Imperial College London

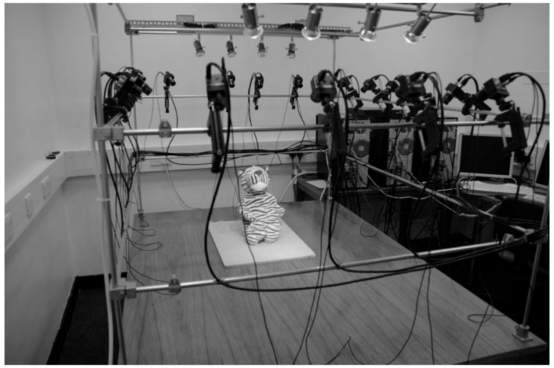
Segmentation of Multiview Data for Scene Analysis and Compression

Jesse Berent and Pier Luigi Dragotti

November 9th, 2006

Camera Sensor Networks

- Camera arrays provide samples of the plenoptic function (multiple viewpoints of a scene)
- Huge amount of data!
- The data is highly correlated and structured
- Unsupervised data analysis
  - Object or layer extraction
    - scene interpretation
    - layer based representations
  - Occlusion detection
    - innovation processes
- Applications
  - Computer vision: automatic scene interpretation
  - 3DTV
  - …



Imperial College London

Talk Outline

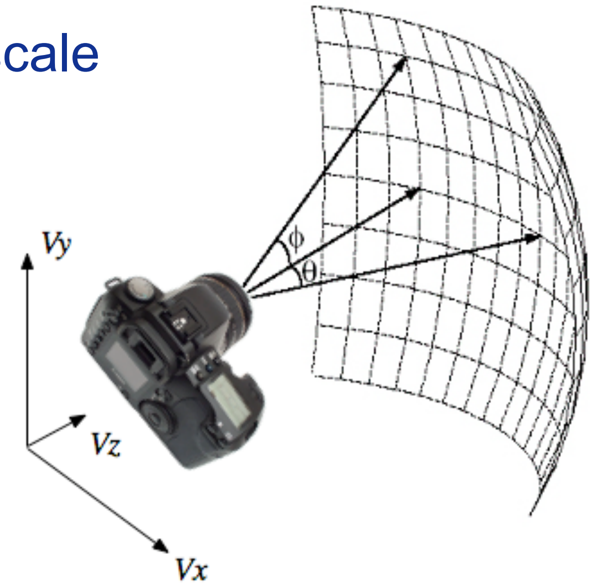
1. Introduction to the plenoptic function
2. Different camera setups and the Epipolar-Plane Image (EPI)
3. A brief review of active contours
4. Derivation of ‘constrained’ evolution equations for the plenoptic function
5. Conclusion and future work

The Plenoptic Function

- 7D function that describes the intensity of each light ray that reaches a point in space [AdelsonB:91]

$$P_7 = I(V_x, V_y, V_z, \phi, \theta, \tau, \lambda)$$

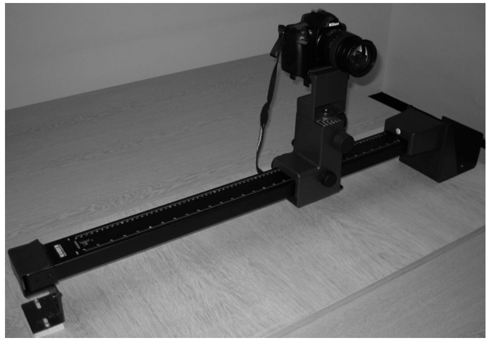
- Assumptions can be made to reduce the high number of dimensions
  - 3 channels for RGB or 1 channel for grayscale
  - Static scenes
  - Viewing position constraints



Different camera setups

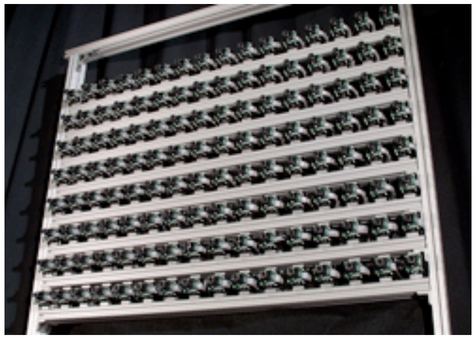
3D

3D



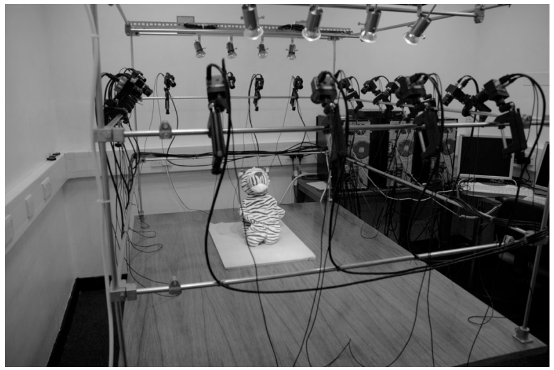
2D

4D



[Stanford multi-camera array]

