Imperial College London

Layer-Based Representation for Image Based Rendering and Compression

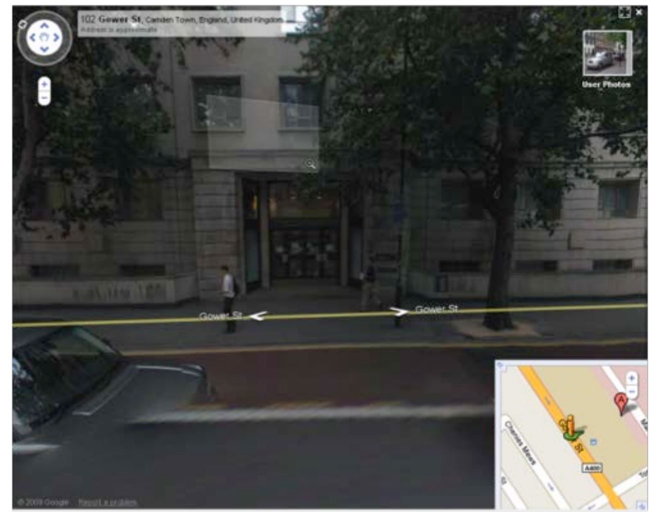
Pier Luigi Dragotti
Imperial College London

Joint work with M. Brookes (ICL), A. Gelman (ICL), J. Onativia (ICL), J. Pearson (ICL).

Imperial College London

The problem: Rendering Novel Views

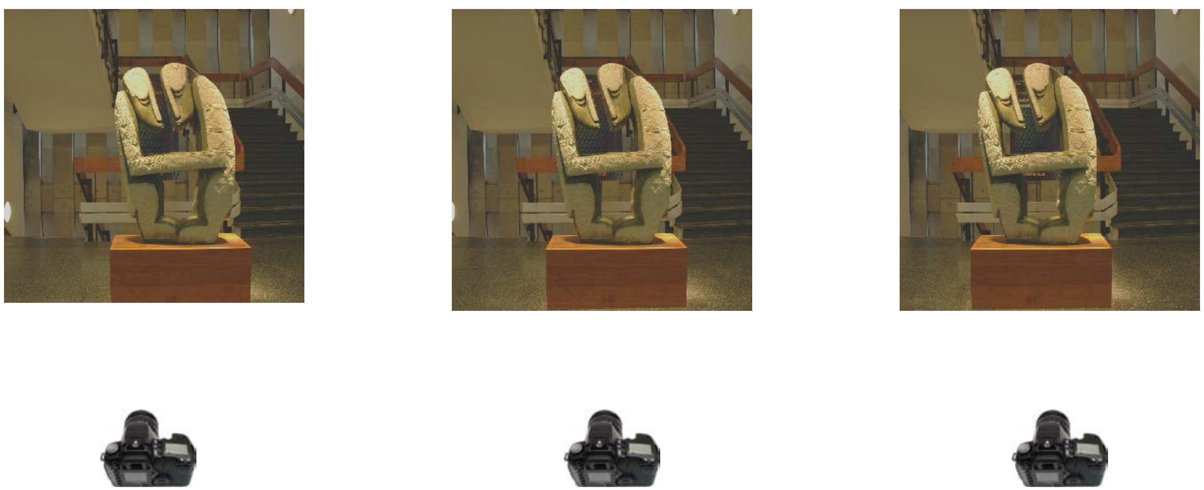
iview BBC Research



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

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



Available images

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
- Many images requires good compression algorithms
- Trade-off between 3-D geometry, number of images and rendering

Fill in the gaps and create a walkthrough environment

Imperial College London

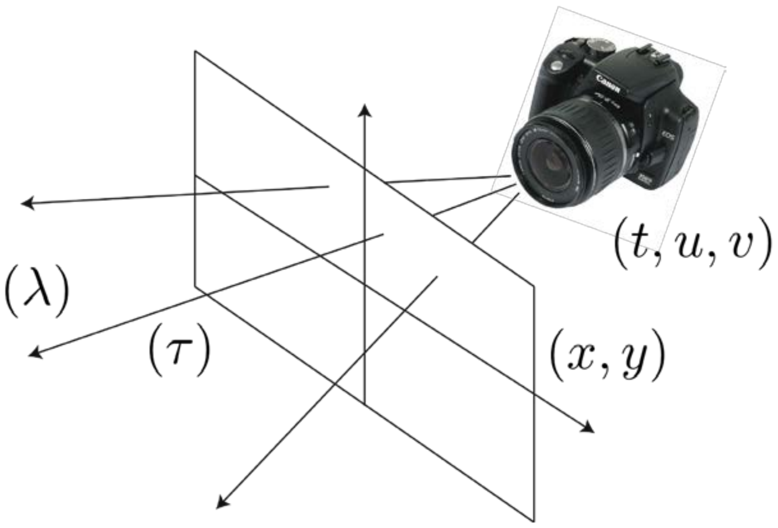
Talk Outline

1. Structure of the data: The plenoptic function, the EPI and the lightfield
2. Layer Based Representation for Compression
3. Results on Compression
4. Layer Based Rendering
5. Geometry Versus Rendering Quality
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

