(This is so ’90s)

This is better!

The brave new media: a plenoptic journey

Dan Lelescu
Micron Technology
Imaging Division R&D Lab

UC Santa Barbara
November 9, 2006

Outline

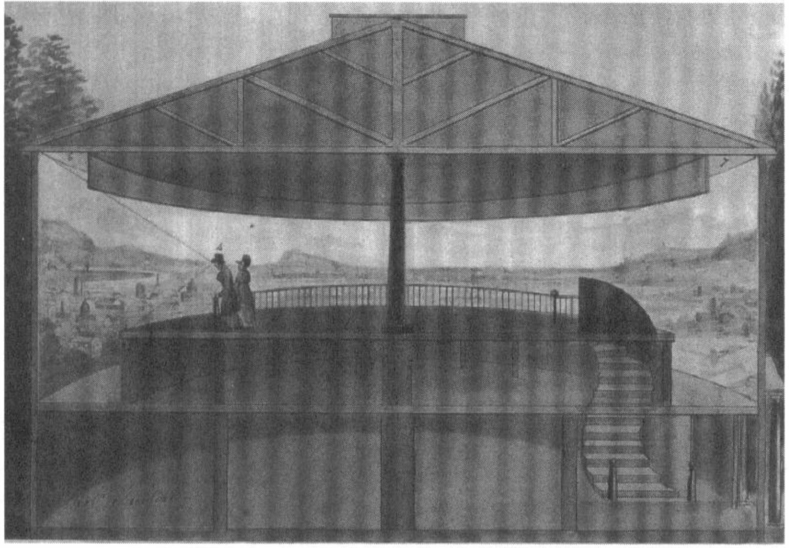
- Introduction.
- Plenoptic acquisition.
- Plenoptic signal processing (and computer vision, and graphics, and image/video coding, and...).
- Summary.
- Time permitting: will show volumetric 3D displays, omnidirectional cameras, QuicktimeVR panoramas.

Introduction

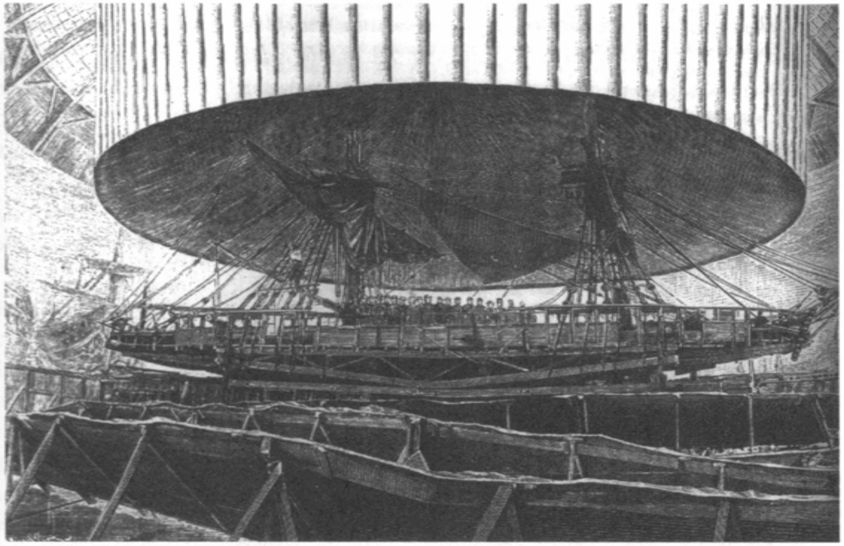
Early inquiries

- Interest in surpassing the limitations of the human visual system (HVS)*.