5D

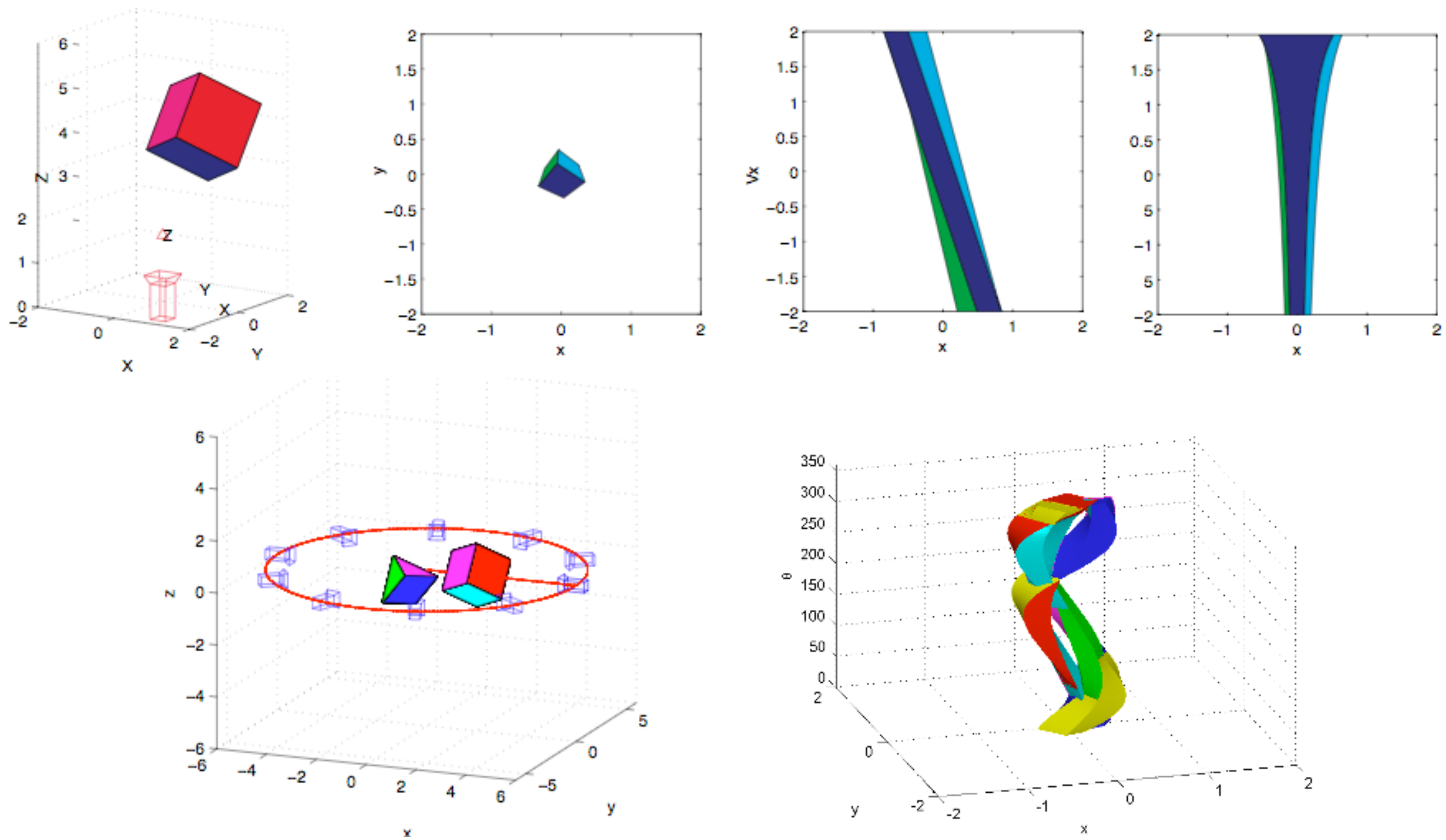


[Imperial College multi-camera array]

