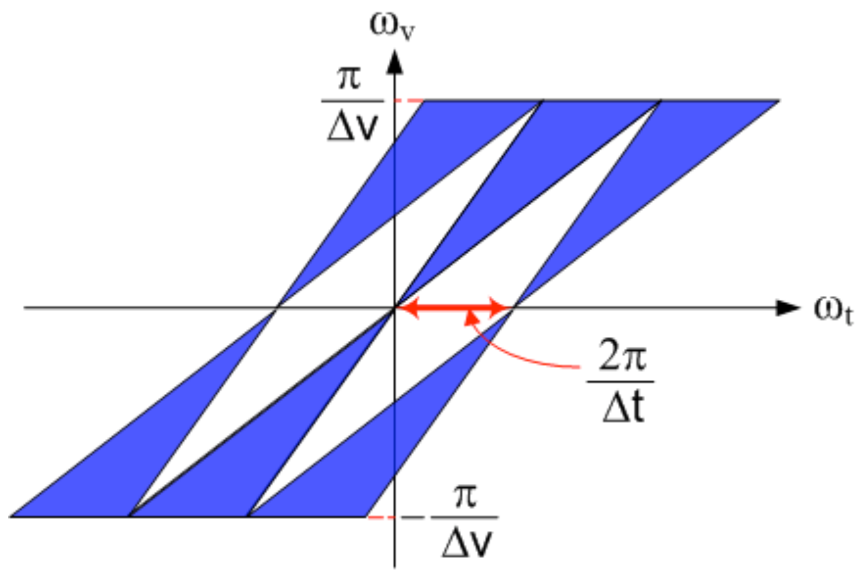


Plenoptic Sampling [ChaiCST:00]

- Finite camera resolution enforces lowpass filtering in ω_v
- Sampling in t of period Δt replicates the Plenoptic spectrum in ω_t



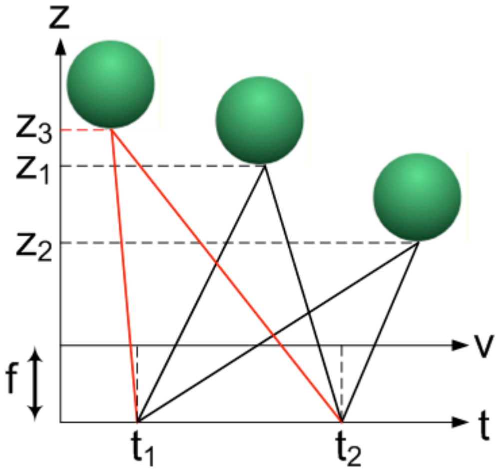
(a) Plenoptic Spectrum Sampled in *v*



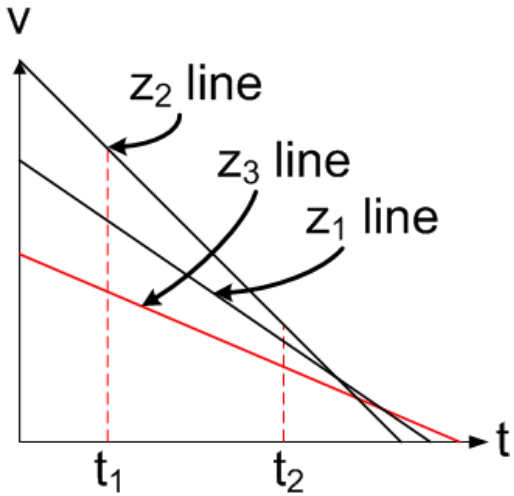
(b) Plenoptic Spectrum Sampled in *t*

Plenoptic Sampling [ChaiCST:00]

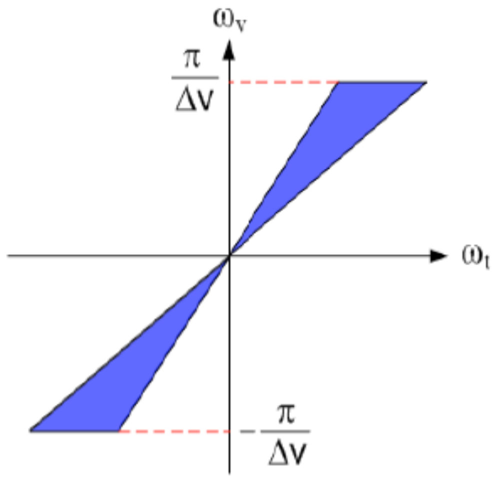
- Finite camera resolution enforces lowpass filtering in ω_v
- Sampling in t of period Δt replicates the Plenoptic spectrum in ω_t



(a) Scene



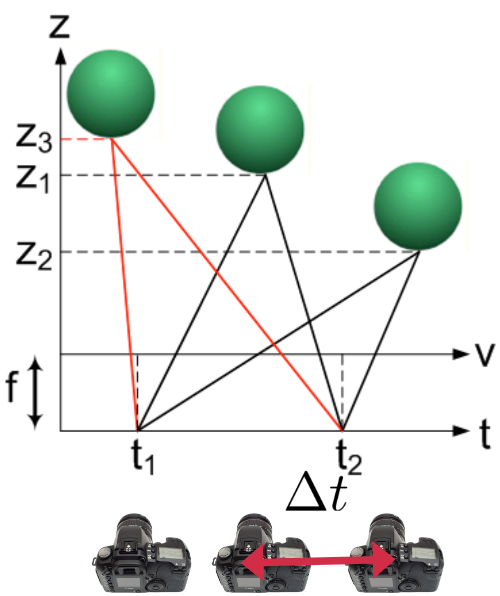
(b) EPI

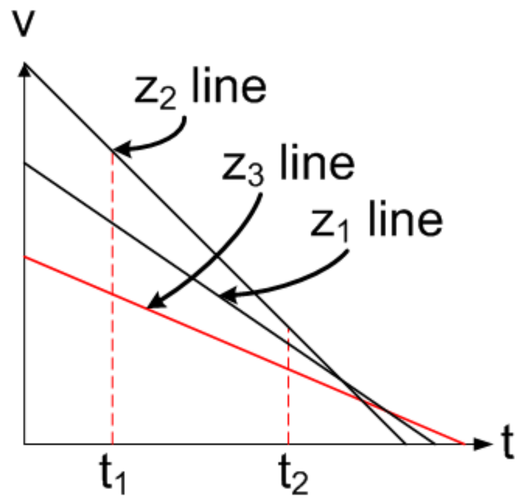


(c) Plenoptic Spectrum

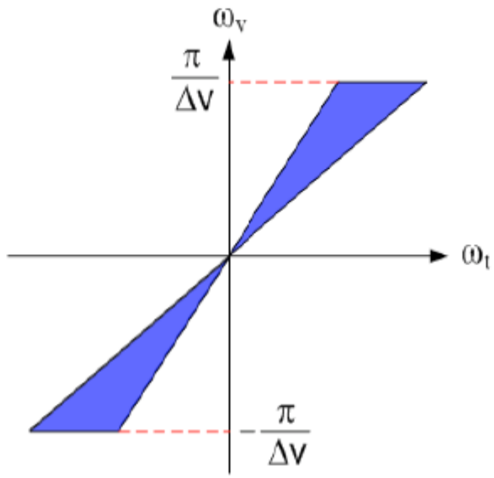
Plenoptic Sampling [ChaiCST:00]

- Finite camera resolution enforces lowpass filtering in ω_v
- Sampling in t of period Δt replicates the Plenoptic spectrum in ω_t





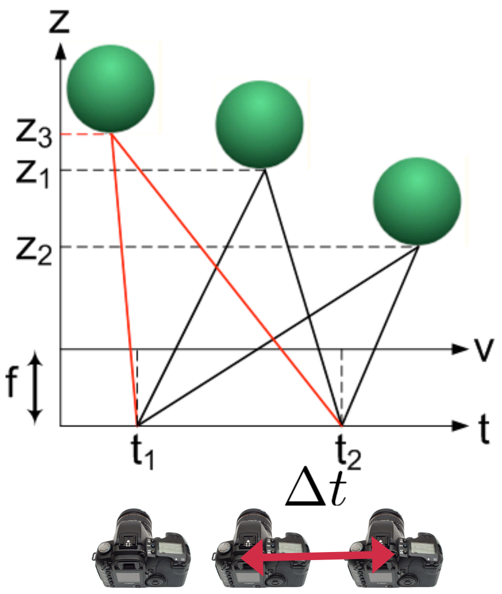
(b) EPI

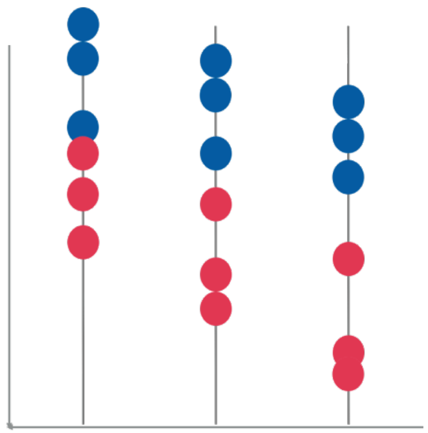


(c) Plenoptic Spectrum

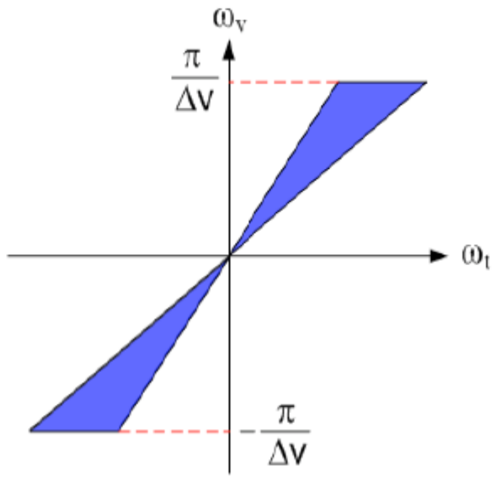
Plenoptic Sampling [ChaiCST:00]

- Finite camera resolution enforces lowpass filtering in ω_v
- Sampling in t of period Δt replicates the Plenoptic spectrum in ω_t





(b) Acquired Data

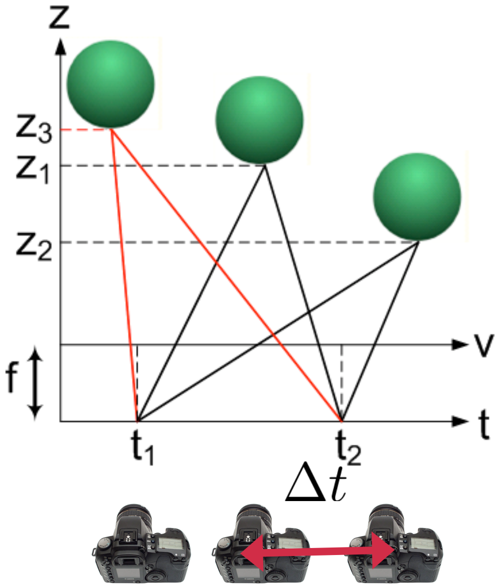


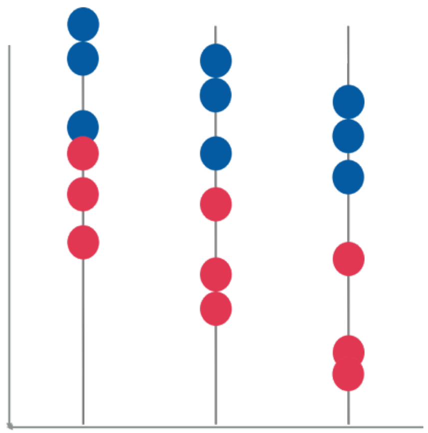
(c) Plenoptic Spectrum

Imperial College London

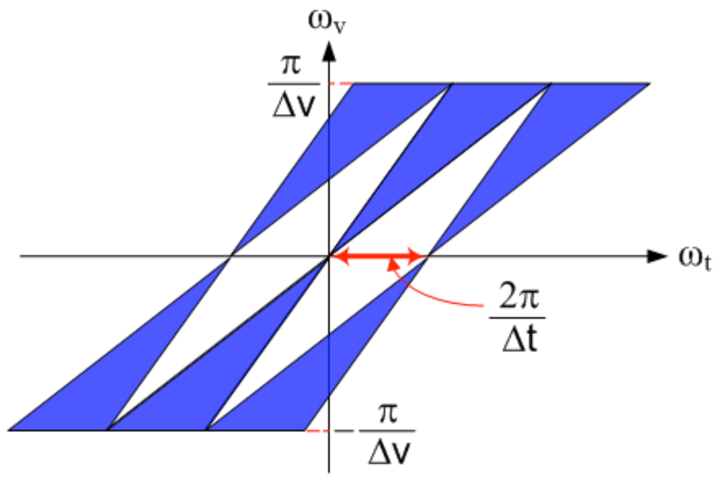
Plenoptic Sampling [ChaiCST:00]