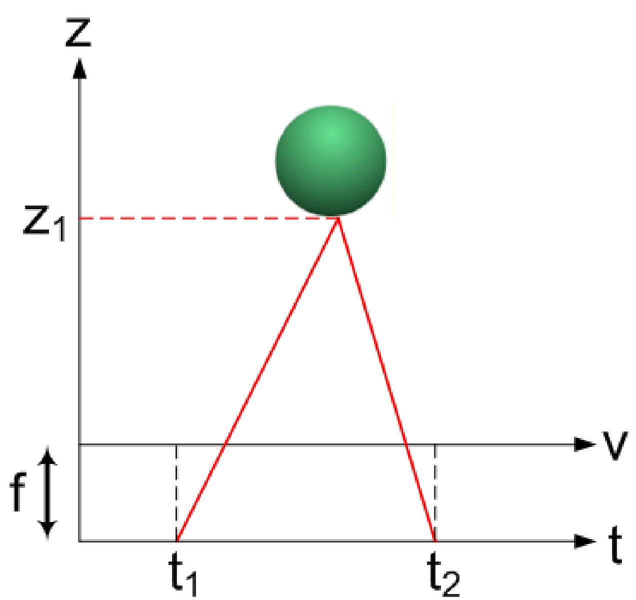
$$I = I(x, y, \lambda, \tau, t, u, v)$$

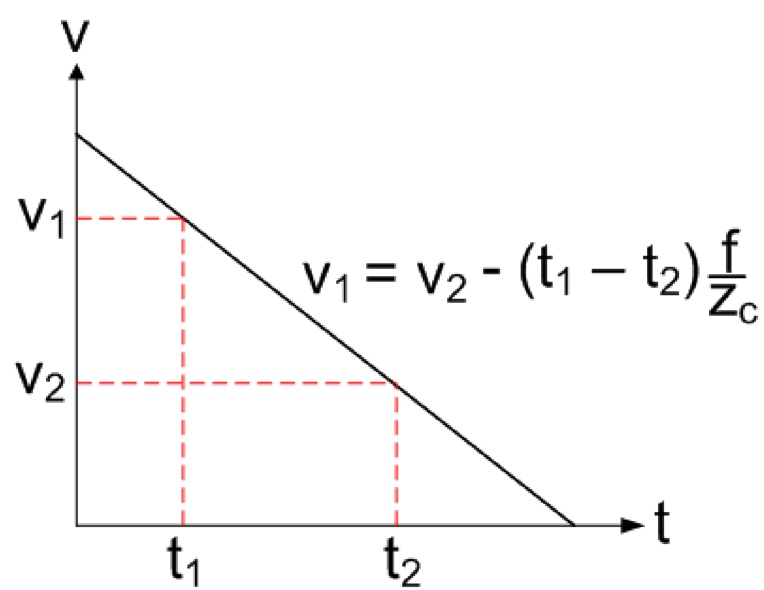


Imperial College London

Epipolar Plane Image (EPI)

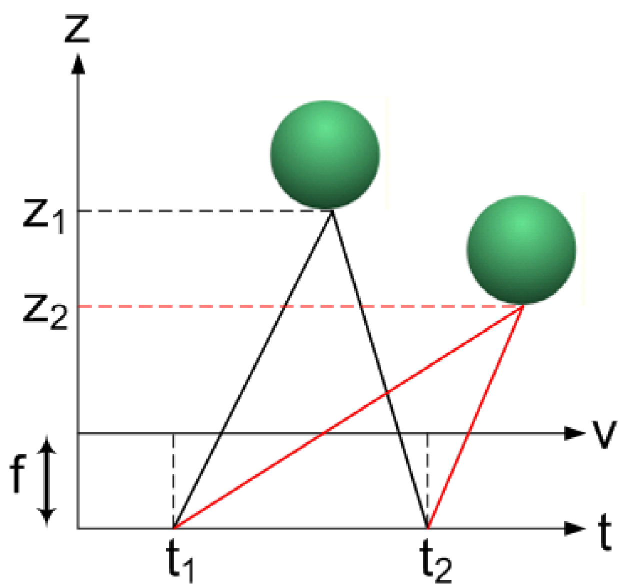
- Pinhole camera model
- Points are mapped onto lines in the (EPI)
- Slope of lines are inversely proportional to the depth
- Lines with larger slopes occlude lines with smaller slopes

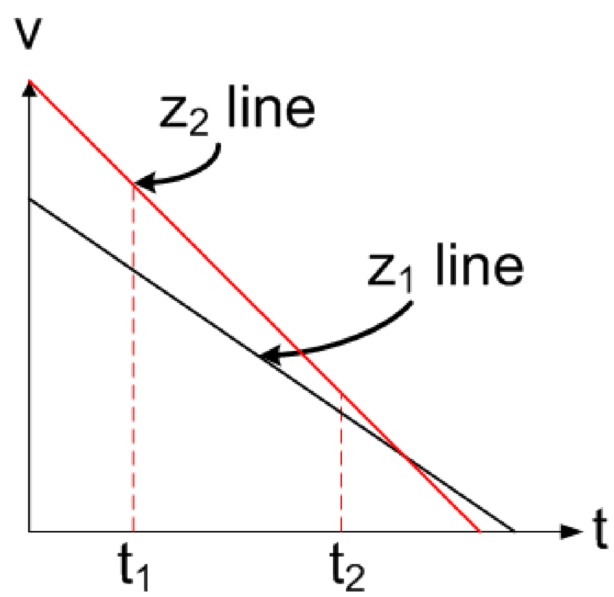




Epipolar Plane Image (EPI)

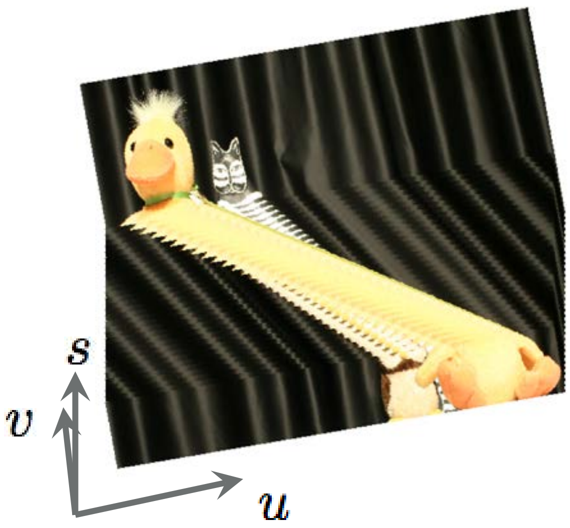
- Pinhole camera model
- Points are mapped onto lines in the (EPI)
- Slope of lines are inversely proportional to the depth
- Lines with larger slopes occlude lines with smaller slopes