Imperial College London

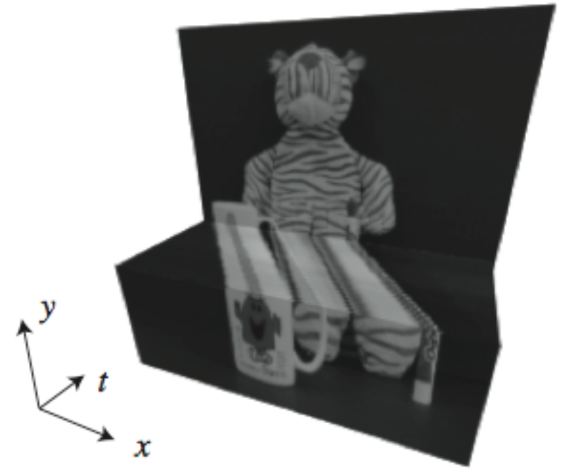
Plenoptic Functions

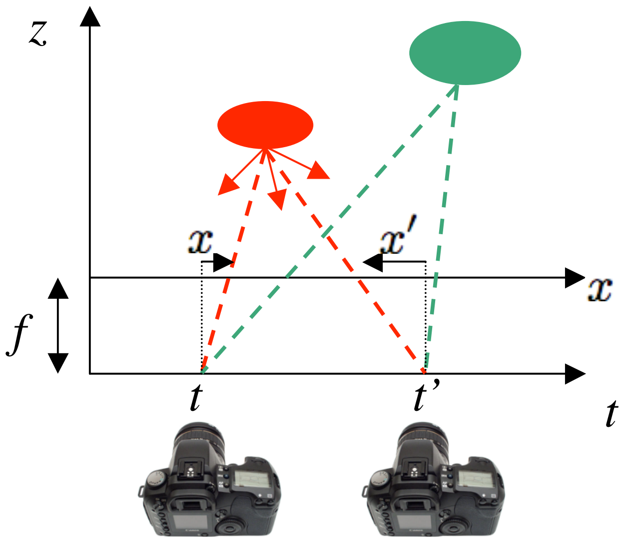
- [Images courtesy of Yizhou Wang]

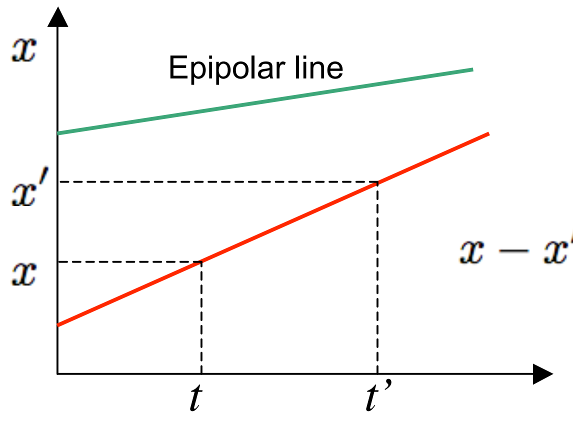


The Epipolar-Plane Image (EPI) Volume

- First introduced in [BollesBM:87]
- Cameras are constrained to a line
- Points in space are mapped on to lines
- The slope of the line ∝ 1/depth
- Objects correspond to 3D tubes