R. Fulton's Description of Panorama, Paris, 1799



Platform of panorama "Le Vengeur", from La Nature,1892

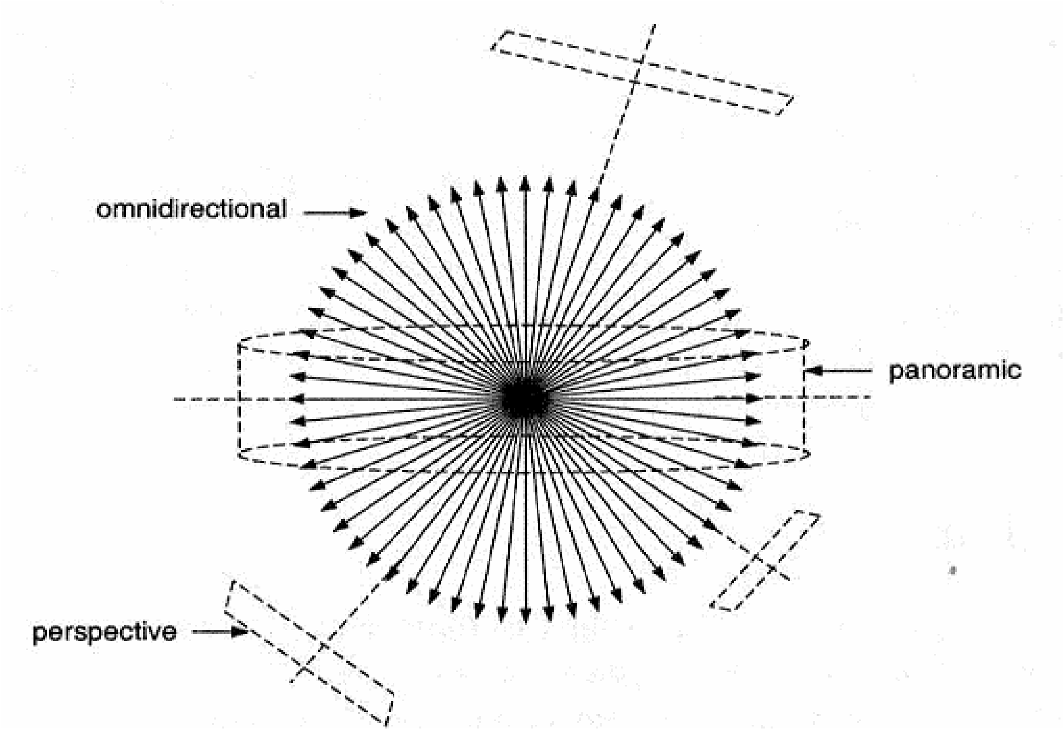


- Definition of "light field", 1939 [1]
- The 50's 3D (red-green glasses): "Creature from the Black Lagoon" (1954)

*Some functions of HVS still cannot be rivaled by machine vision today: e.g., object discrimination/recognition, depth perception etc.

The ray space

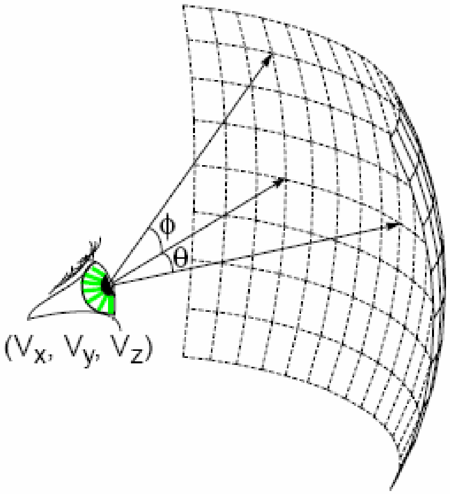
- The light rays passing through a point $\{V_x, V_y, V_z\}$ in space form a pencil of rays.
- By taking a subset of these rays various types of views can be generated.



Plenoptic function

- The *plenoptic function* [Latin, *plenum*] was introduced formally in [2] in 1992.
  - Describes all light information collected at a point in space-time

- The plenoptic function is originally a 7D function, $\boxed{f(V_x, V_y, V_z, \theta, \varphi, \lambda, t)}$, where
  - Vx, Vy, Vz - viewpoint coords.
  - θ, ϕ - ray direction
  - λ - wavelength
  - t - time



- By fixing various parameters in the plenoptic function, one obtains different, more restrictive representations.

Challenges: a first set

- Plenoptic image acquisition
  - Sensor design, calibration, syncronization
  - Space/time sampling
  - Acquisition speed
  - Huge amount of data generated
- Plenoptic processing→ present in the other topics
  - Mappings, plenoptic representation
  - Coding
- Plenoptic signal communication
  - Transport issues (e.g., error resilience) specific to this domain
- Rendering and display
  - View reconstruction/rendering
  - Display devices (to take advantage of new imaging capability)
  - Comfort in visualization (nausea should not be part of experience).