- Finite camera resolution enforces lowpass filtering in ω_v
- Sampling in t of period Δt replicates the Plenoptic spectrum in ω_t





(b) Acquired Data

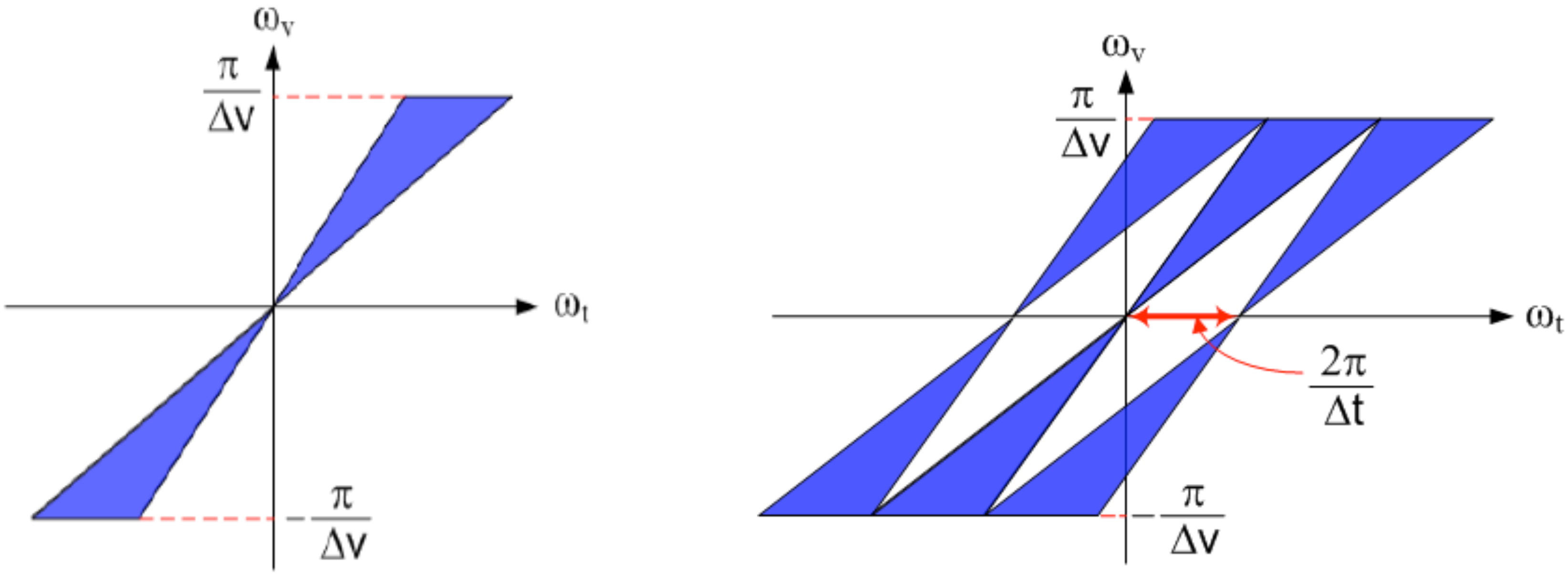


(c) Spectrum Sampled in t

Imperial College London

Plenoptic Sampling [ChaiCST:00]

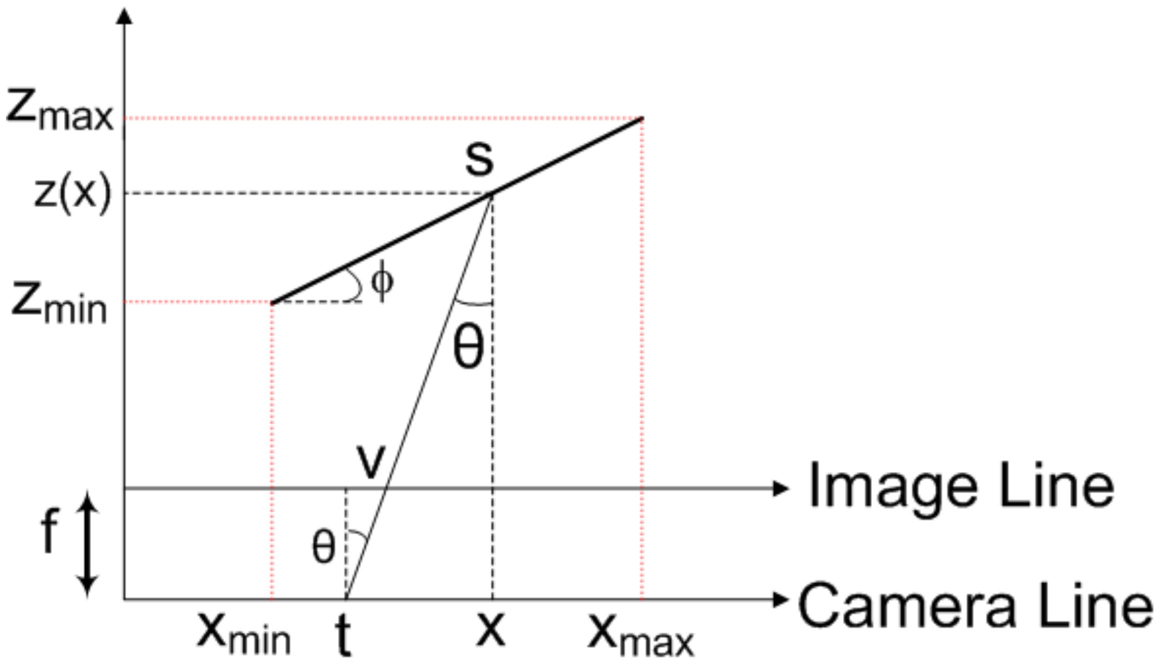
- Finite camera resolution enforces lowpass filtering in ω_v
- Sampling in t of period Δt replicates the Plenoptic spectrum in ω_t



(a) Plenoptic Spectrum Sampled in *v* (b) Plenoptic Spectrum Sampled in *t*

The sampling is exact only with an infinite flat plane and Infinite Field of View (FoV)

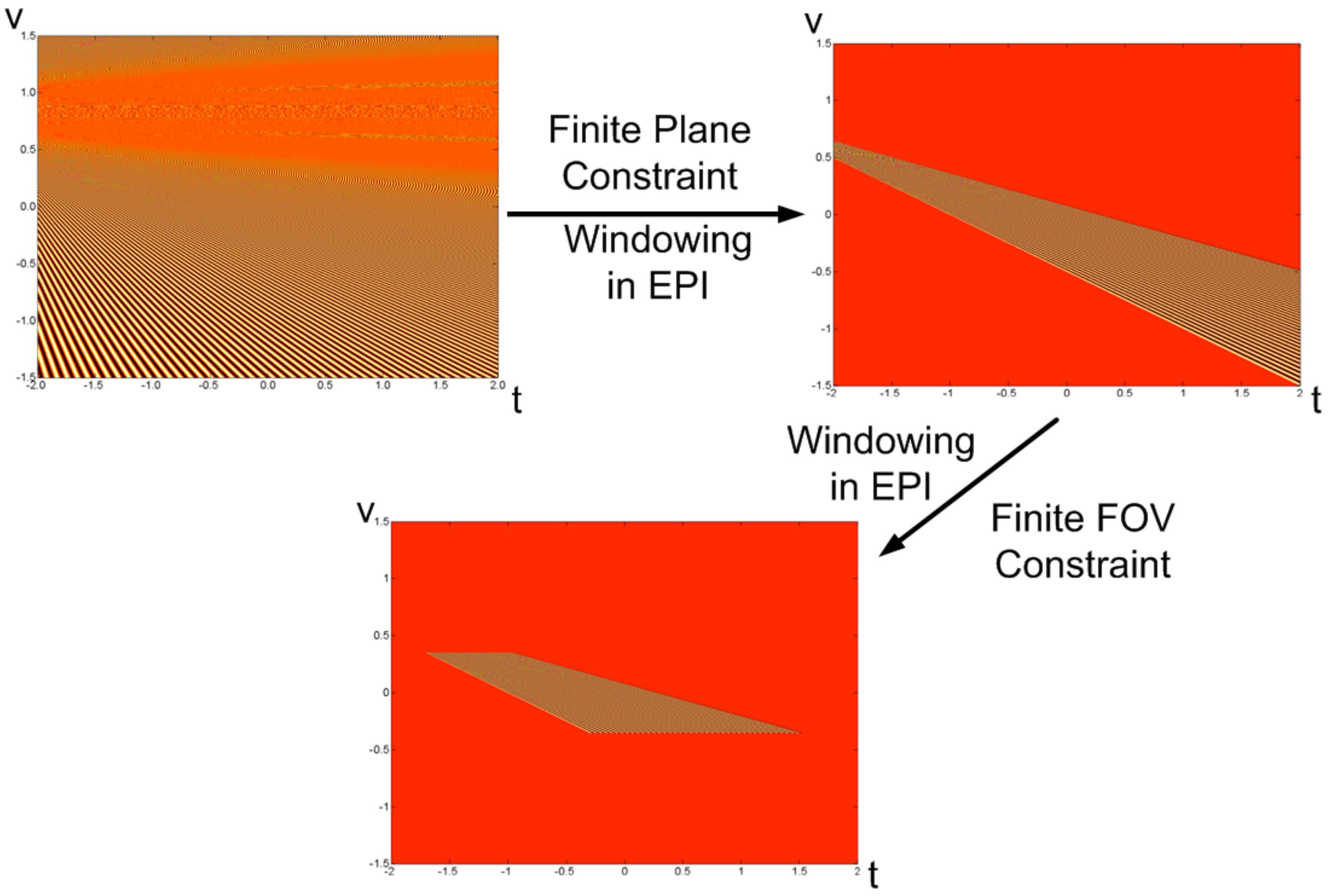
Slanted Plane Geometry



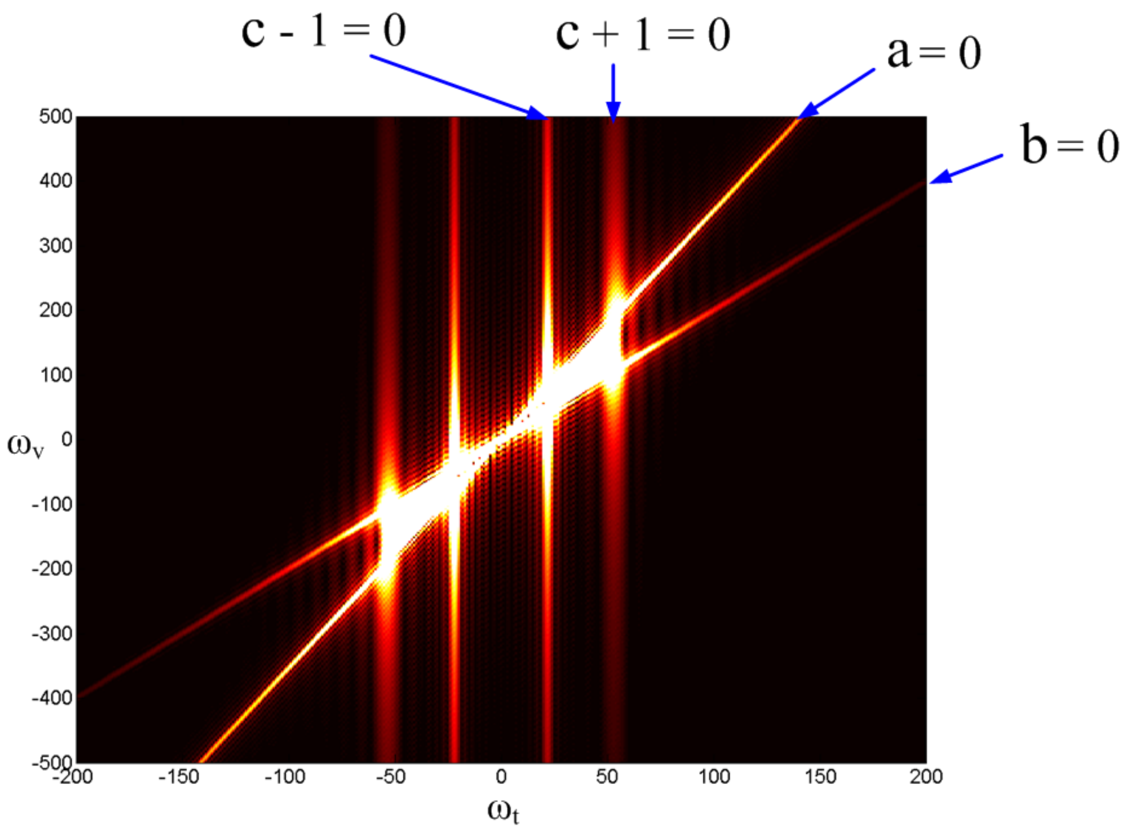
Set-up:

- Finite Field of View (FoV) for the Cameras => $v \in [-v_m, v_m]$
- Finite Plane Width
- Sinusoidal Texture Signal Pasted to Scene Surface
- Lambertian Scene