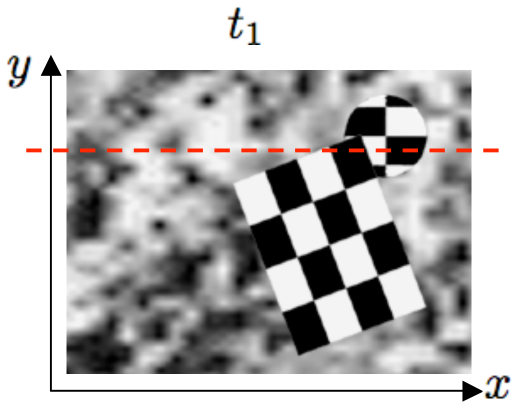


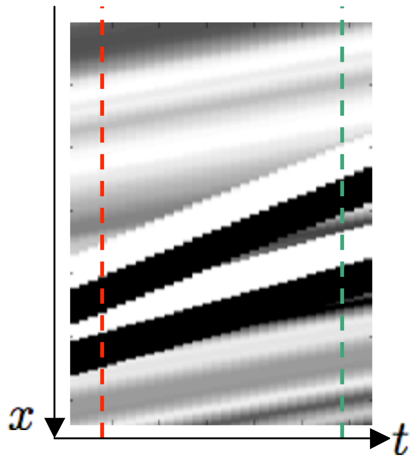


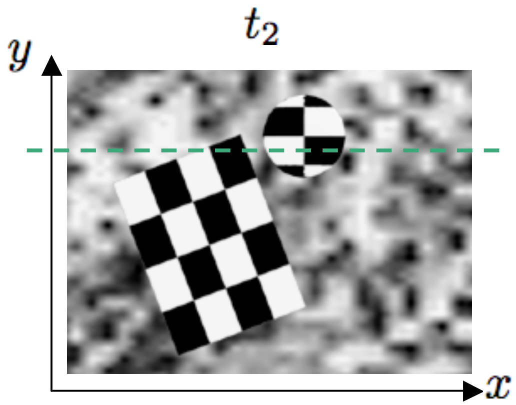
$$x - x' = (t - t')\frac{f}{z}$$

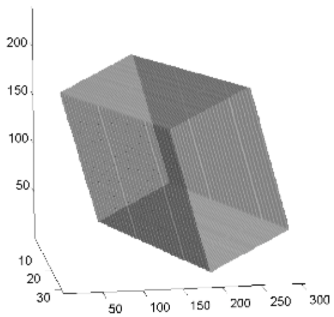
Object Tubes

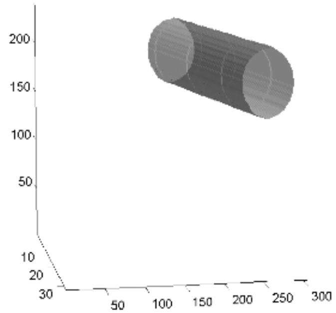
- EPI is made of a collection of tubes

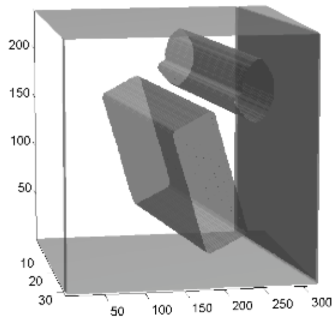








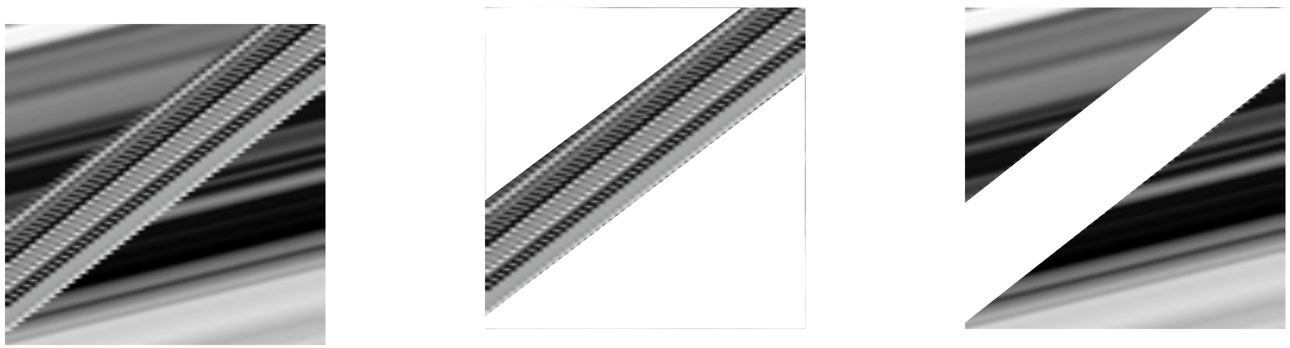




Occlusions

- A line with a larger slope will always occlude a line with a smaller one
- Occlusions occur at line intersections
- Occlusions are explicit
- Object tubes can be 'orthogonalized'

$$\mathcal{V}_1^{\perp} = \mathcal{V}_1 \qquad\qquad \mathcal{V}_2^{\perp} = \mathcal{V}_2 \cap \overline{\mathcal{V}_1^{\perp}}$$



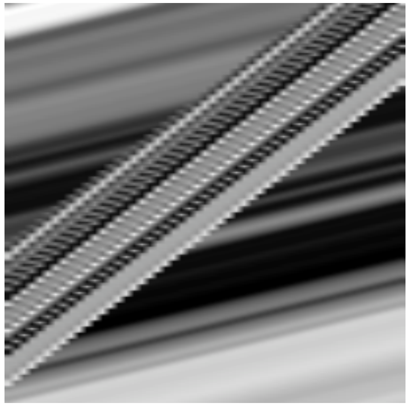
Imperial College London

Some of the Related Work

- Extracting layers using EPI analysis: Criminisi et al. 2002
- Image Cube Trajectories (ICT): Feldmann et al. 2003
- Space-time video analysis, object and occlusion volumes: Konrad and Ristivojevic 2006
- Layered stereo with occlusions: Tomasi-Lin-Birchfield 1999