How is all this different from Computer Graphics?

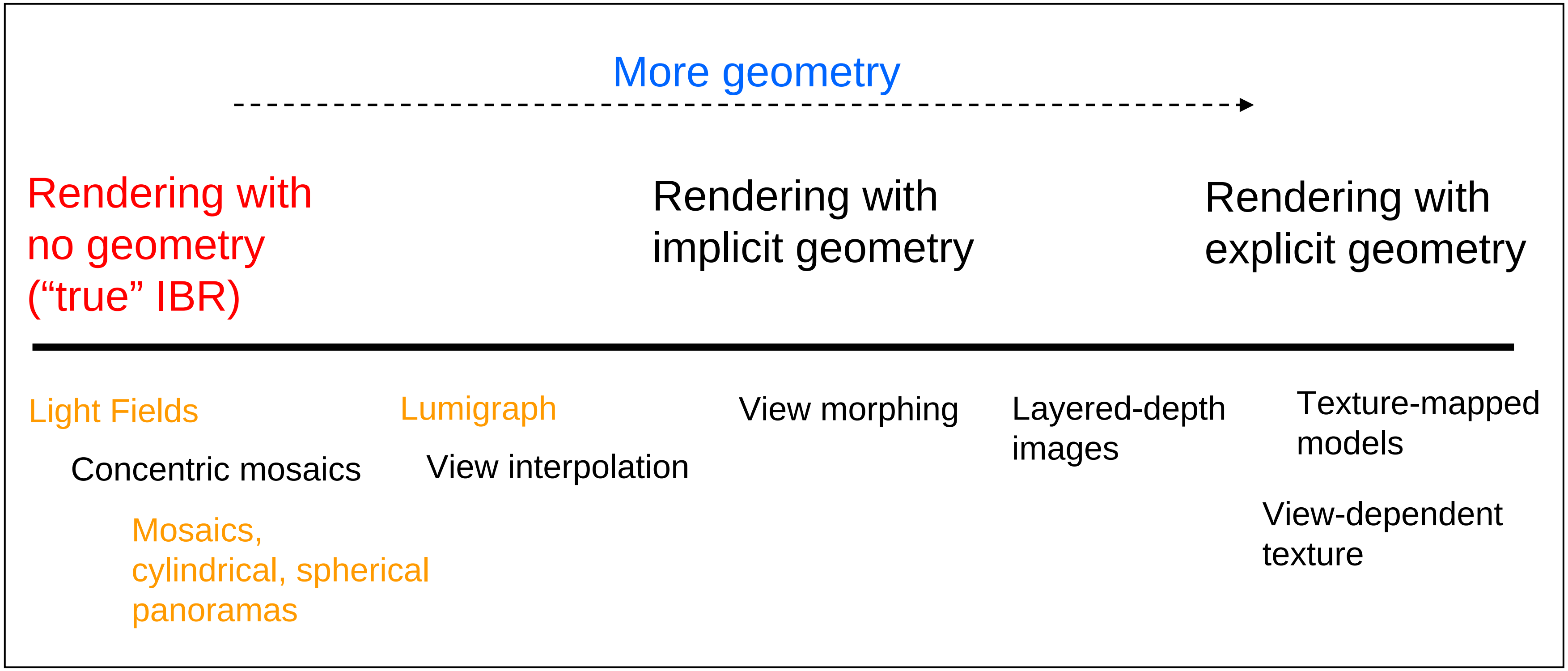
- Computer graphics (CG) is a mature field
  - Geometry modelling + texture mapping for rendering
    - Therein lies the problem for rendering natural objects/scenes

- CG techniques have limitations:
  - Natural and generic objects/scenes are extremely difficult to model
    - Even if possible, heavy computation cost

- Image-based rendering techniques (IBR) [3]
  - Attempt to use acquired images for rendering (although they can, and should, use geometry if available)
  - Elements of CG and IBR are often combined (but weight is heavily in favor of new IBR techniques)

From CG to IBR (and stations in-between)

- There is a continuum of methods spanning the IBR and CG fields



More geometry

Rendering with no geometry (“true” IBR)