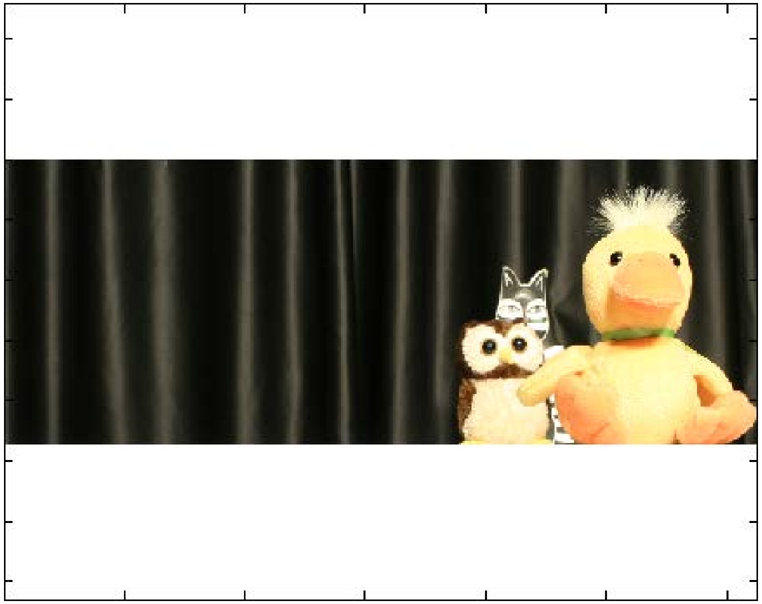




EPI Structure

- 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

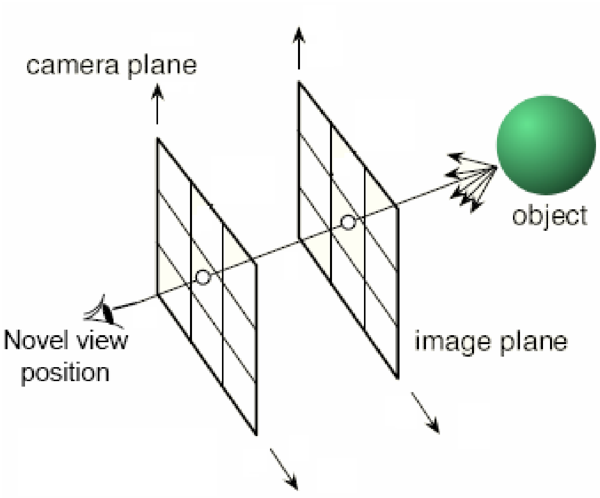


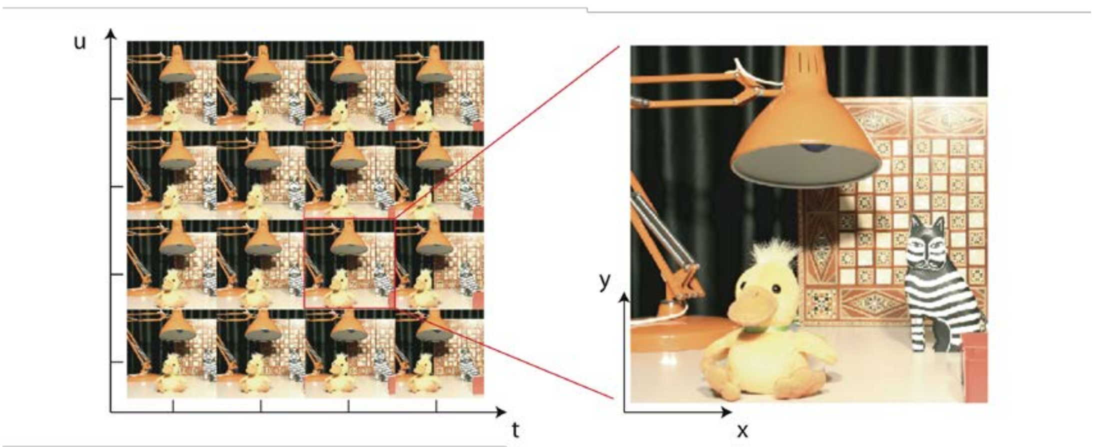


Imperial College London

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



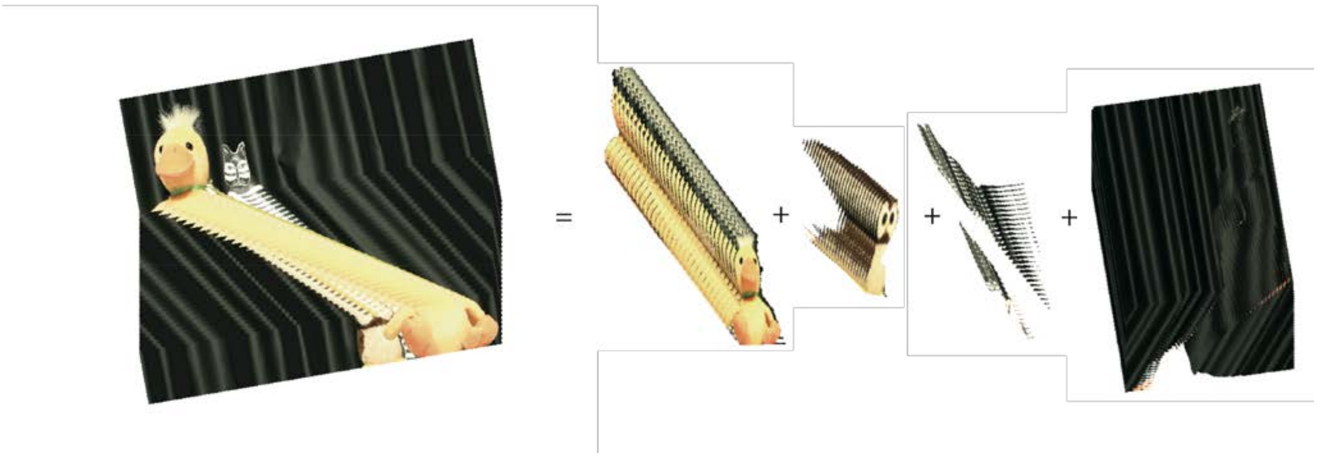


Imperial College London

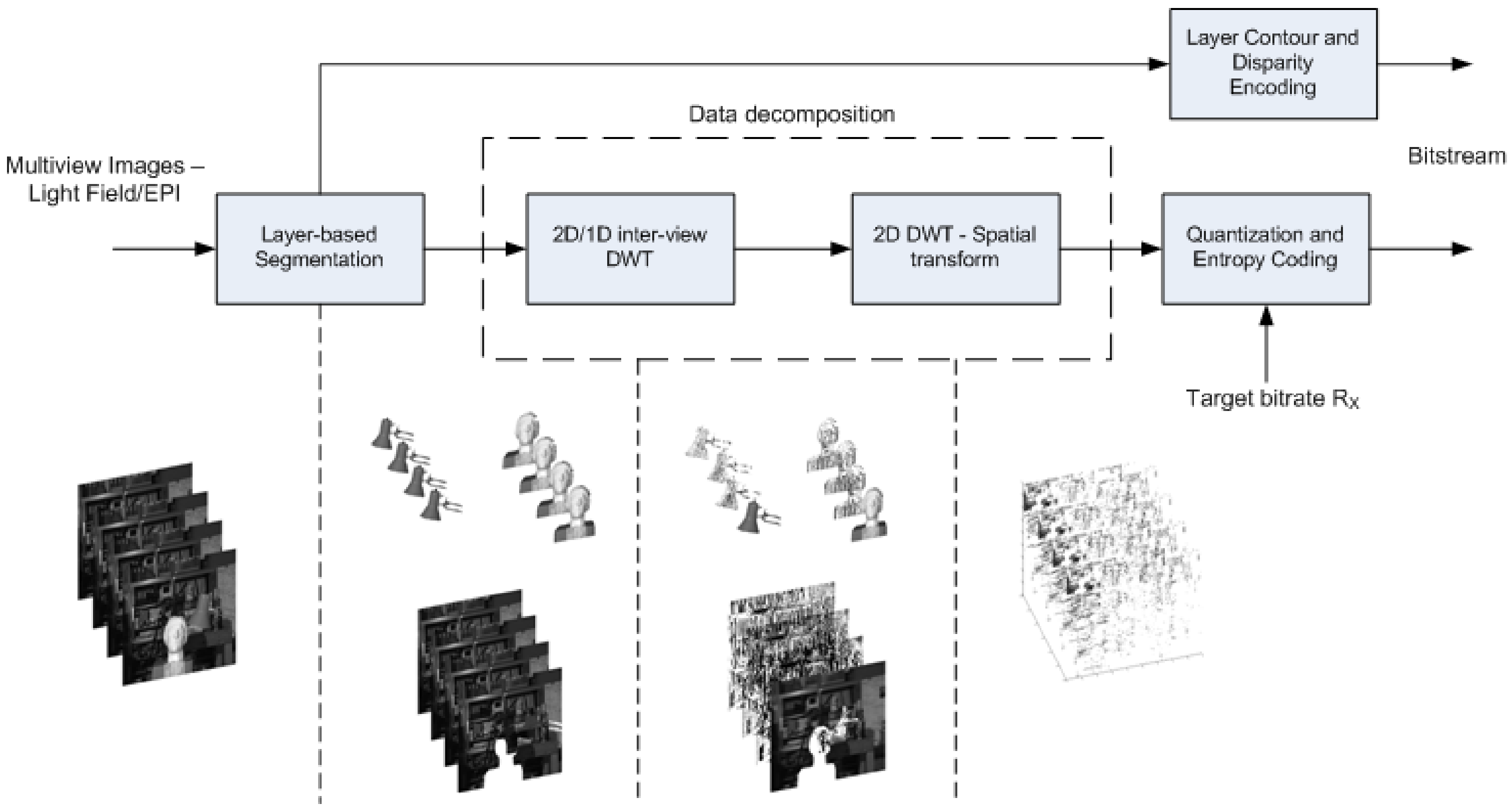
Layer-Based Compression of the Multi-view Images [GelmanDV:12]

Exploit the structure of the data in order to maximize compression efficiency:

- Decompose the EPI into layers with similar depths.