Object Tube Extraction

- We assume Lambertian and opaque surfaces
- Minimize a global cost function



$$E_{tot} = \sum_{n=1}^{N} E_n = \sum_{n=1}^{N} \iiint_{\mathcal{V}_n^{\perp}} f_n(\vec{x}) d\vec{x}$$

$f_n(\vec{x})$ is a measure of consistency with tube n

- Separated in 2 sub-problems:
  - Estimation of contours given the slopes of the lines
  - Estimation of the slopes given the contour

Imperial College London

Estimation of the Region Boarders: Active Contours

- Consider a cost function of the type:

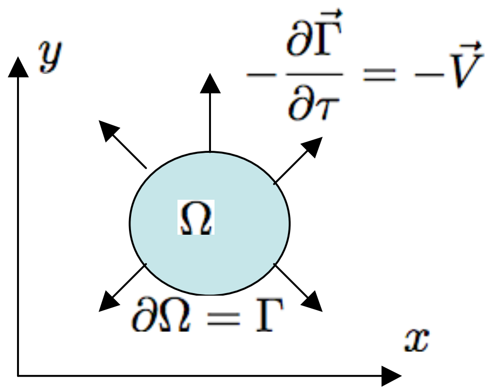
$$E(\Gamma) = \iint_{\Omega(\Gamma)} f(x,y)dxdy + \iint_{\overline{\Omega}(\Gamma)} g(x,y)dxdy + \int_{\Gamma} \lambda ds$$

- Gradient [KassWT:88, CasellesKS:97, ChanV:01, Jehan-BessonBA:01]

$$\frac{dE(\tau)}{d\tau} = \int_{\partial\Omega} [f(x,y) - g(x,y) + \lambda\kappa](\vec{V}\cdot\vec{N})ds$$

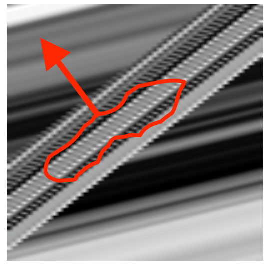
- Steepest descent

$$\frac{\partial\vec{\Gamma}(\tau)}{\partial\tau} = [f(x,y) - g(x,y) + \lambda\kappa]\vec{N} = F\vec{N}$$



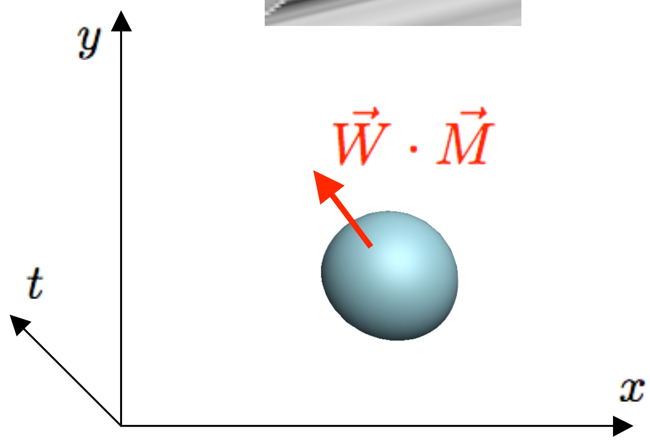
Estimation of EPI Tube Contours

- We assume the slope of the lines are known
- In the case where there are 2 layers (i.e. 1 layer and the background)



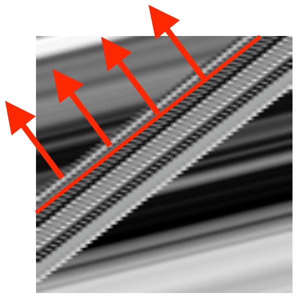
$$E_{tot}(\tau) = \iiint_{\mathcal{V}_1^{\perp}(\tau)} f_1(\vec{x})d\vec{x} + \iiint_{\mathcal{V}_2^{\perp}(\tau)} f_2(\vec{x})d\vec{x}$$