Rendering with implicit geometry

Rendering with explicit geometry

Light Fields

Concentric mosaics

Mosaics, cylindrical, spherical panoramas

Lumigraph

View interpolation

View morphing

Layered-depth images

Texture-mapped models

View-dependent texture

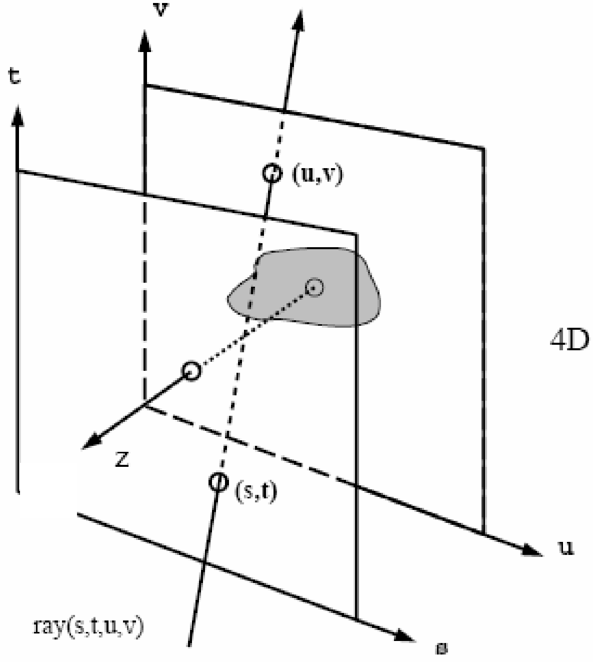
*adapted from **[3]**

- Note that stereo visualization can augment the figure above
  - e.g., one can generate stereo panoramas
  - there are specifics in terms of inducing stereopsis (for depth perception)

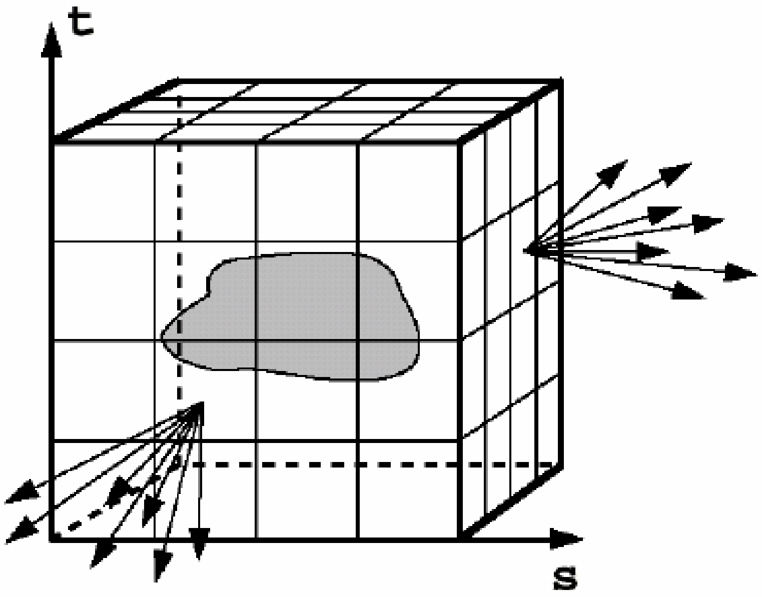
Plenoptic acquisition

Light Fields

- Light fields represent 4D parameterizations of the plenoptic function
  - Light Fields[4] and Lumigraphs[5]: a ray is indexed by its intersection with two parallel planes. [Not the only approach!]
  - Assumption of space free of occluders; six pairs of planes surrounding the convex hull of the object being imaged