- *The disparity of each depth layer is constant*
- *Occlusion ordering can be inferred from the layer depth*

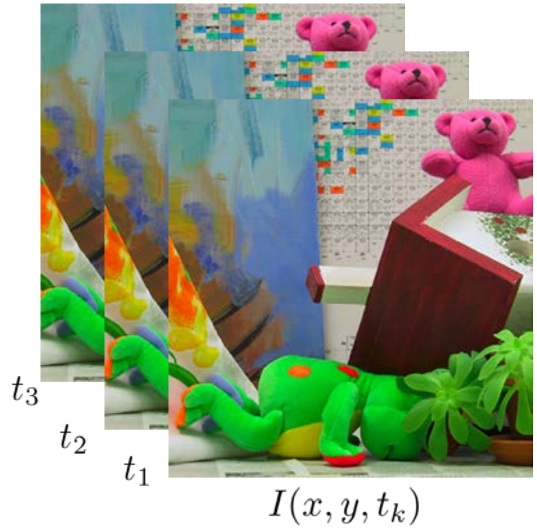


Overview of Layer-Based Compression Algorithm

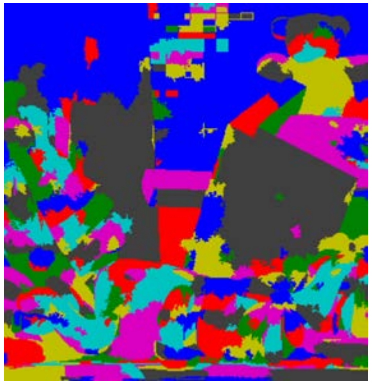


Unsupervised Layer Extraction [BerentDB:09, PearsonDB:11, PearsonBD:12]

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: $E_n(m) = \sum_{\mathbf{p} \in S_n} f(\mathbf{p}, m)$

$$\mathbf{p}_{m,k} = (x_{\mathbf{p}} - d_m t_k, y_{\mathbf{p}}, t_k)$$

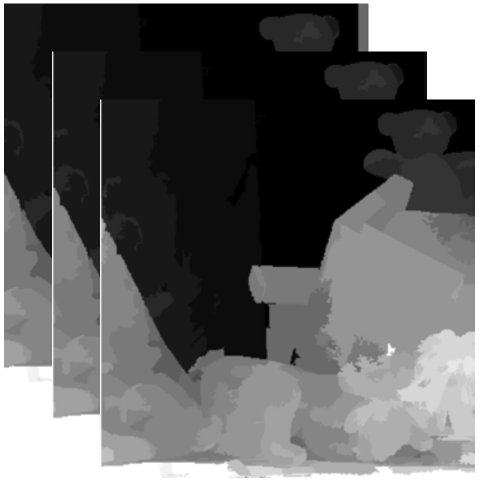
$$f(\mathbf{p}, m) = \sum_{k=1}^{K-1} |I(\mathbf{p}_{m,k}) - I(\mathbf{p}_{m,k+1})|$$