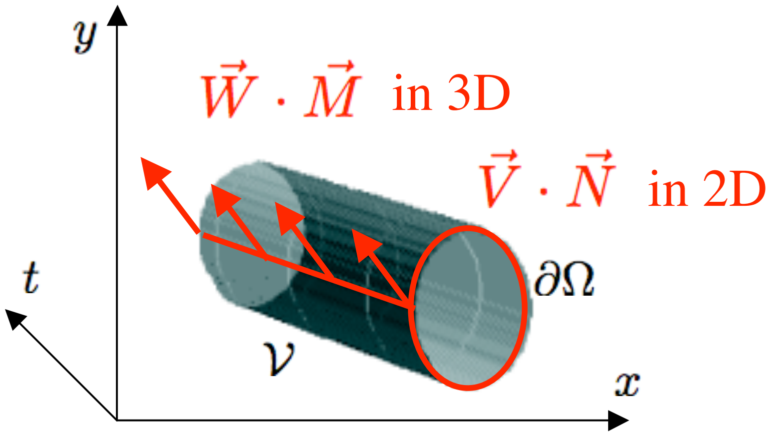
$$\frac{dE_{tot}(\tau)}{d\tau} = \iint_{\partial\mathcal{V}_1^{\perp}} (f_1(\vec{x}) - f_2(\vec{x}))(\vec{W} \cdot \vec{M})d\vec{\sigma}$$



From 3D to 2D Using Epipolar Geometry

- The positions of the cameras are known
- The shape of the tubes are constrained
- Leads to ‘constrained’ surface evolution that can be implemented in a 2D subspace





$$\vec{W} \cdot \vec{M} = \alpha(s,t)(\vec{V} \cdot \vec{N})$$

$$\iint_{\partial\mathcal{V}} f_1(\vec{x})(\vec{W} \cdot \vec{M})dtds = \int_{\partial\Omega} (\vec{V} \cdot \vec{N}) \int_t \alpha(s,t) f_1(\vec{x})dtds$$

The speed function

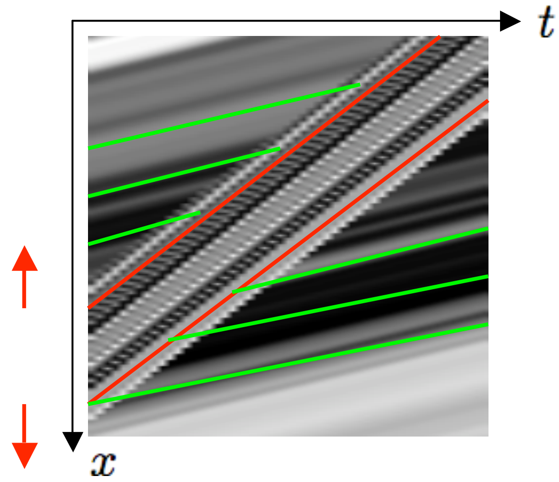
- The gradient becomes (α=1 for fronto-parallel planes)

$$\frac{dE(\tau)}{d\tau} = \int_{\partial\Omega} (\vec{V} \cdot \vec{N}) \Big[\underbrace{\int_t f_1(\vec{x}) dt}_{F_1(s)} - \underbrace{\int_t f_2(\vec{x}) dt}_{F_2(s)} \Big] ds$$

- The functional is set to be the normalized squared difference between the intensity and the mean of the line the layer belongs to

$$f_n(\vec{x}) = \frac{[I(\vec{x}) - \mu_n(\vec{x})]^2}{L_n(\vec{x})}$$

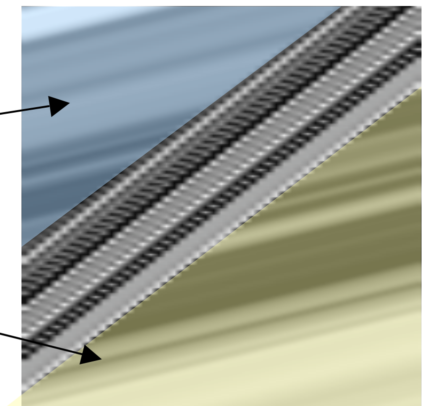
$$\vec{V} = [F_1(x,y) - F_2(x,y)]\vec{N}$$



Occlusion/Disocclusion

- Its not that simple…
- For an occluded layer, the functional depends also on τ and the evolution equation has additional terms that are extremely complex.
- We alleviate the problem by separating tubes into ‘to be occluded’ and ‘disoccluded’ regions (similar to [KonradR] for video)

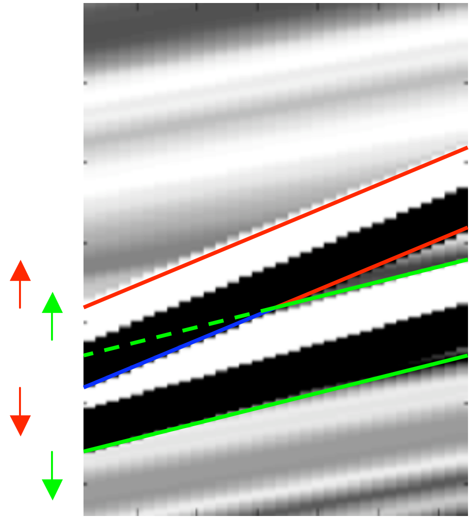
$$\mathcal{V}_2^{\perp} = \mathcal{V}_{2_{occ}} \cup \mathcal{V}_{2_{dis}}$$