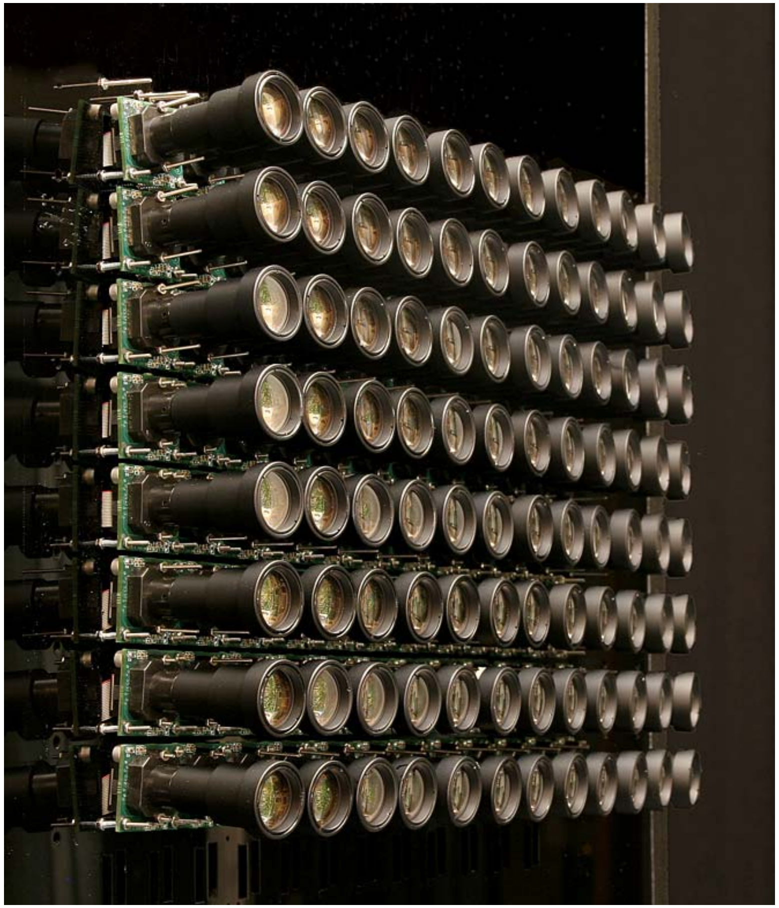


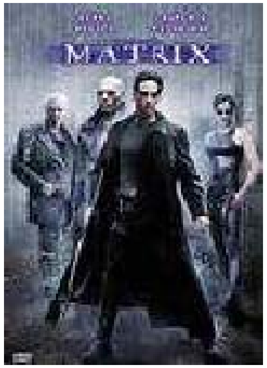
4D parameterization of rays



Light Field acquisition

- Arrays of cameras on a surface (or more restrictive arrangements).





Light Field rendering

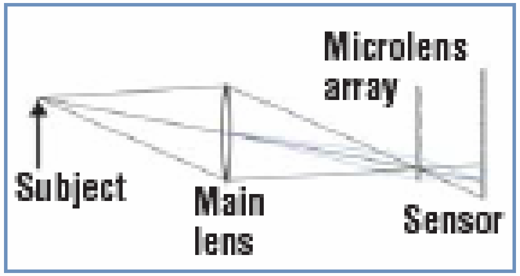
- Imaged views are trivially rendered from the ray database.
- Novel, virtual views rendering consists of two main steps:
  - Determine the coordinates of rays in the desired virtual view (within the specific light field parameterization)
  - Interpolate “neighboring” rays from database to generate the new view