Imperial College London

Unsupervised Layer Extraction

5. Generate layers:

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



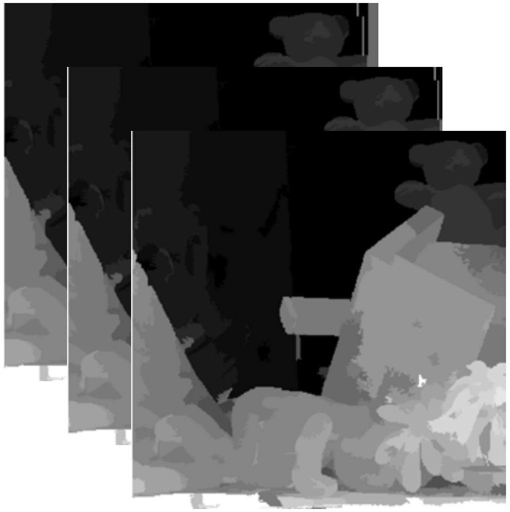
6. Run matching with occlusions:

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:

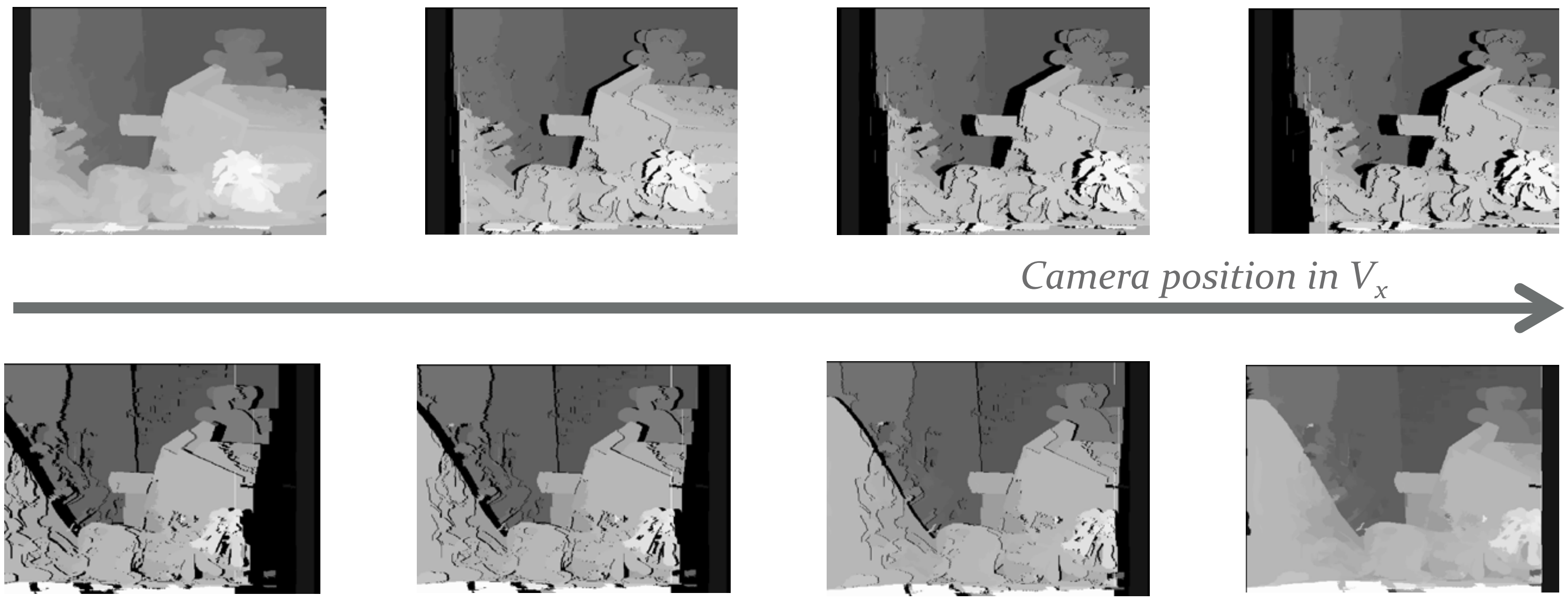
$$f(\mathbf{p}, m) = \frac{\sum_{k=1}^{K-1} |I(\mathbf{p}_{m,k}) - I(\mathbf{p}_{m,k+1})| V(\mathbf{p}, m, k) V(\mathbf{p}, m, k+1)}{\sum_{k=1}^{K-1} V(\mathbf{p}, m, k) V(\mathbf{p}, m, k+1)},$$