Slanted Plane Geometry: Effect on the EPI

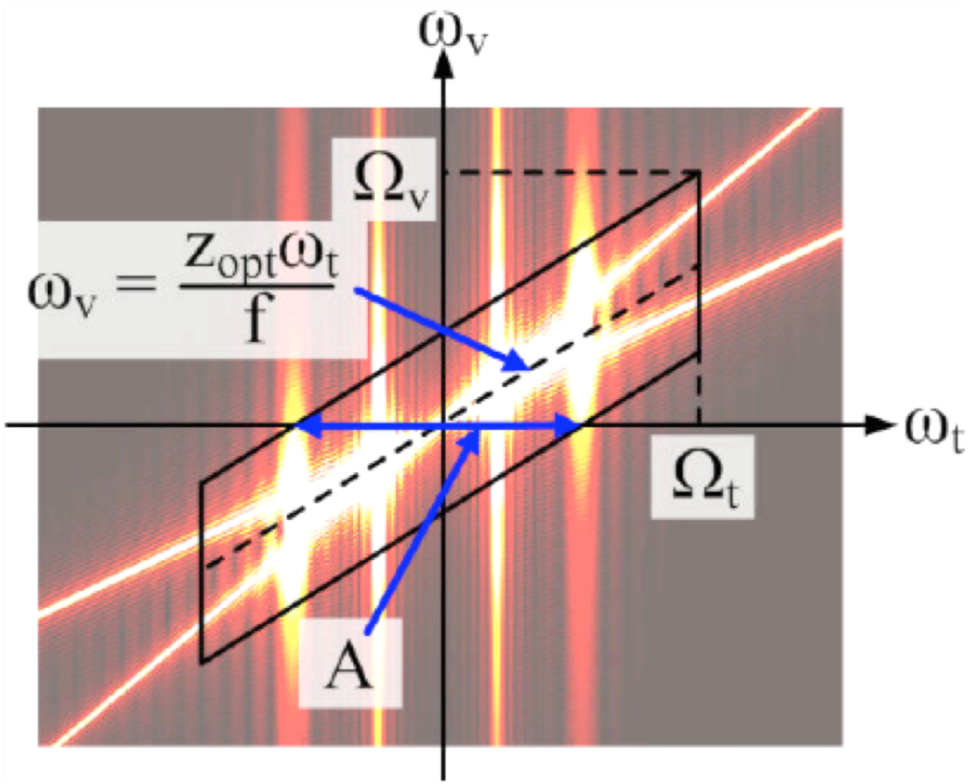


Slanted Plane Plenoptic Spectrum [GilliamDB:10]



- Lines *a=0* and *b=0* are related to the maximum and minimum depth
- Line *c-1=0* is related to the minimum projected frequency captured at $-v_m$
- Line *c+1=0* is related to the maximum projected frequency captured at $+v_m$

Essential Bandwidth of a Slanted Plane



$$\Omega_t = \frac{\omega_s f}{f\cos(\phi) - v_m\,|\sin(\phi)|} + \frac{2\pi}{T},$$

$$\Omega_v = \Omega_t \frac{z_{max}}{f} + \frac{\pi}{v_m},$$

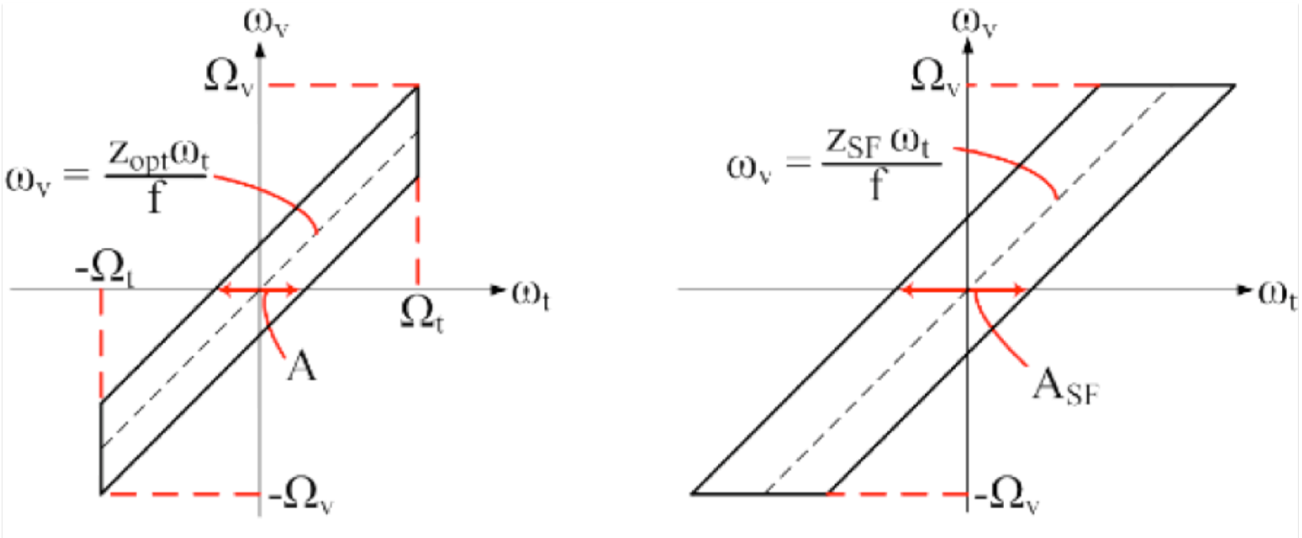
$$z_{opt} = \frac{z_{max} + z_{min}}{2},$$

$$A = \frac{T\,|\sin(\phi)|\,\Omega_t}{z_{opt}} + \frac{2\pi f}{z_{opt} v_m}$$

Imperial College London

Plenoptic Sampling of a Slanted Plane

Comparison between the essential bandwidth and the standard filter



(a) Essential Bandwidth *(b) Standard Filter*

$$z_{opt} = \frac{z_{max} + z_{min}}{2}$$

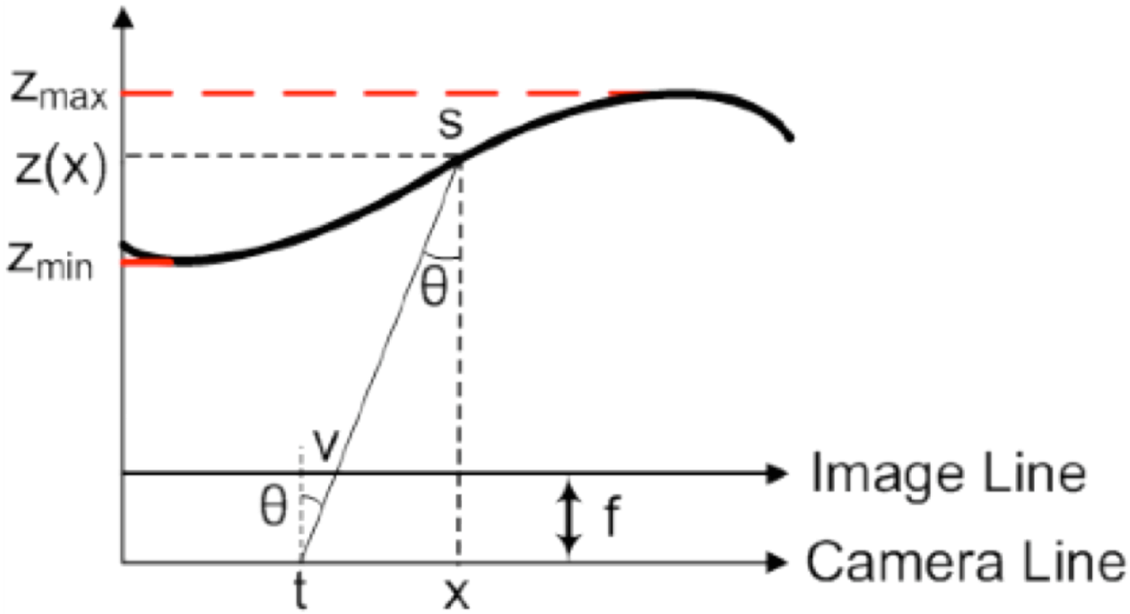
$$z_{SF} = \frac{2}{z_{max}^{-1} + z_{min}^{-1}}$$

$$A = \frac{T\,|\sin(\phi)|\,\Omega_t}{z_{opt}} + \frac{2\pi f}{z_{opt} v_m}$$

$$A_{SF} = \Omega_v f \left(\frac{1}{z_{min}} - \frac{1}{z_{max}} \right)$$

Imperial College London

Adaptive Plenoptic Sampling [GilliamDB:11]



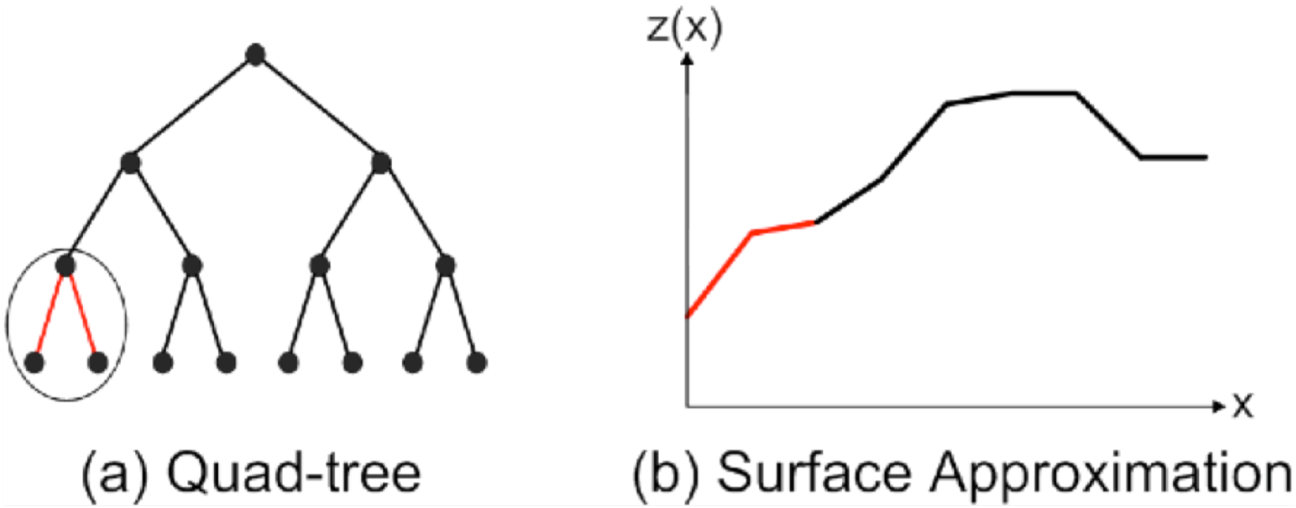
•Finite field of view implies local sampling

•Use the previous theory to sample a smoothly varying surface with bandlimited texture

•Approximate the surface using a set of slanted plane

•Use a pruned quadtree decomposition to find the approximation with the best trade-off between distortion due to aliasing and geometrical error, and number of samples.

Adaptive Plenoptic Sampling (cont'd)

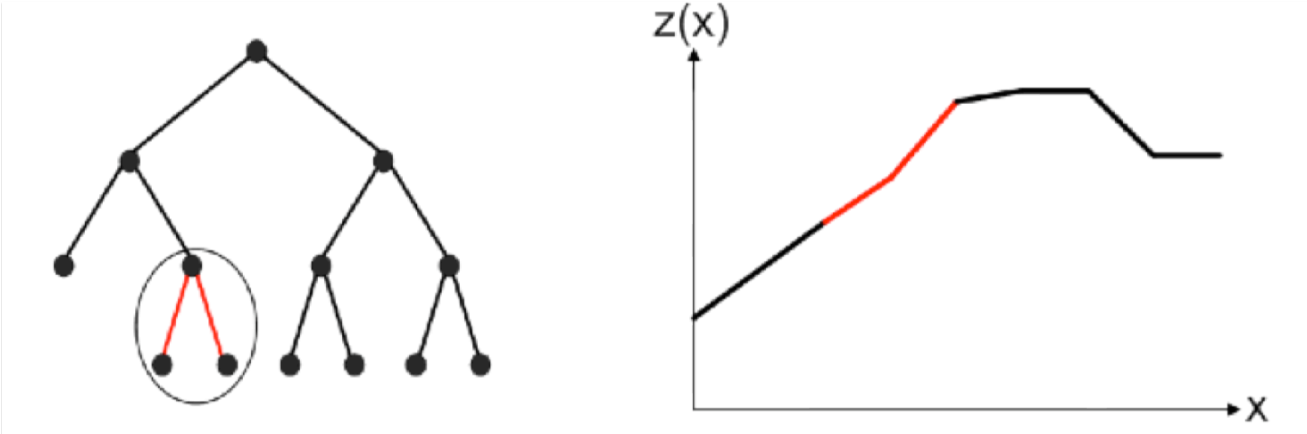


•Approximate the surface using a set of slanted plane
•Use a quadtree decomposition to find the approximation with the best trade-off between distortion due to aliasing and geometrical error, and number of samples.