Dealing with Multiple Tubes

- Evolve one tube at a time
- By construction, tubes compete only with the other tubes they are occluding or disoccluding
- For example:

$$
\begin{aligned}
\mathcal{V}_1^{\perp} &= \mathcal{V}_1 \\
\mathcal{V}_2^{\perp}(\tau) &= \mathcal{V}_2(\tau) \cap \overline{\mathcal{V}_1^{\perp}} \\
\mathcal{V}_3^{\perp}(\tau) &= \mathbb{R}^3 \cap (\overline{\mathcal{V}_1^{\perp}} \cap \overline{\mathcal{V}_2^{\perp}}(\tau))
\end{aligned}
$$



$$
E_{tot}(\tau) = \iiint_{\mathcal{V}_2^{\perp}(\tau)} f_2(\vec{x})d\vec{x} + \underbrace{\iiint_{\mathcal{V}_1^{\perp}} f_1(\vec{x})d\vec{x} + \iiint_{\mathcal{V}_3^{\perp}(\tau)} f_3(\vec{x})d\vec{x}}_{\iiint_{\overline{\mathcal{V}_2^{\perp}(\tau)}} f(\vec{x})d\vec{x}}
$$

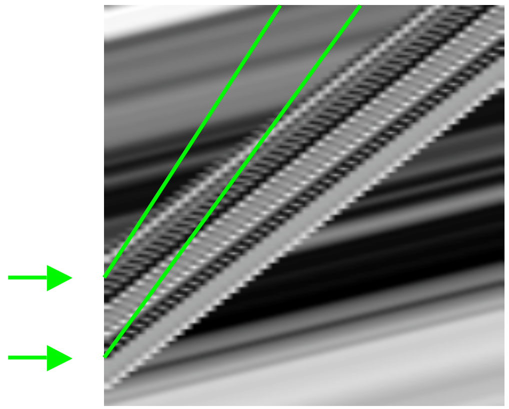
Estimation of Line Slopes (i.e. Disparity)

- Contours are fixed
- Find the slopes of the lines
- Done jointly over all the images
- Takes into account occlusions

$$E_{tot} = \sum_{n=1}^{N} E_n = \sum_{n=1}^{N} \iiint_{\mathcal{V}_n^{\perp}} f_n(\vec{x}) d\vec{x}$$

$$f_n(\vec{x}) = \frac{[I(\vec{x}) - \mu_n(\vec{x})]^2}{L_n(\vec{x})}$$

$$\mu_n(\vec{x}) = \mu(x, y, t, d_n(x, y))$$

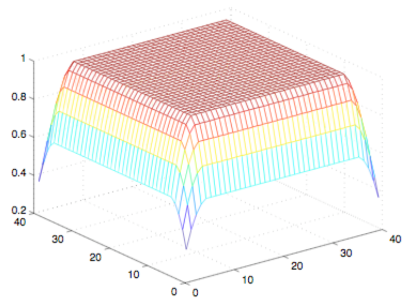


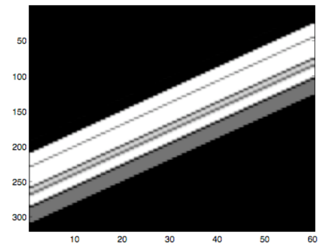
- Non-linear optimization problem

Imperial College London

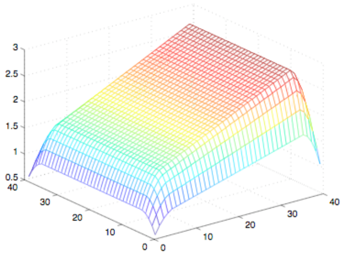
Layer Disparity Model

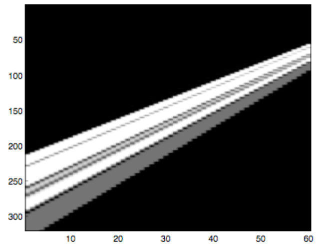
- Disparity map can be modeled as a bicubic spline [LinT:03]