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

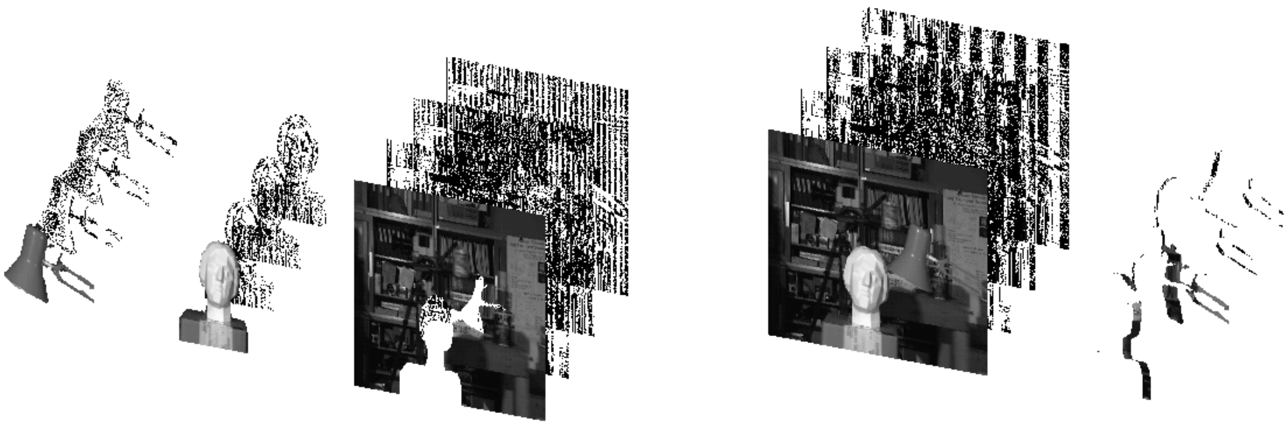


Rendered from RHS

Imperial College London

Disparity Compensated Wavelet Transform

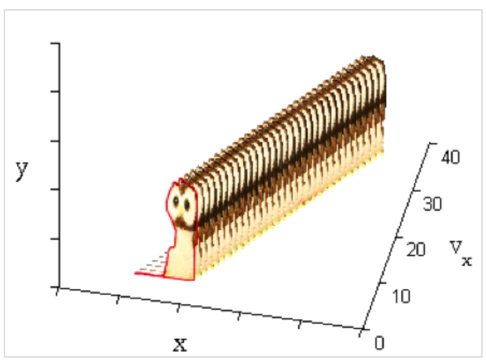
- *Apply a disparity compensated wavelet transform along the view domain*
- *Apply a 2-D WT on the recombined layer after the view-domain transform*
- *Wavelet Coefficients are quantized using a method similar to the one in EBCOT*



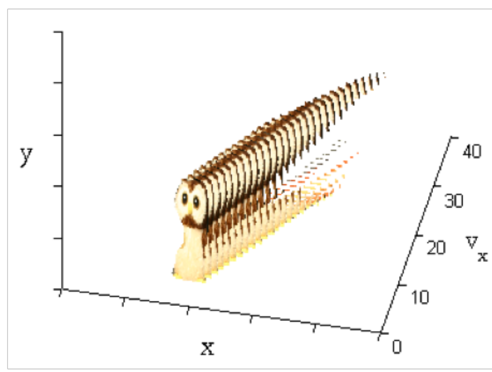
(a) (b)

Contour Encoding

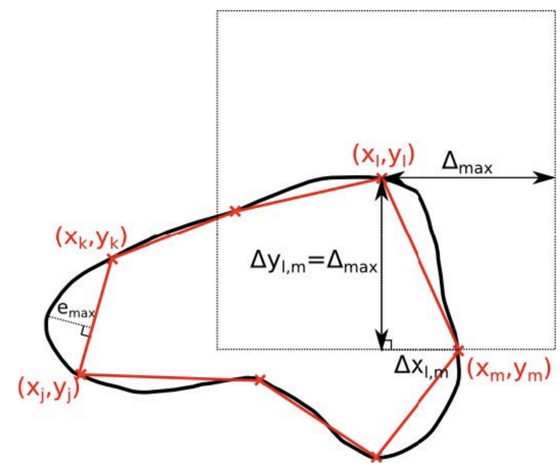
- *Only 2-D contour + disparity information needs to be encoded*
- *Contours of the layers are lossy or lossless compressed according to the bit budget*
- *Piecewise Linear Approximation of contours*



(a)



(b)



Imperial College
London

Layer-Based Compression: Simulation Results

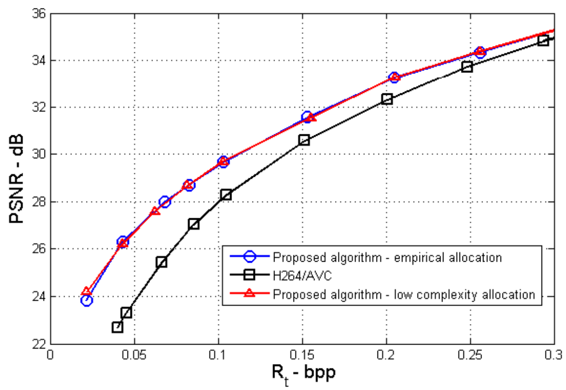
(a) Animal Farm

(b) Teddy

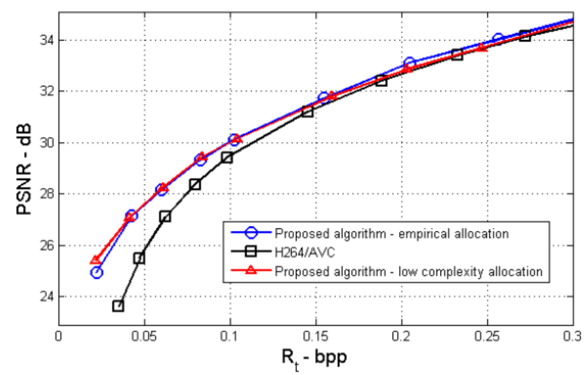
(c) Tsukuba EPI

(d) Cones

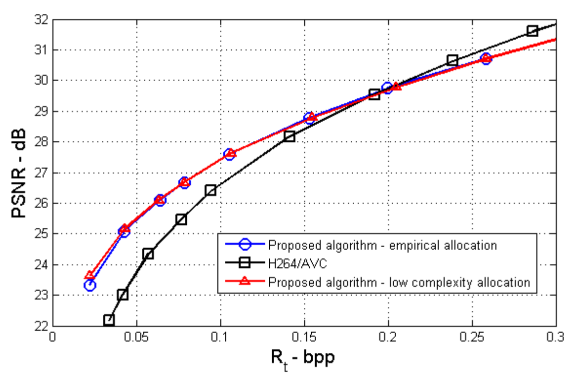
Layer-Based Compression: Simulation Results