.

Imperial College
London

Adaptive Plenoptic Sampling (cont'd)

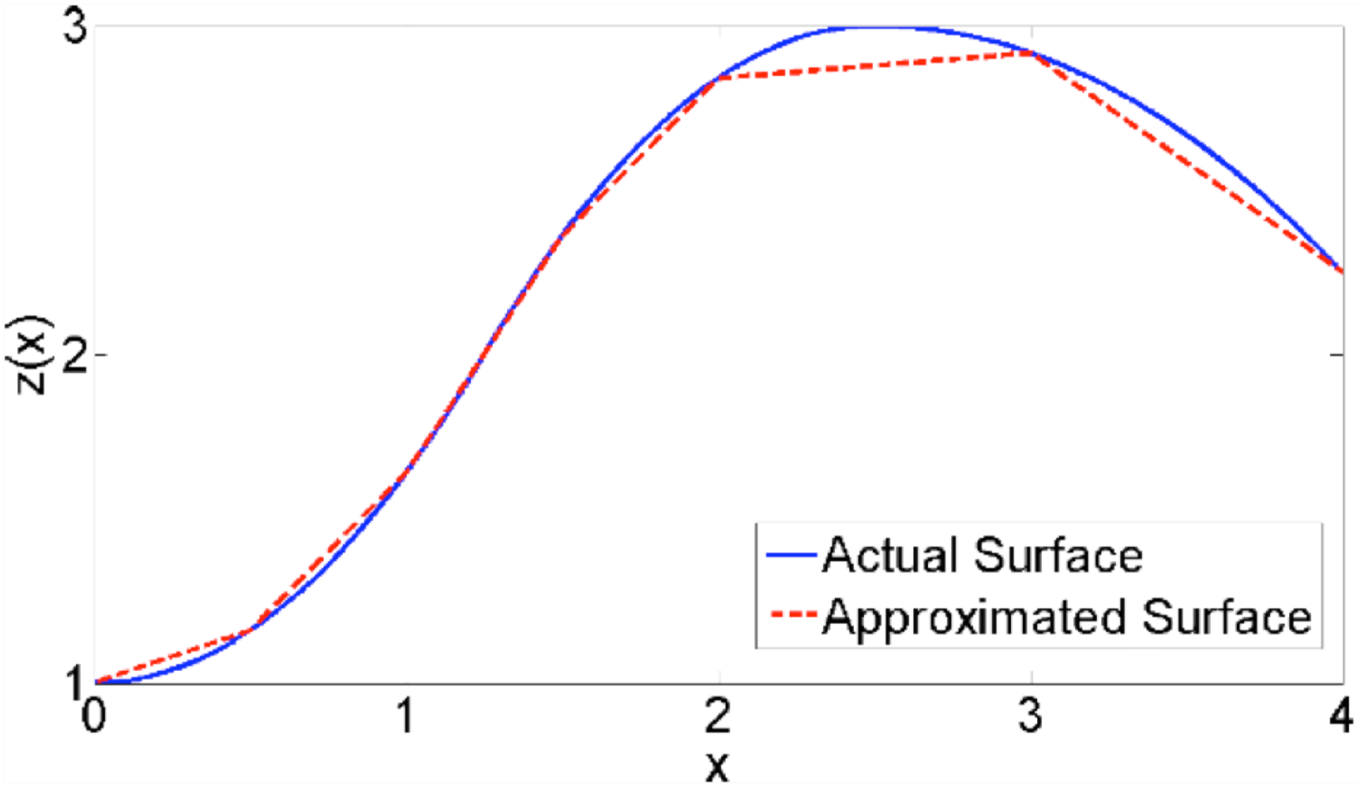


(a) Quad-tree (b) Surface Approximation

•Approximate the surface using a set of slanted plane
•Use a quadtree decomposition to find the approximation with the best trade-off between distortion due to aliasing and geometrical error, and number of samples.

Numerical Results

Approximation of a piecewise quadratic surface using N=20 samples

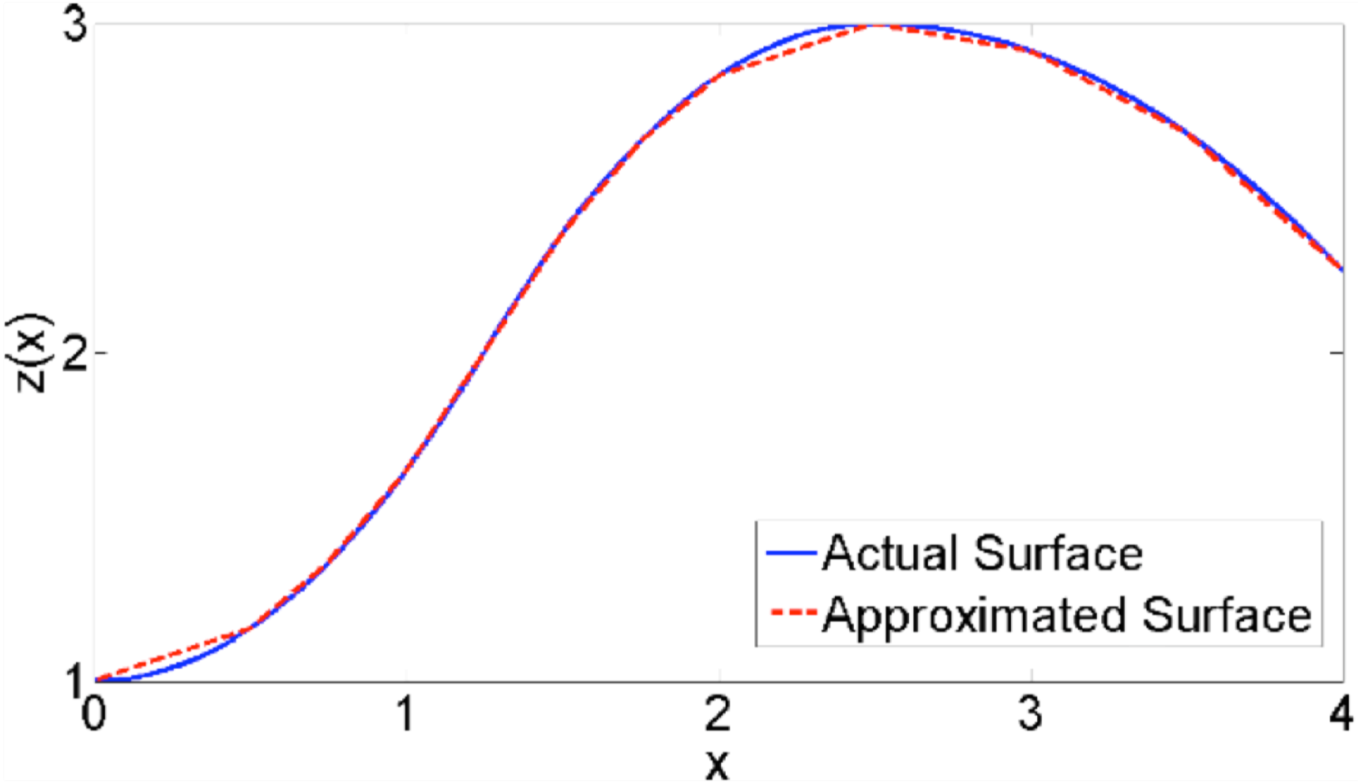


Initial number of planes=16, final number of planes=6

Imperial College London

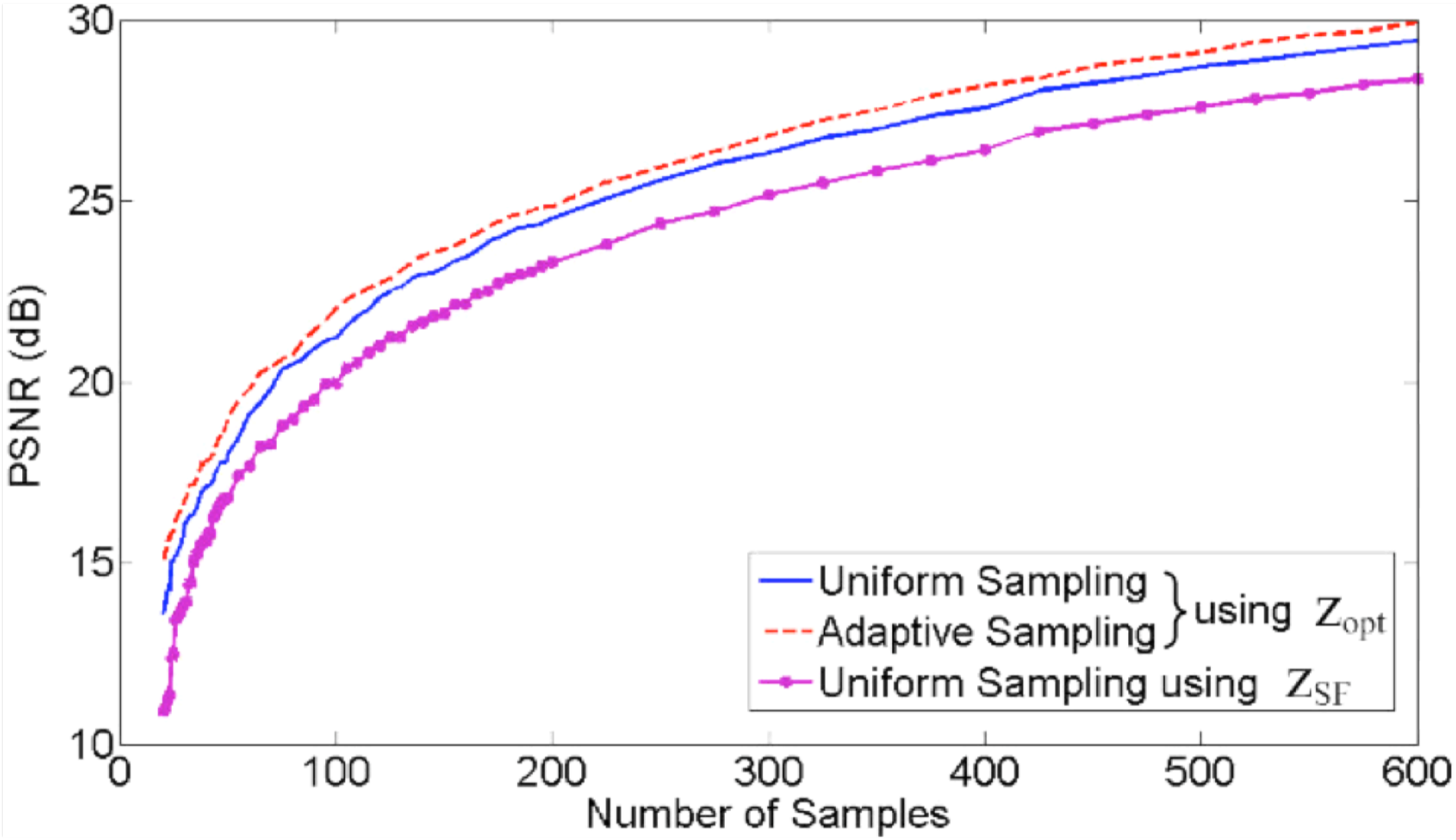
Numerical Results

Approximation of a piecewise quadratic surface using N=300 samples



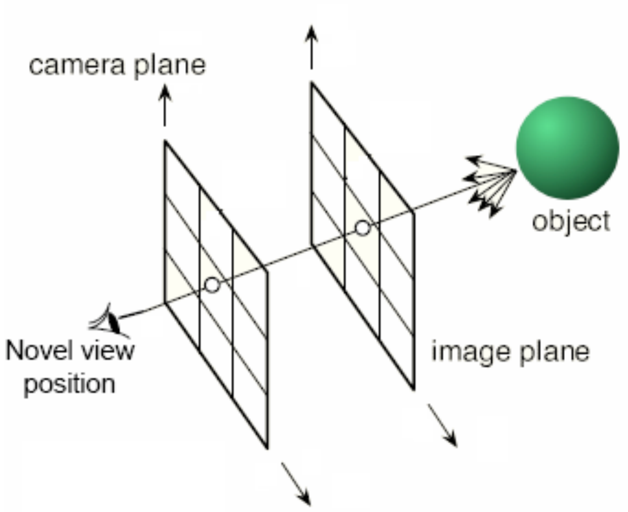
Initial number of planes=16, final number of planes=10