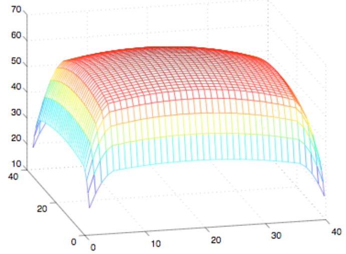


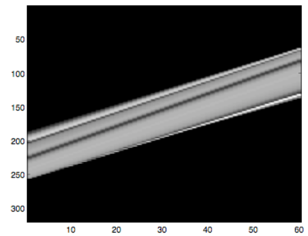


$$d_n(x,y) = \sum_{i,j} D_n(i,j)b(x-i, y-j)$$









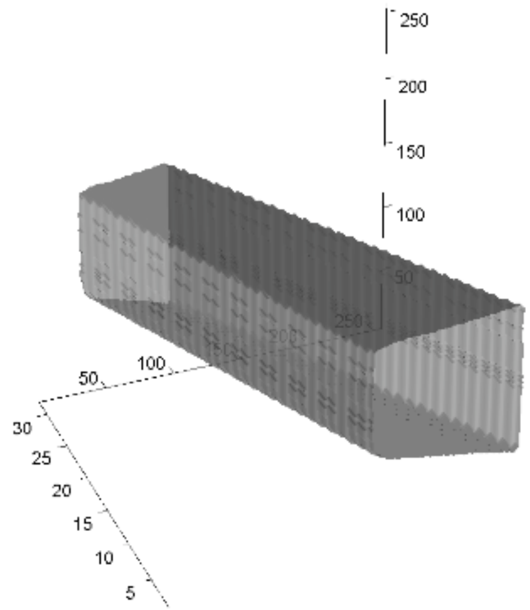
Overall optimization

- Initialize
- Iteratively alternate
  - Segmentation given layer depth maps
    - Evolve each contour iteratively with the level set method
  - Estimation of depth maps given segmentation
    - using classical optimization methods
- End when there is no significant decrease in energy

Imperial College London

Simulation Results

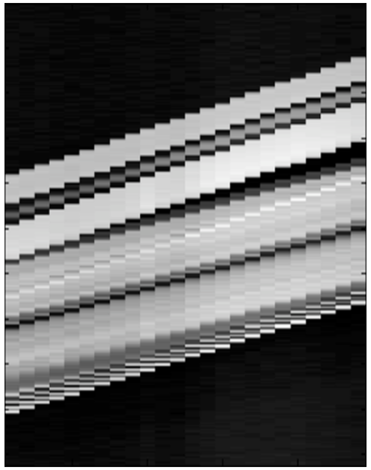
- Slanted planes can be a problem in classical stereo since there is not a 1 to 1 mapping.
- Not a problem here since slanted planes are taken into account in the model

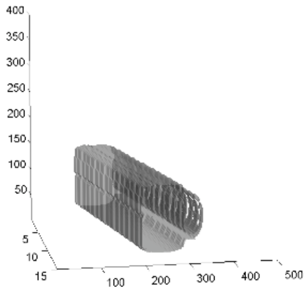


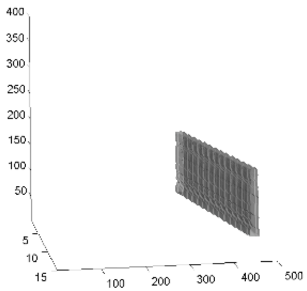
Imperial College London

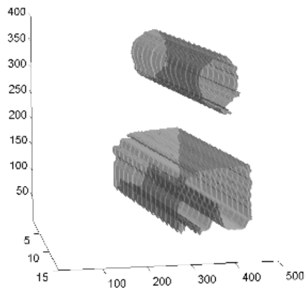
Preliminary Experimental Results

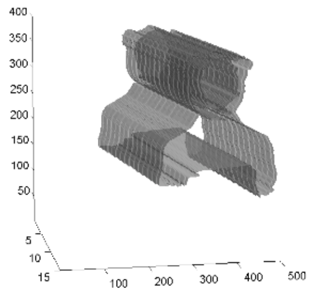
- Tiger image sequence (15 images covering 5 degrees)











Imperial College London

Conclusions

- The plenoptic function provides a nice framework for multiview image analysis!
- New segmentation scheme for the Epipolar-Plane Image volume
  - Constrained surface evolution (uses knowledge of camera setup for added robustness)
  - Takes into account all the images simultaneously
  - Handles occlusions
  - Is scalable to higher dimensions

Imperial College London

Ongoing and Future Research

- Extension to the 4D and 5D cases: More degrees of freedom to the camera locations
  - Segmentation of hyper-volumes
- Scene interpretation
  - What can we learn about the scene from the shape of the tubes?
- Compression
  - Layer based representations and/or linear transforms taking into occlusions and disparities (along the EPI lines)

Imperial College London

Questions?