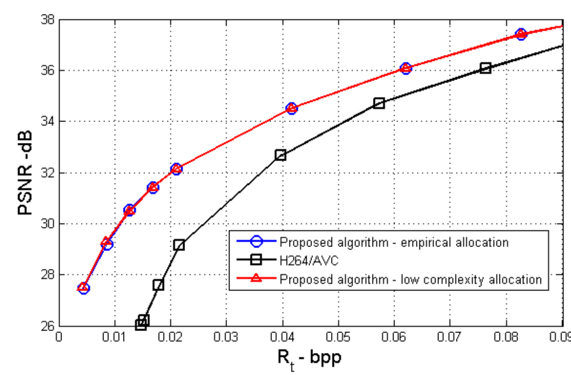
(a) Tsukuba EPI



(b) Teddy



(c) Cones



(d) Animal Farm

Imperial College London

Layer-Based Compression: Simulation Results

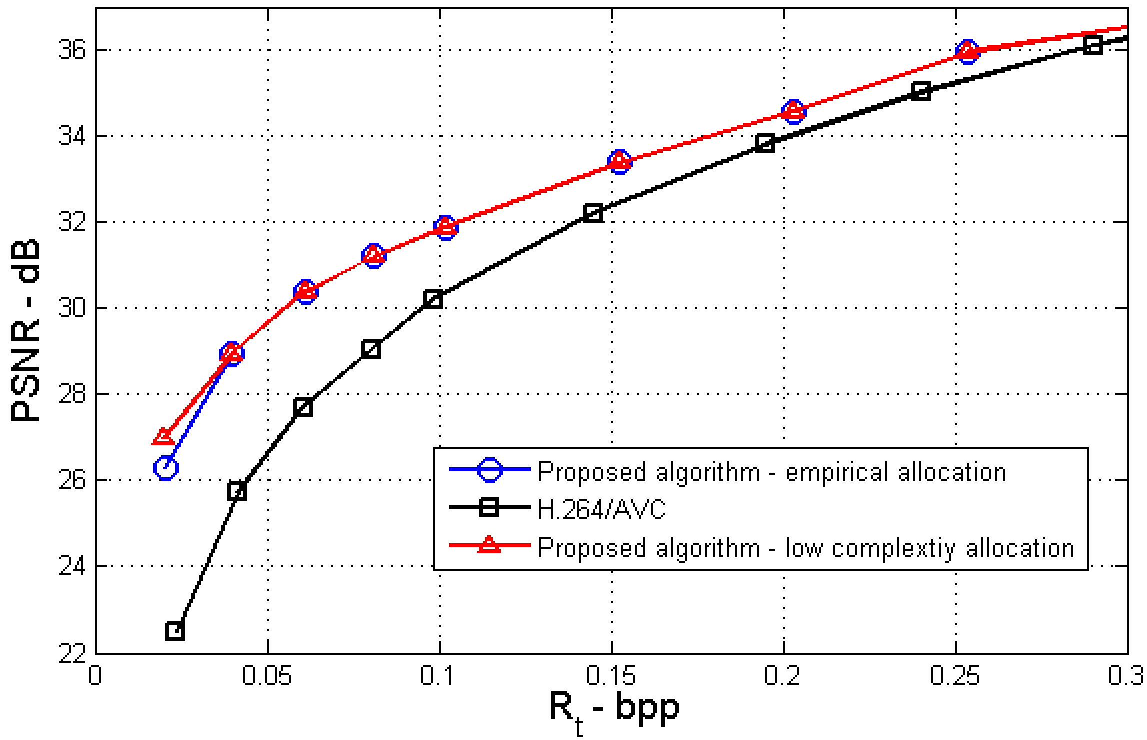
(a)

H.264/AVC

(b)

Layer-based Compression

Layer-Based Compression: Simulation Results



Tsukuba Light Field

Imperial College London

Layer-Based Compression: Simulation Results

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

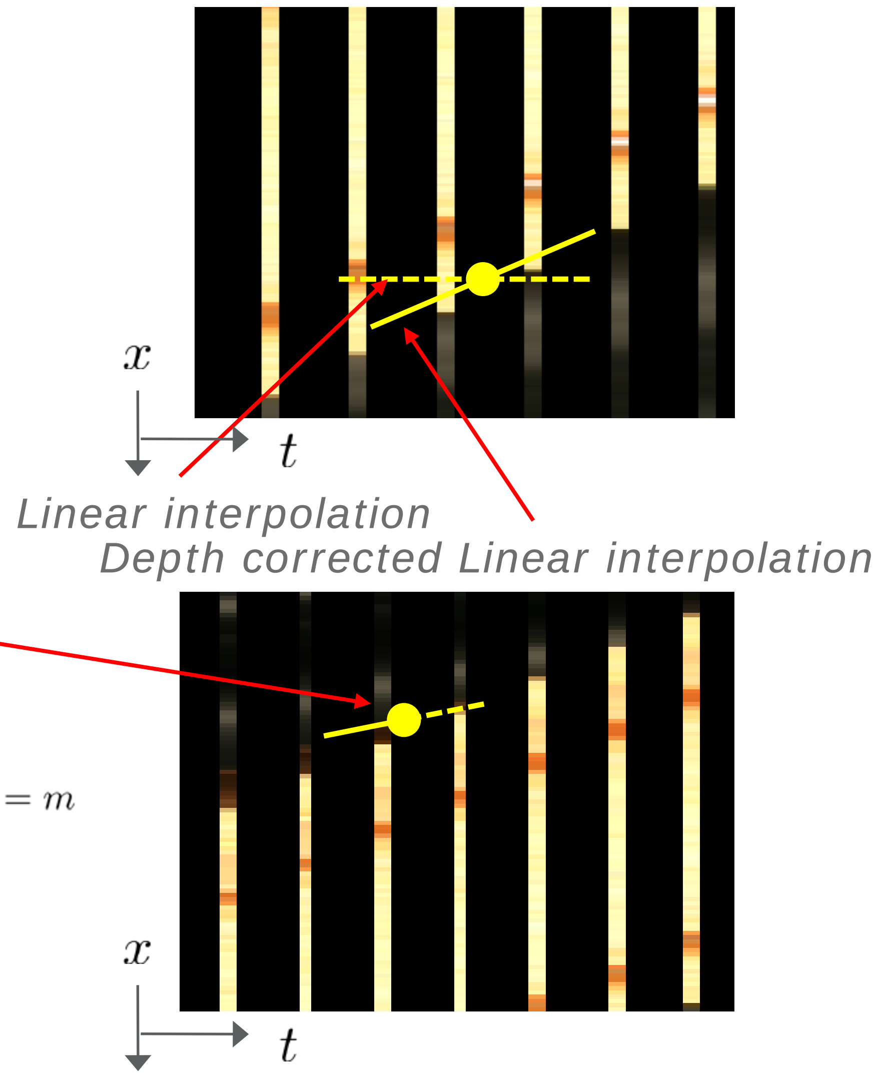
Layer-based Compression
(PSNR: 29.8dB, 0.05bpp)

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



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

IBR Results: Rendering Quality versus Layers

Layers:

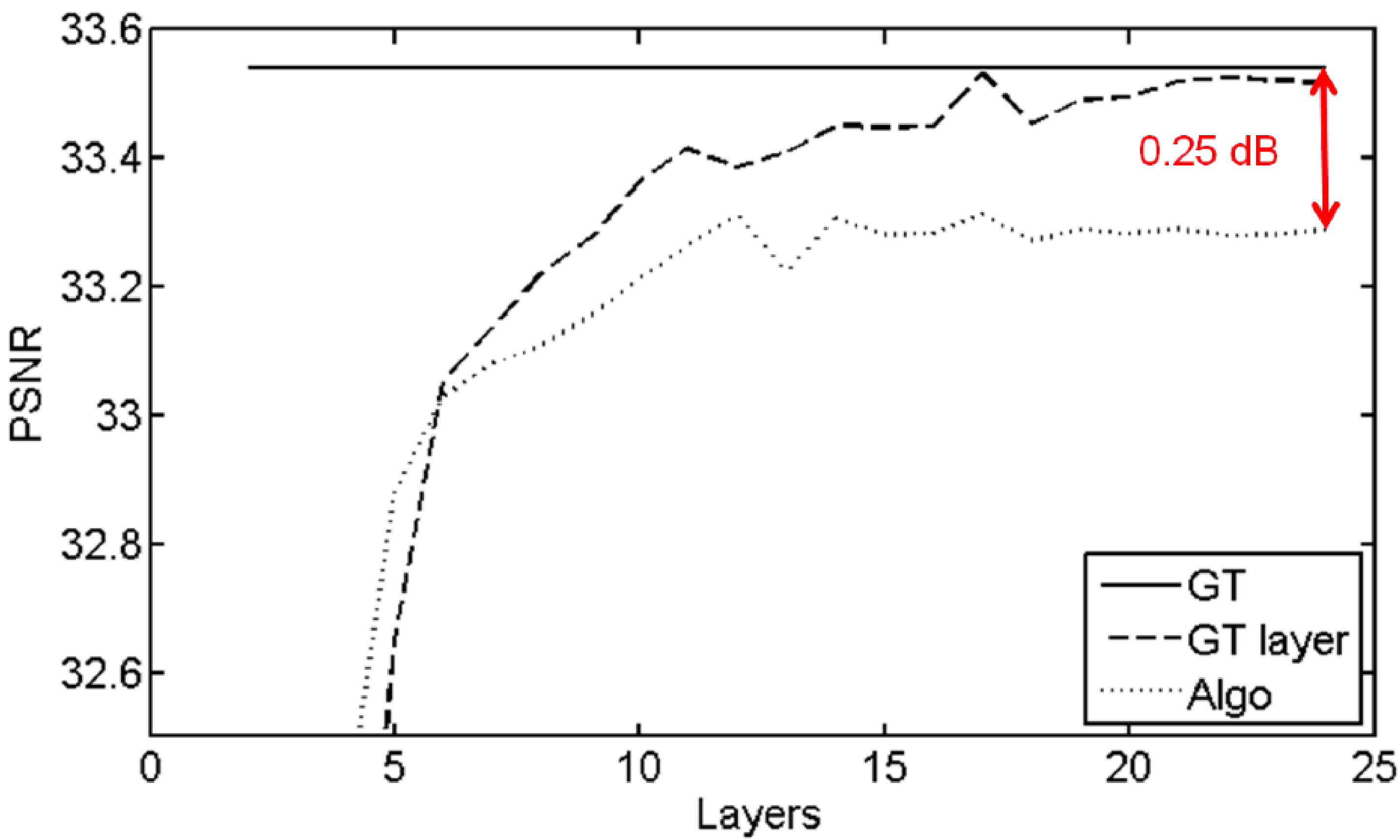
Synthesized view:

Error:

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

Imperial College London

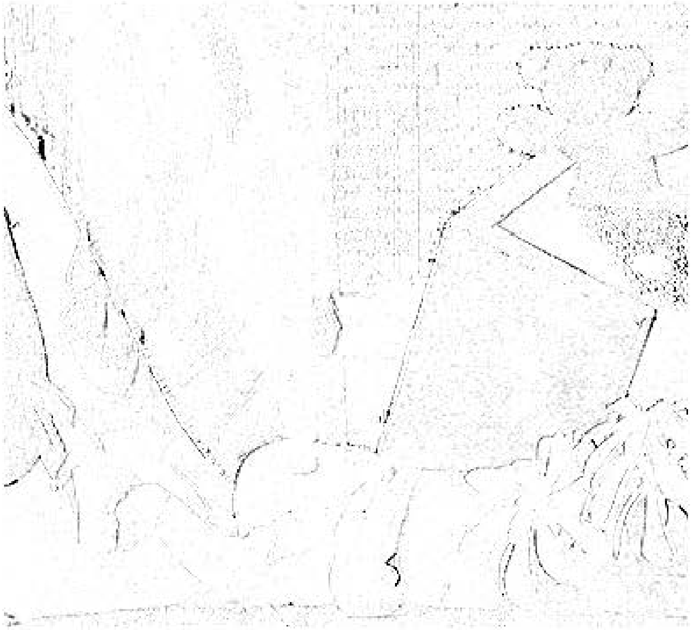
IBR Results: Rendering Quality versus Geometry



Imperial College
London

IBR Results

(a) Output



(b) Error

IBR Results on the Lightfield

http://www.commsp.ee.ic.ac.uk/jpearson/ICASSPdemo

Imperial College London

Conclusions and Outlook

- Image-based rendering is more relevant now than ever
- Plenoptic domain viewpoint is helpful
- Use the peculiar structure of Multi-view image for efficient Rendering and Compression Algorithms

- On compression:
  - Competitive algorithms for centralized 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

References

- On depth layer extraction and IBR:
  - J.Pearson, M. Brookes and P.L. Dragotti, Plenoptic layer-based model for image based Rendering, IEEE Trans. on Image Processing, 2012, submitted
  - 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 coding using depth layers and an optimized bit allocation, IEEE Trans. on Image Processing, September 2012
  - V. Chaisinthop and P.L. Dragotti, ` Centralized and Distributed Semi-Parametric Compression of Piecewise Smooth Functions' Semi-Parametric Compression of Piecewise-Smooth Functions', IEEE Trans. on Signal Processing, July 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.