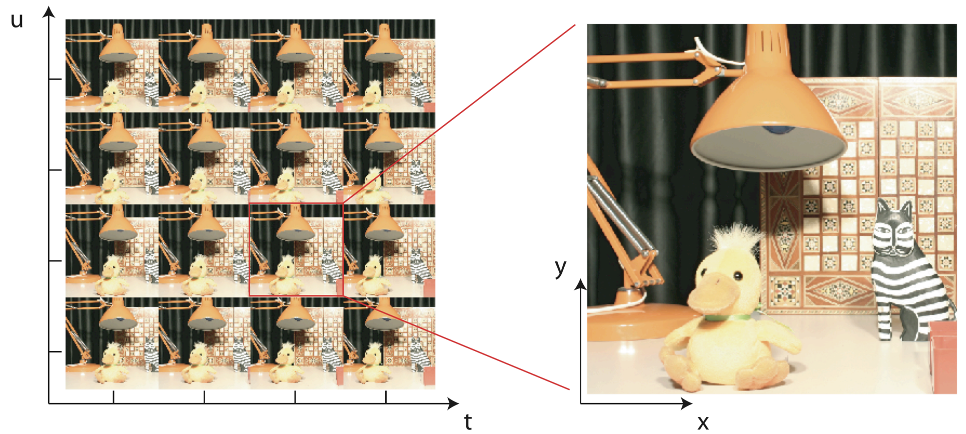
Numerical Results



Imperial College
London

The Light Field





Linear Interpolation of a Sparse Light Field

$d_m = 0$

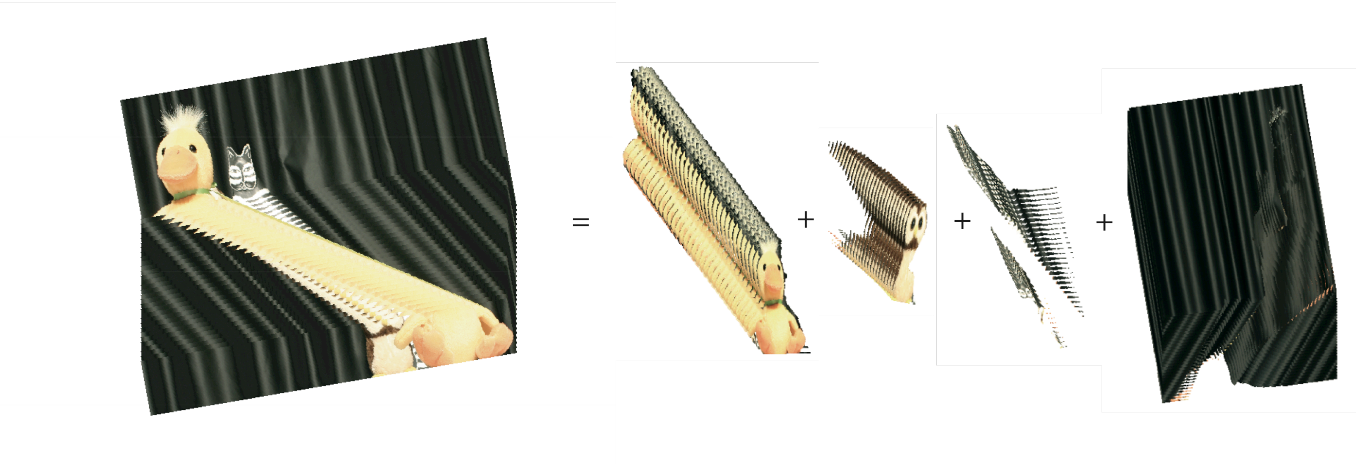
$d_m = d_{min}$

$d_m = d_{average}$

$d_m = d_{max}$

Imperial College London

Depth Layer Extraction

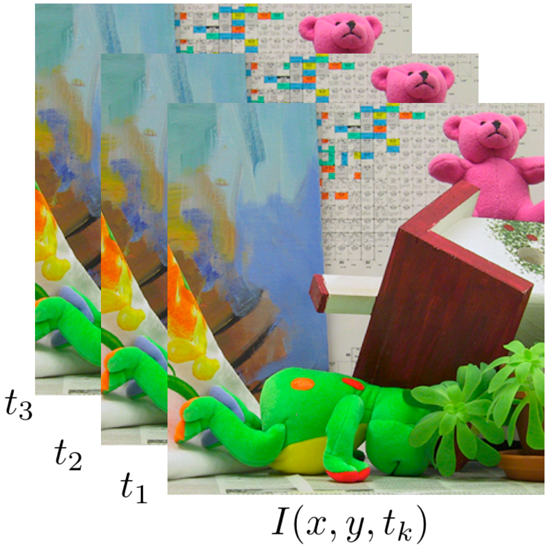


- Decompose the EPI into layers with similar depths.
- Anticipate occlusion ordering in order to make the layers independent

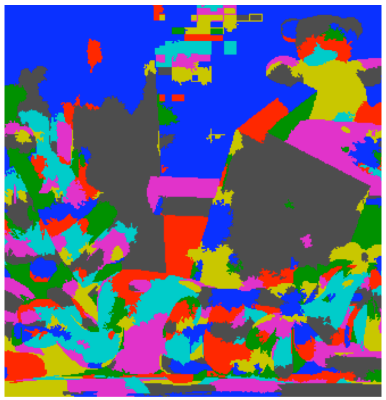
Imperial College London

Unsupervised Layer Extraction [BerentDB:09, PearsonDB:11]

1. Input images (3 in this example):



2. Color segmentation of reference image (Mean-Shift):



Layer boundaries usually occur at color changes

Set of patches S_n

3. Choose number of depths (i.e. layers): $(d_1, d_2, \ldots, d_M)$

4. Assign each patch to a layer using a matching function: $E_n(m) = \sum_{\mathbf{p} \in S_n} f(\mathbf{p}, m)$

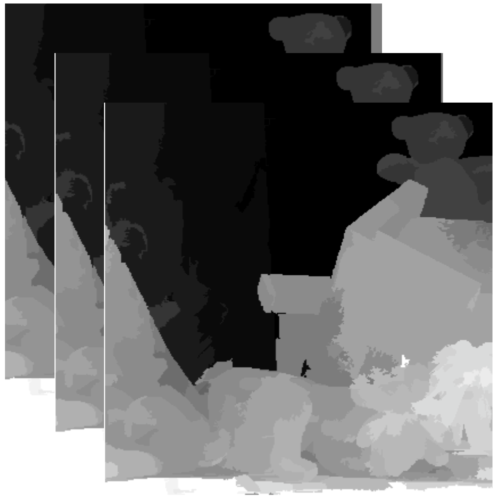
$$\mathbf{p}_{m,k} = (x_{\mathbf{p}} - d_m t_k, y_{\mathbf{p}}, t_k) \qquad f(\mathbf{p}, m) = \sum_{k=1}^{K-1} |I(\mathbf{p}_{m,k}) - I(\mathbf{p}_{m,k+1})|$$

Imperial College London

Unsupervised Layer Extraction

5. Generate layers:

$$L(\mathbf{p}_{m,k}) = m_n \text{ for } \mathbf{p} \in S_n, k \in [1, K]$$



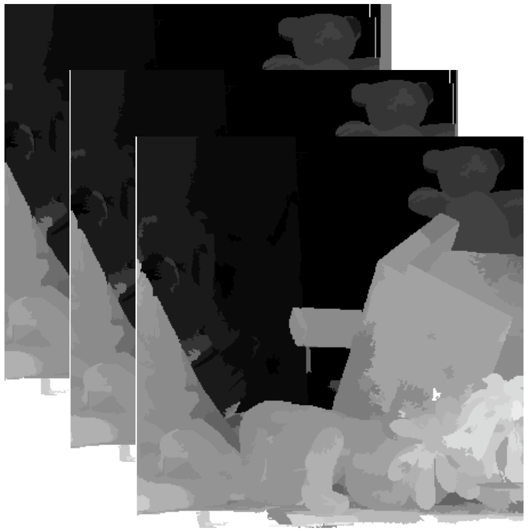
6. Run matching with occlusions:

Visibility function for each pixel on a layer:

$$V(\mathbf{p}, m, k) = \begin{cases} 1, & d_{L(\mathbf{p}_{m,k})} < d_m \text{ or if } L(\mathbf{p}_{m,k}) = m_n \text{ for } \mathbf{p} \in S_n \\ 0, & \text{otherwise.} \end{cases}$$

New matching functional:

$$f(\mathbf{p}, m) = \frac{\sum_{k=1}^{K-1} |I(\mathbf{p}_{m,k}) - I(\mathbf{p}_{m,k+1})| V(\mathbf{p}, m, k) V(\mathbf{p}, m, k+1)}{\sum_{k=1}^{K-1} V(\mathbf{p}, m, k) V(\mathbf{p}, m, k+1)},$$

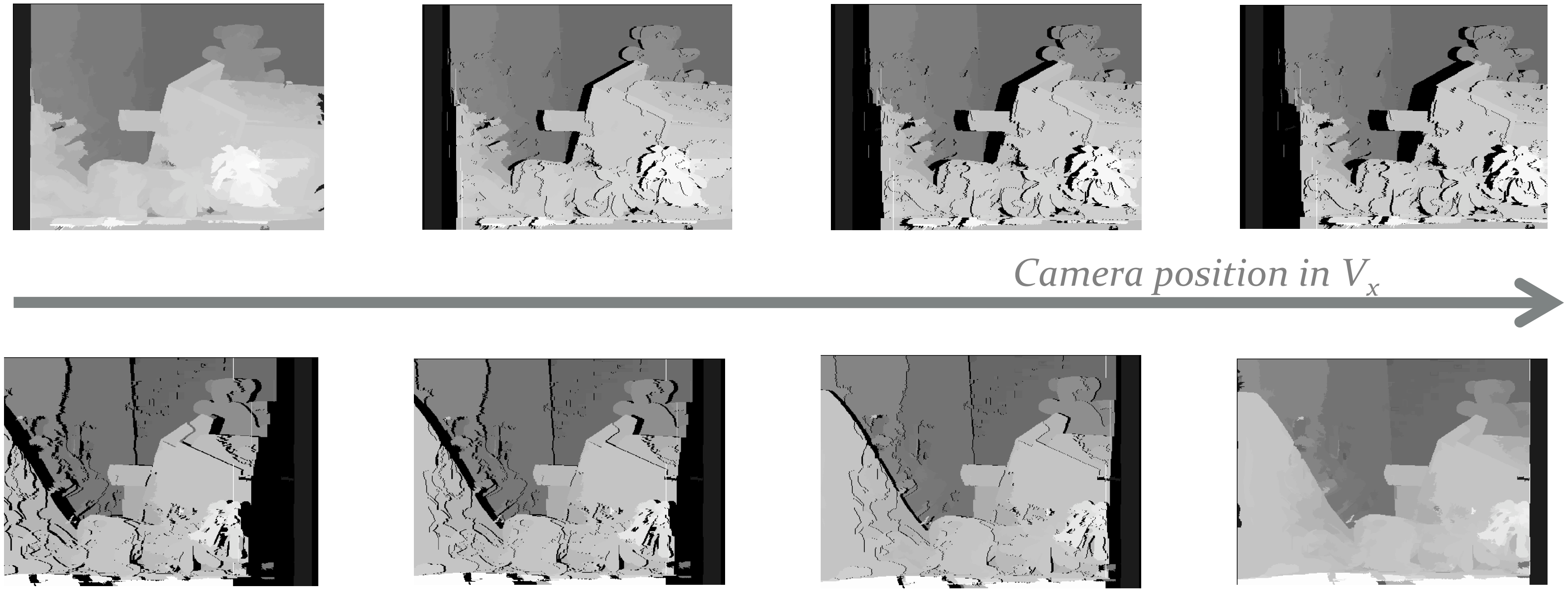


Imperial College
London

Two Key Frames

- Some types of dis-occlusions are inevitable with one key image and a complex scene
- By taking two key images from opposite ends the parallax between them is maximised
- Dis-occlusions in one direction are often covered from the other

Rendered from LHS



Rendered from RHS

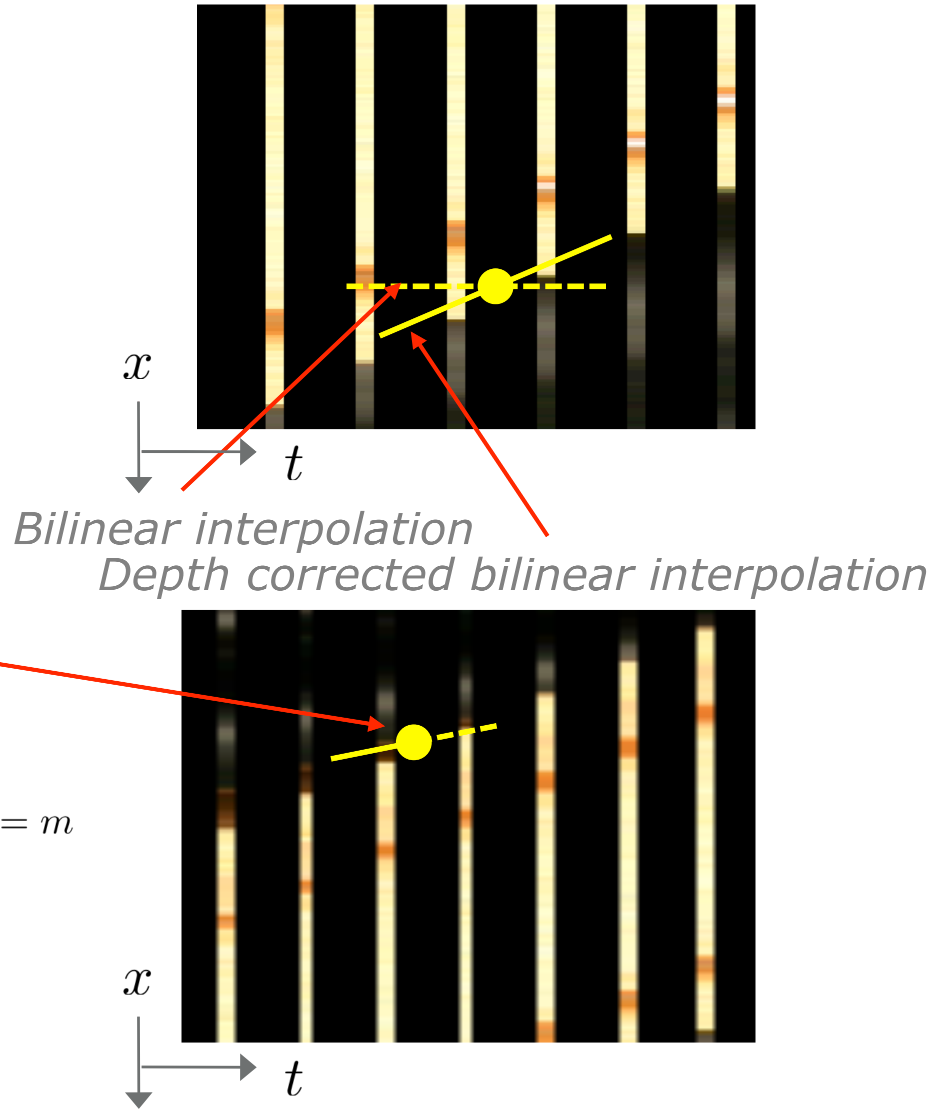
Plenoptic Layer Interpolation

- Build layers for the view:

$$L(\mathbf{p}_{m,i}) = m_n \text{ for } \mathbf{p} \in S_n$$

- Use linear interpolation with a skewed filter according to the depth of the layer
- Use nearest neighbor if the point is occluded in one of the sample images

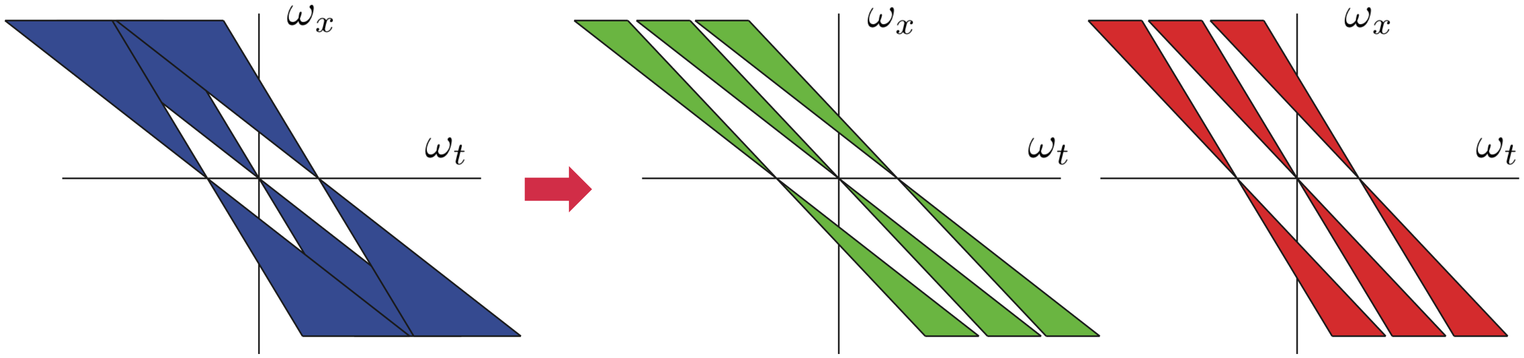
$$I(\mathbf{p}_{m,i}) = \begin{cases} \beta I(\mathbf{p}_{m,k}) + \alpha I(\mathbf{p}_{m,k+1}), L(\mathbf{p}_{m,k}) = L(\mathbf{p}_{m,k+1}) = m \\ I(\mathbf{p}_{m,k}), L(\mathbf{p}_{m,k}) = m, L(\mathbf{p}_{n,k+1}) \neq m \\ I(\mathbf{p}_{m,k+1}), L(\mathbf{p}_{m,k}) \neq m, L(\mathbf{p}_{m,k+1}) = m, \end{cases}$$



How Many Layers?

- We can use layers to reduce aliasing but how many should we use given a set of images?
- Layers slice the spectra into pieces
- Each layer has smaller depth variation (i.e. a tighter bow-tie)
- Individually render each layer and blend them to create synthesized view

$$\frac{\Delta t}{M} = \frac{1}{Bf[1/z_{min} - 1/z_{max}]}$$



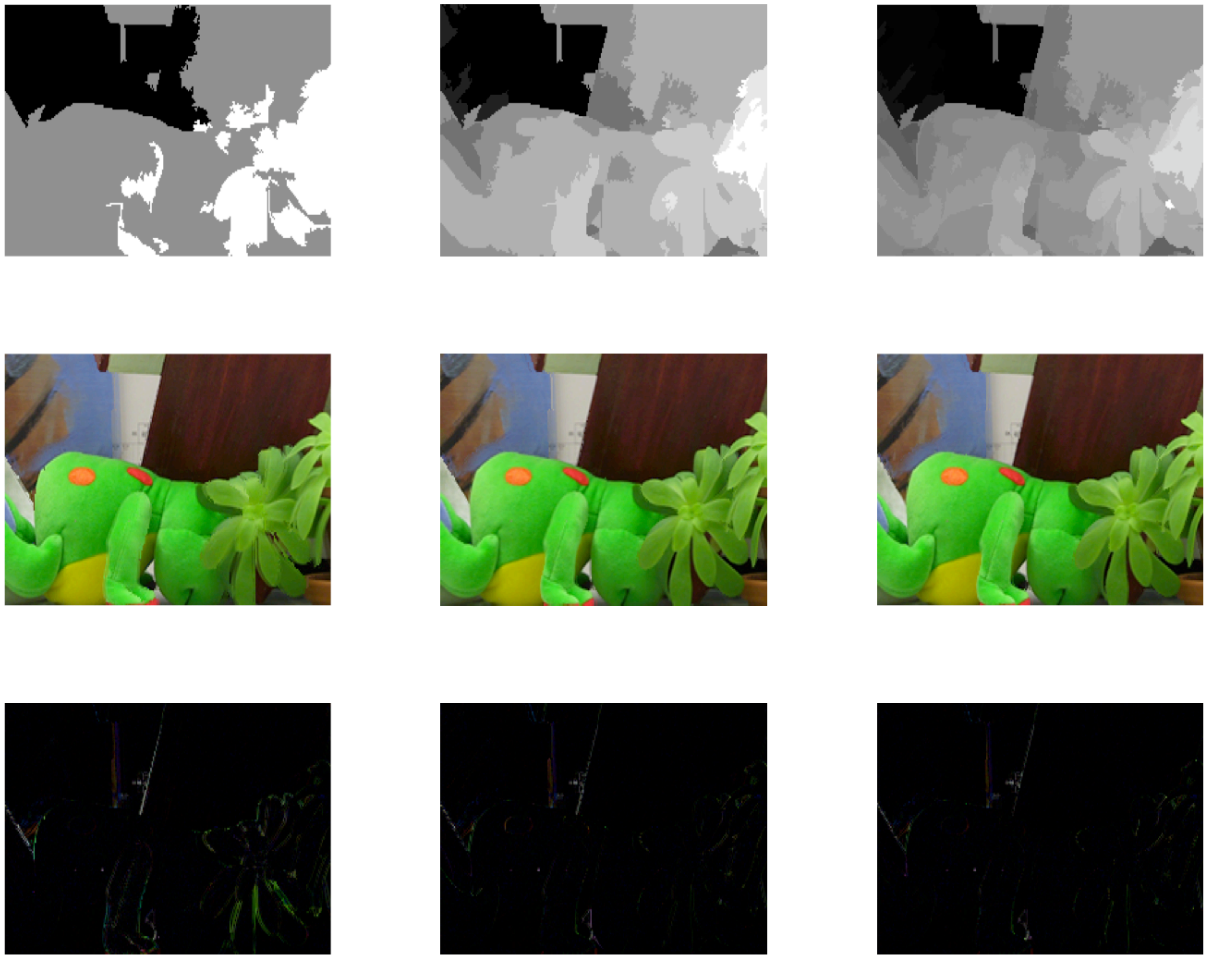
Aliasing so we need more layers *Layer 1* *Layer 2*

IBR Results: Rendering Quality versus Layers

Layers:

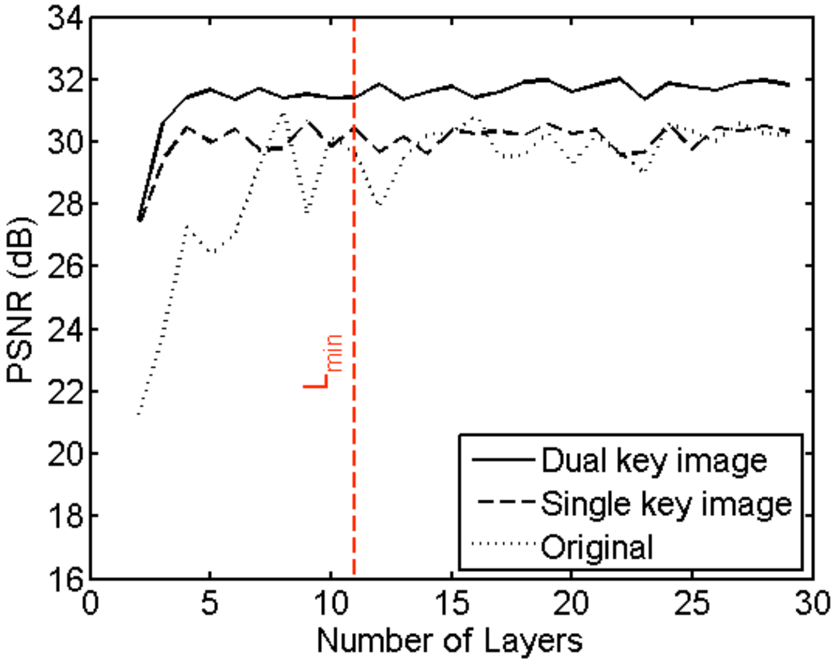
Synthesized view:

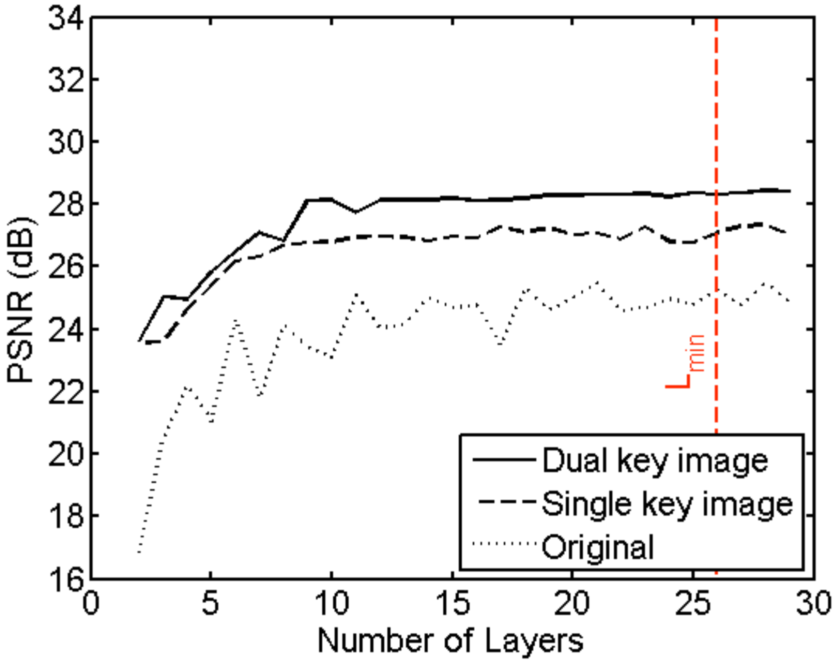
Error:



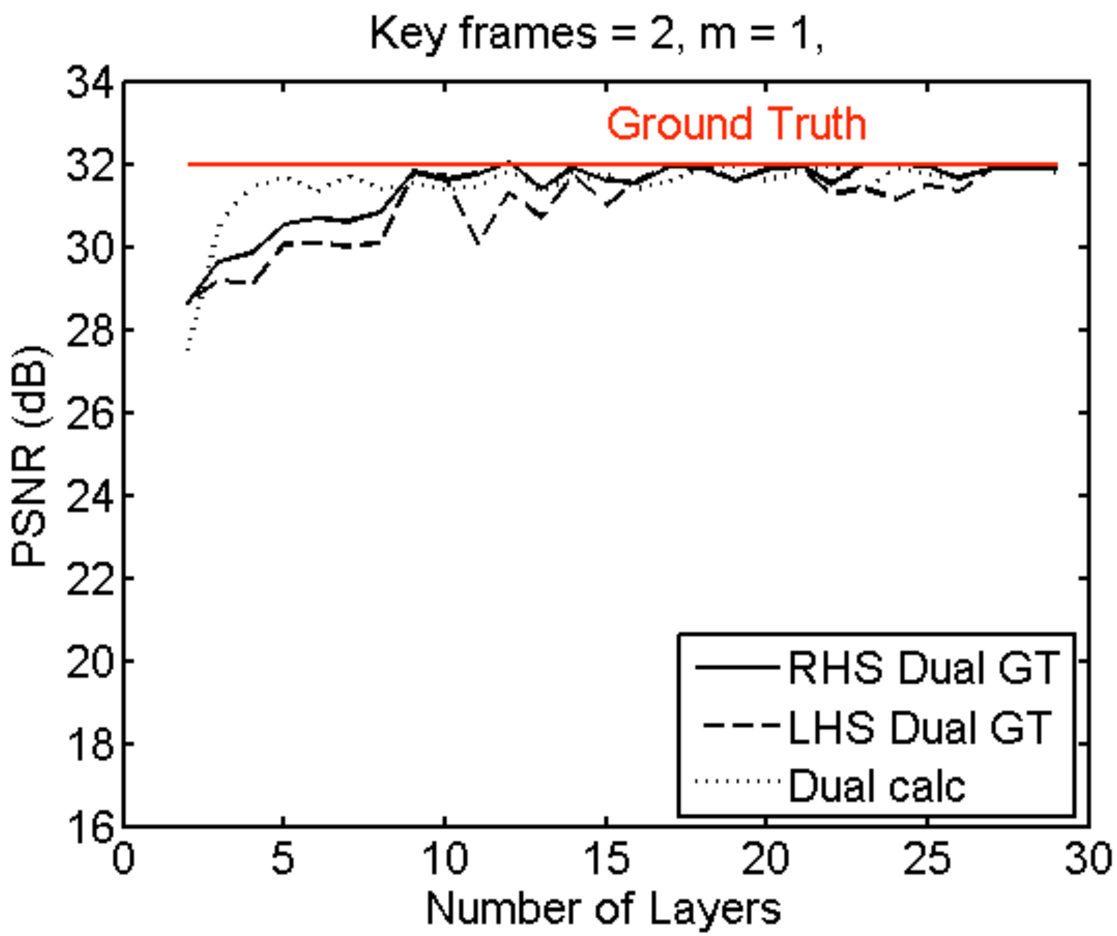
M=3 layers	M=11 layers	M=30 layers
SNR 23.49 dB	SNR 27.02 dB	SNR 27.45 dB

IBR Results: Rendering Quality versus Layers





IBR Results: Rendering Quality versus Geometry



Adaptive Layer Extraction

- Adaptive rendering system: can use the interplay between complexity - rendering quality - number of images
- Interpolation algorithm can be made to automatically estimate the amount of geometry required in the images and extract layers accordingly.

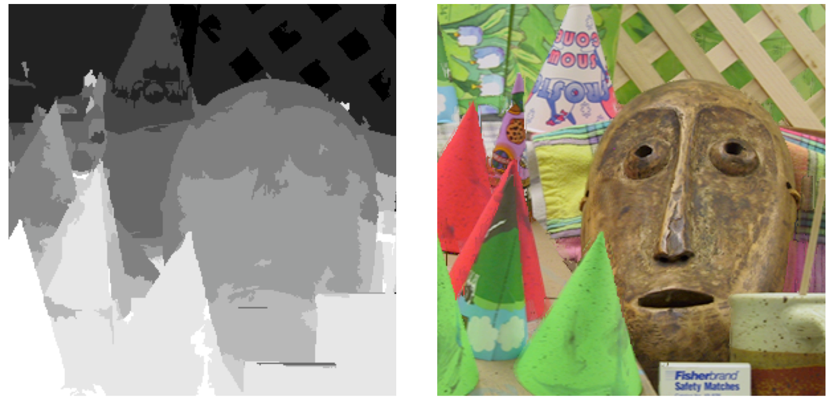
$$M = \frac{\Delta t}{2}(d_{max} - d_{min})$$

Estimated using block matching and outlier rejection

$$d_m = \frac{m - 0.5}{M} d_{max} + (1 - \frac{m - 0.5}{M}) d_{min} \qquad m = 1, \ldots, M$$

Feed dm to the layer extraction algorithm

IBR Results: Non-uniform arrays



M=10 layers, MS: 2.47 sec, LE: 2.48 sec

(Matlab implementation)

M=5 layers, MS: 2.55 sec, LE: 1.7 sec

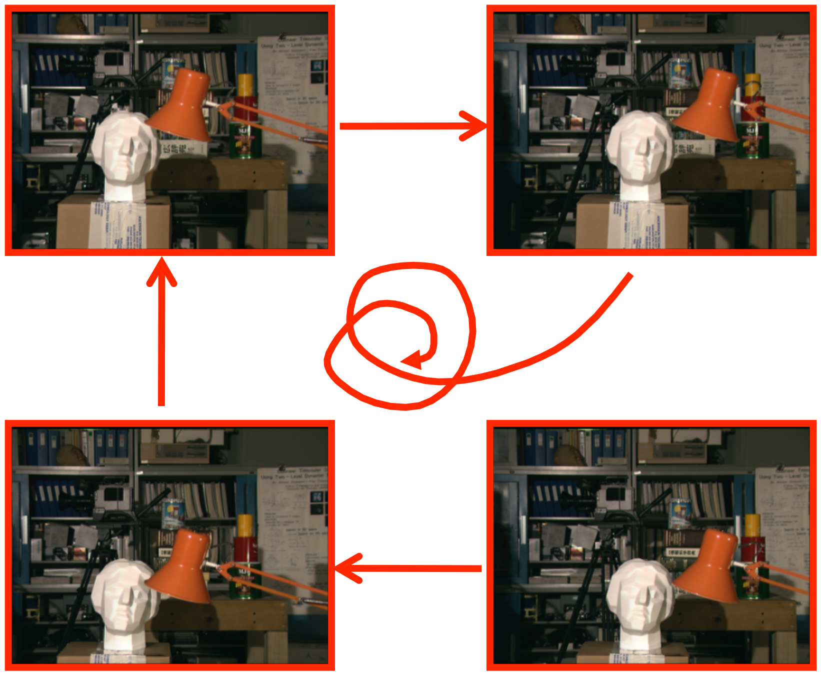
Imperial College London

IBR Results: Adaptive Layer Extraction

5 Images Acquired

25 Images Rendered

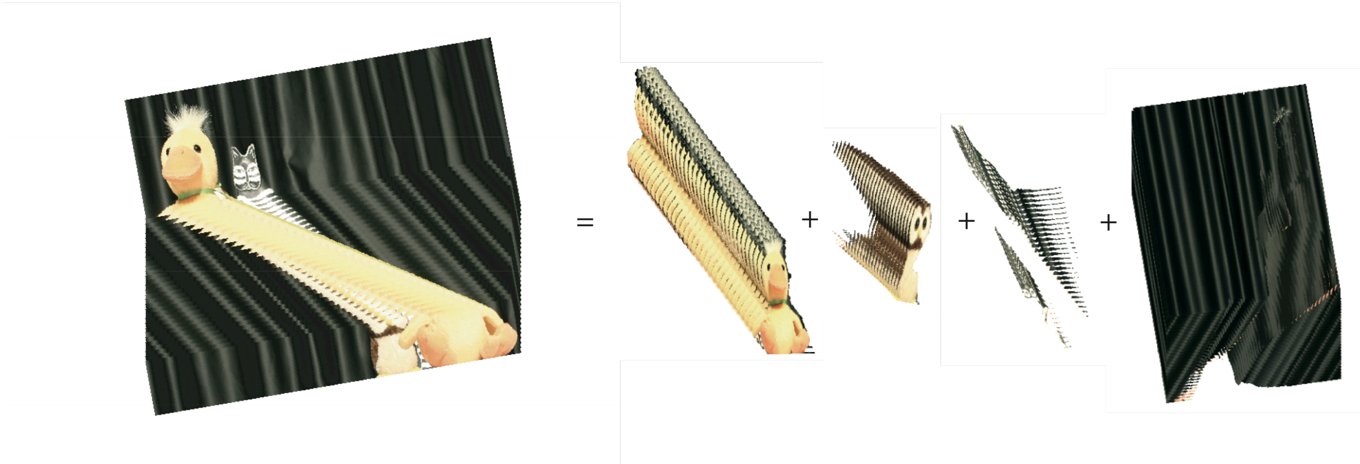
IBR Results on the Lightfield



Demo due to James Pearson (ICL)

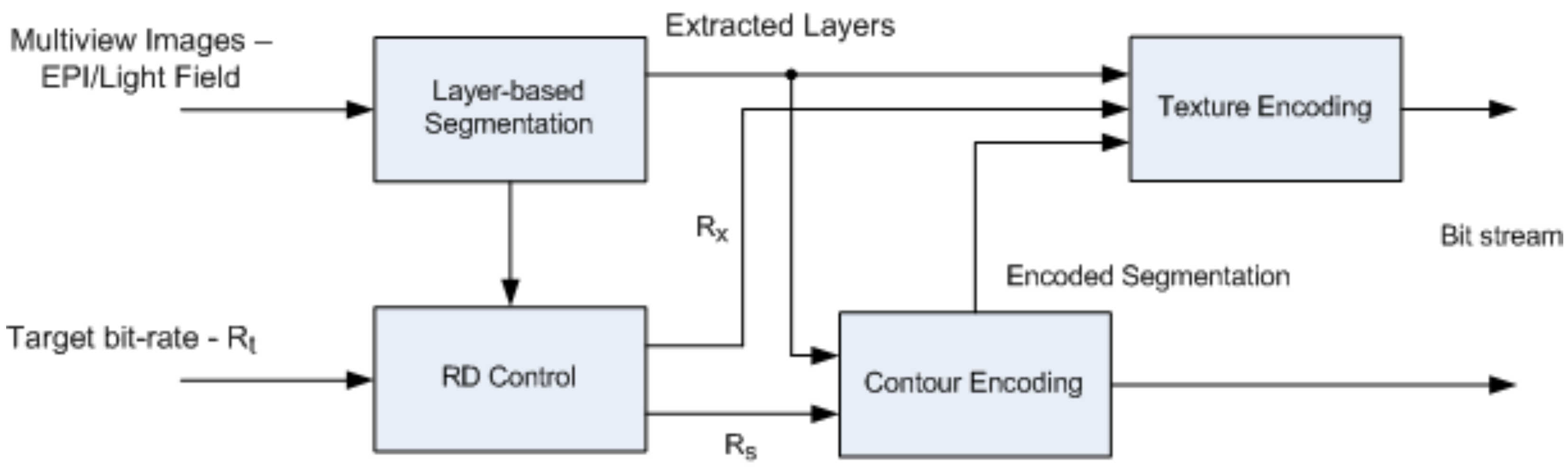
Layer-Based Compression of the Plenoptic Function [GelmanDV: 10]

Exploit the structure of the data in order to maximize compression efficiency:
- *The disparity of each depth layer is constant*
- *Occlusion ordering can be inferred from the layer depth*



Layer-Based Compression of the Plenoptic Function

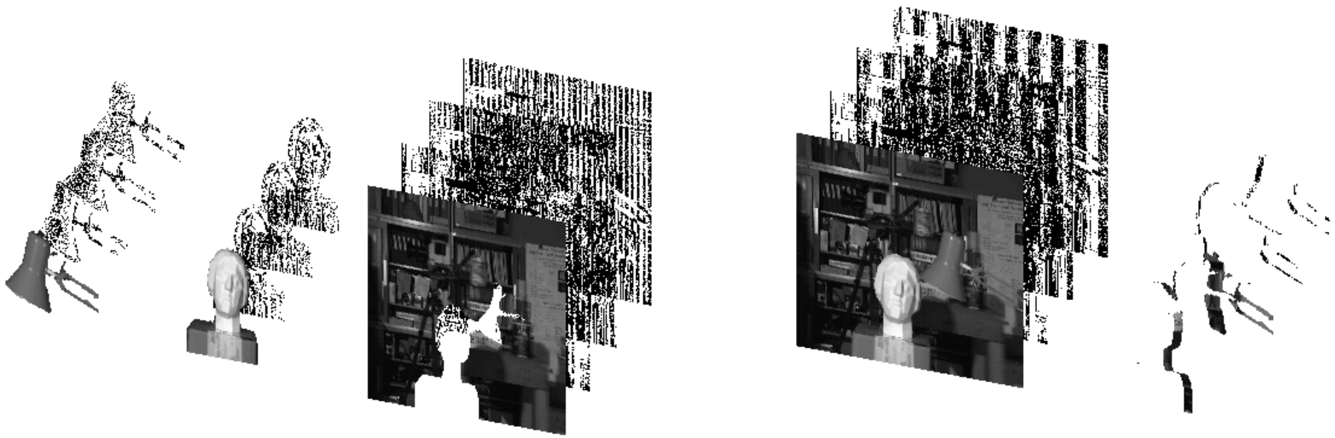
Exploit the structure of the data in order to maximize compression efficiency:
- *The disparity of each depth layer is constant*
- *Occlusion ordering can be inferred from the layer depth*



Imperial College London

Layer-Based Compression of the Plenoptic Function [GelmanDV: 10]

- *Apply a disparity compensated wavelet transform along the view domain*
- *Apply a 2-D WT on the recombined layer after the view-domain transform*
- *Contours of the layers are lossy or lossless compressed according to the bit budget*



(a) (b)

Layer-Based Compression of the Plenoptic Function: Simulation Results

(a) Animal Farm

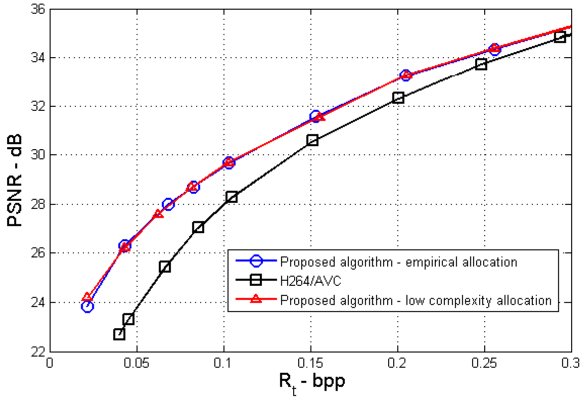
(b) Teddy

(c) Tsukuba EPI

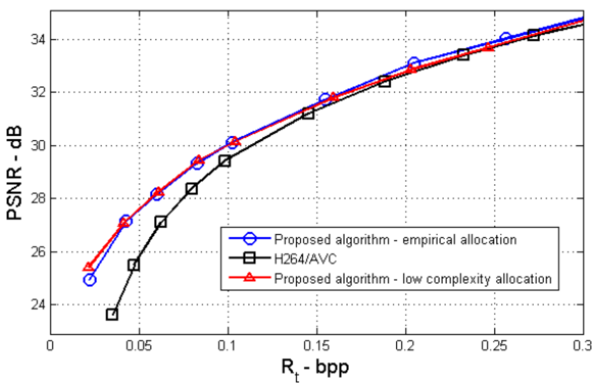
(d) Cones

Imperial College
London

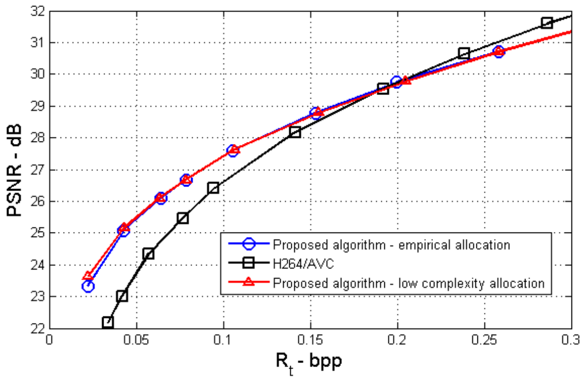
Layer-Based Compression of the Plenoptic Function: Simulation Results



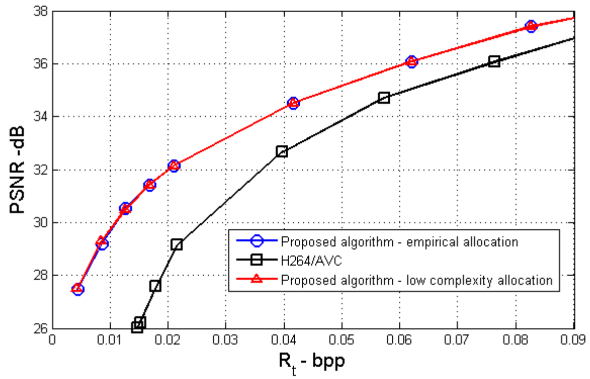
(a) Tsukuba EPI



(b) Teddy



(c) Cones



(d) Animal Farm

Layer-Based Compression of the Plenoptic Function: Simulation Results

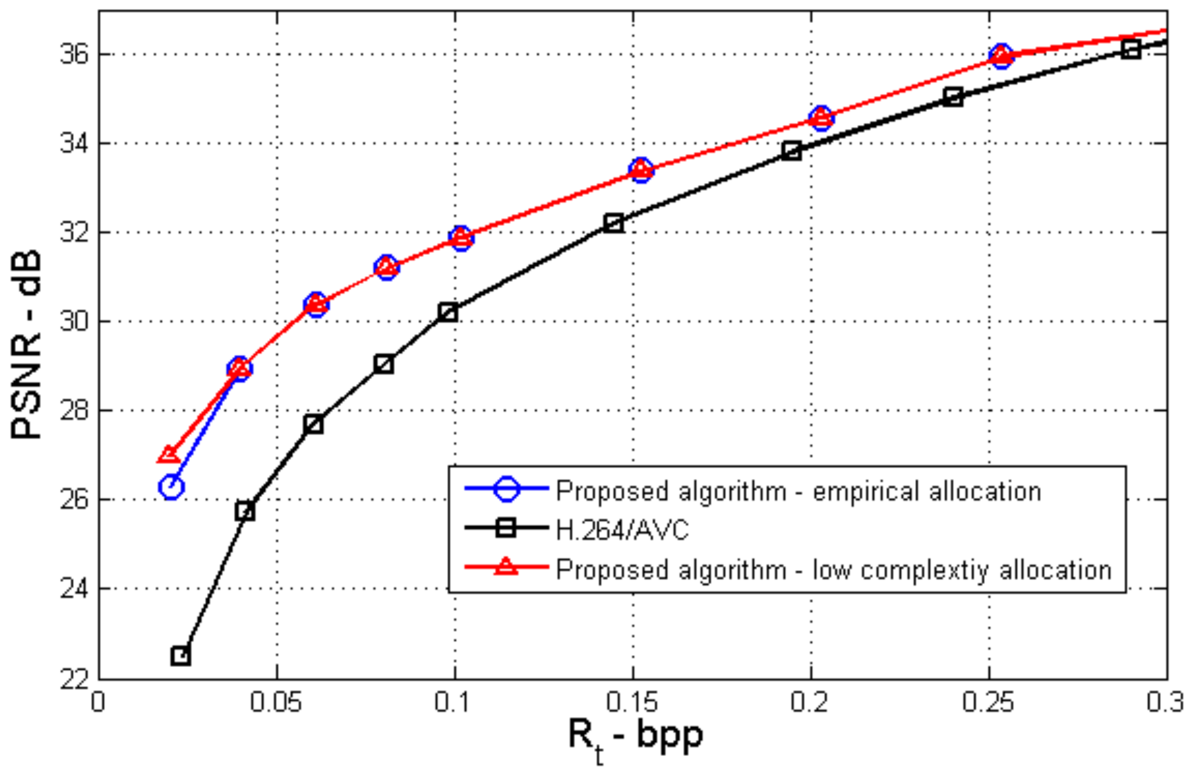
(a)

H.264/AVC

(b)

Layer-based Compression

Layer-Based Compression of the Plenoptic Function: Simulation Results



Tsukuba Light Field

Imperial College London

Layer-Based Compression of the Plenoptic Function: Simulation Results

H.264/AVC (PSNR: 26.9dB, 0.05bpp)

Layer-based Compression (PSNR: 29.8dB, 0.05bpp)

Imperial College London

Conclusions and Outlook

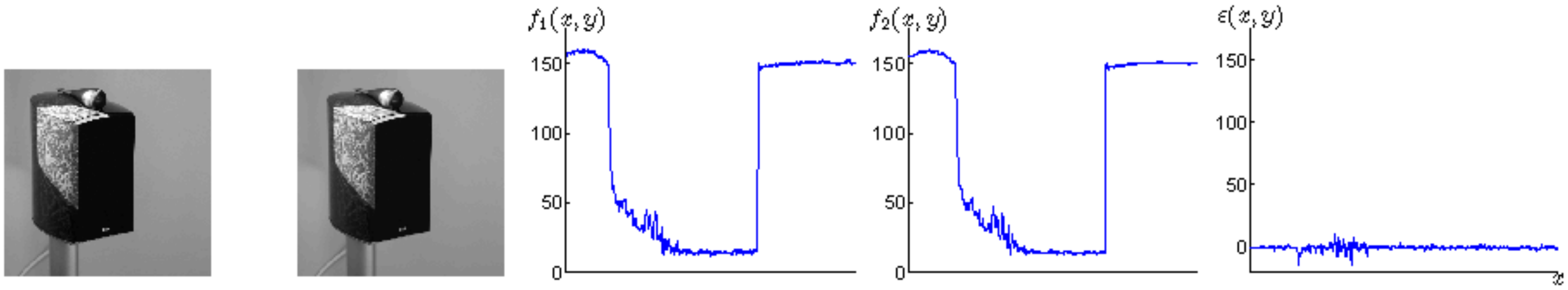
- Image-based rendering is more relevant now than ever
- Plenoptic domain viewpoint is helpful
- Still many open questions from theory to practical implementations
- On sampling:
  - Strong theoretical results only for very limited cases
  - A complete plenoptic sampling theory with piecewise smooth models is still an open challenge.
  - Rendering with depth cameras and cameras with different resolution.
- On compression:
  - Competitive algorithms for joint compression of the lightfield
  - Need to derive methods with the correct trade-off between complexity, efficiency in an R-D sense and with random access capabilities.

Imperial College London

Publications

- On sampling:
  - C. Gilliam, P. L. Dragotti, M. Brookes, A closed-form expression for the bandwidth of the plenoptic function under finite field of view constraints, Proc. of IEEE ICIP, September 2010.
  - C.Gilliam, P.L. Dragotti and M. Brookes, Adaptive Plenoptic Sampling, to appear in Proc. of IEEE ICIP, Brussels, Belgium, September 2011.
- On depth layer extraction:
  - J. Pearson, P.L. Dragotti and M. Brookes, Accurate non-iterative depth layer extraction algorithm for image based rendering, Proc. of IEEE ICASSP, Prague, Czech Republic, May 2011.
  - J. Berent, P.L. Dragotti and M. Brookes, Adaptive Layer Extraction for Image Based Rendering, in Proc. of International Workshop on Multimedia Signal Processing (MMSP), Brazil, October 2009.
  - J. Berent and P.L. Dragotti, Plenoptic Manifolds: Exploiting Structure and Coherence in Multiview Images, IEEE Signal Processing Magazine, vol. 24 (6), pp.34-44, November 2007.
- On compression:
  - A. Gelman, P.L. Dragotti, V. Velisavljevic, Multiview Image compression using layer-based representation, Proc. of IEEE ICIP, September 2010.
  - V. Chaisinthop and P.L. Dragotti, ` Centralized and Distributed Semi-Parametric Compression of Piecewise Smooth Functions' Semi-Parametric Compression of Piecewise-Smooth Functions', accepted for publication in the IEEE Trans. on Signal Processing, January 2011.
  - N.Gehrig and P.L. Dragotti, Geometry-Driven Distributed Compression of the Plenoptic Function: Performance Bounds and Constructive Algorithms, IEEE Trans. on Image Processing, Vol. 18(3), pp. 457-470, March 2009.

Modelling Multi-view Images



(a) (b) (c) (d) (e)

- *Each scan-line of the images is a piecewise smooth signal.*
- *The residual is assumed to be globally smooth.*
- Compress the first signal independently
- Transmit the low-pass coefficients of the second signal together with the less significant bits of the wavelet coefficients.

Modelling Multi-view Images – Simulation Results

(b) distributed semi-parametric compression;

(c) independent SPIHT algorithm.

• *Distributed: PSNR=28.1dB at 0.34bpp*
• Independent: PSNR=26.5dB at 0.31bpp

Distributed Compression of Multi-view Images

- *Each image is compressed using an algorithm based on quadtree decomposition*
- Transmit most of the quad-tree structure from each encoder
- Transmit partial information of the texture in each tiles to deal with non-Lambertian scenes.

Distributed Compression of Multi-view Images

Independent Encoding
PSRN: 28.24 dB

Distributed Encoding
PSNR: 30.02 dB