*from **[6]**

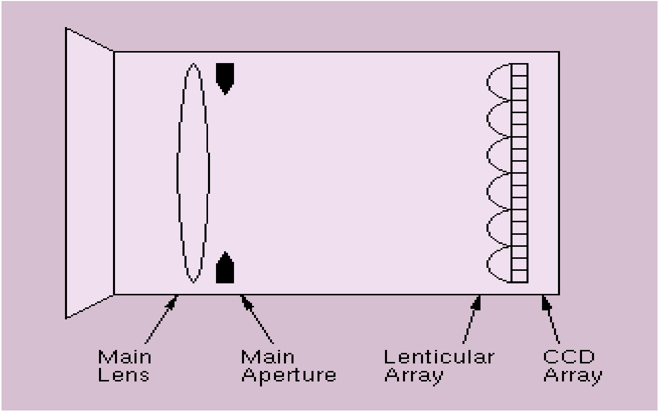
Light Field camera

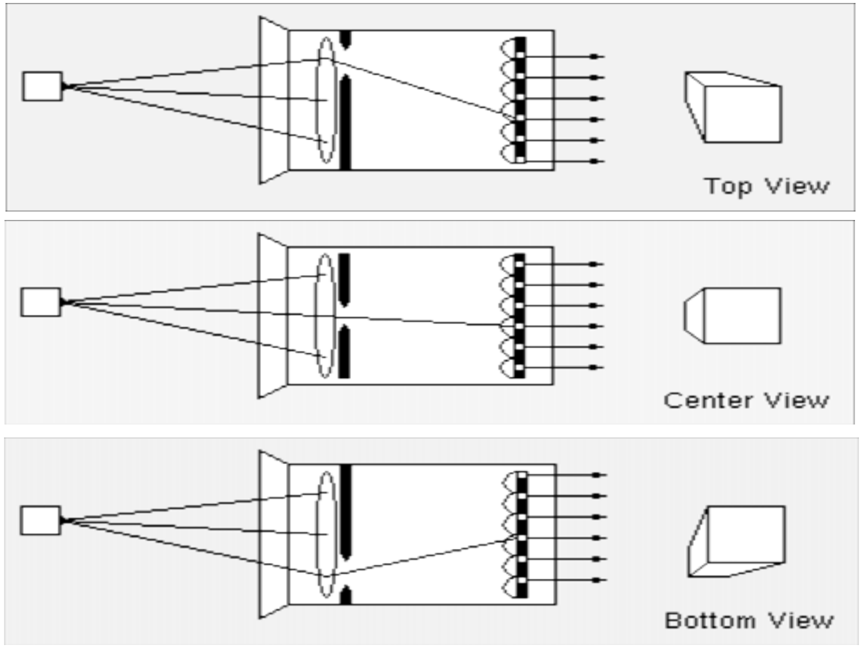
- Plenoptic (or light field) cameras, use lenticular arrays on top of the sensor array.
  - Mimic a “camera array” on a small scale
  - Capable of 3D reconstruction, variable focus (after capture!)-very promising.

[7]



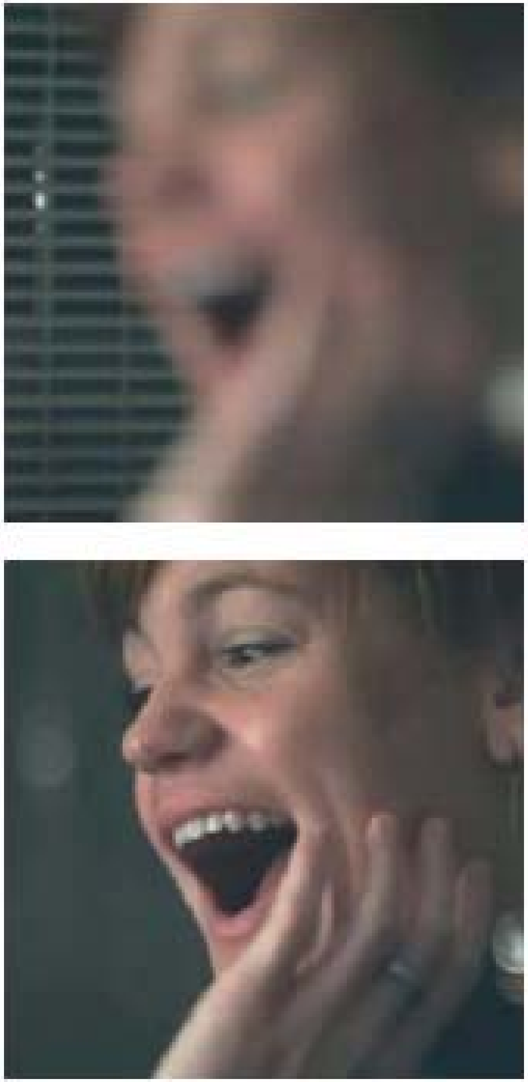
[8]





Light Field camera images

- Example ([7]).
- Focus change is a re-sampling of the light field.
- Once picture at top is taken with a “normal” camera with fixed settings, you wouldn’t be able to obtain the other image as a post-processing operation.



Omnidirectional acquisition

- For dynamic environments, “omniview” systems are needed (rotating cameras will not do).
- Catadioptric systems: lens+mirror elements
  - E.g., parabolic mirror and telecentric lens coupled with a video cam.

ParaMax Reality 360

Omnidirectional display

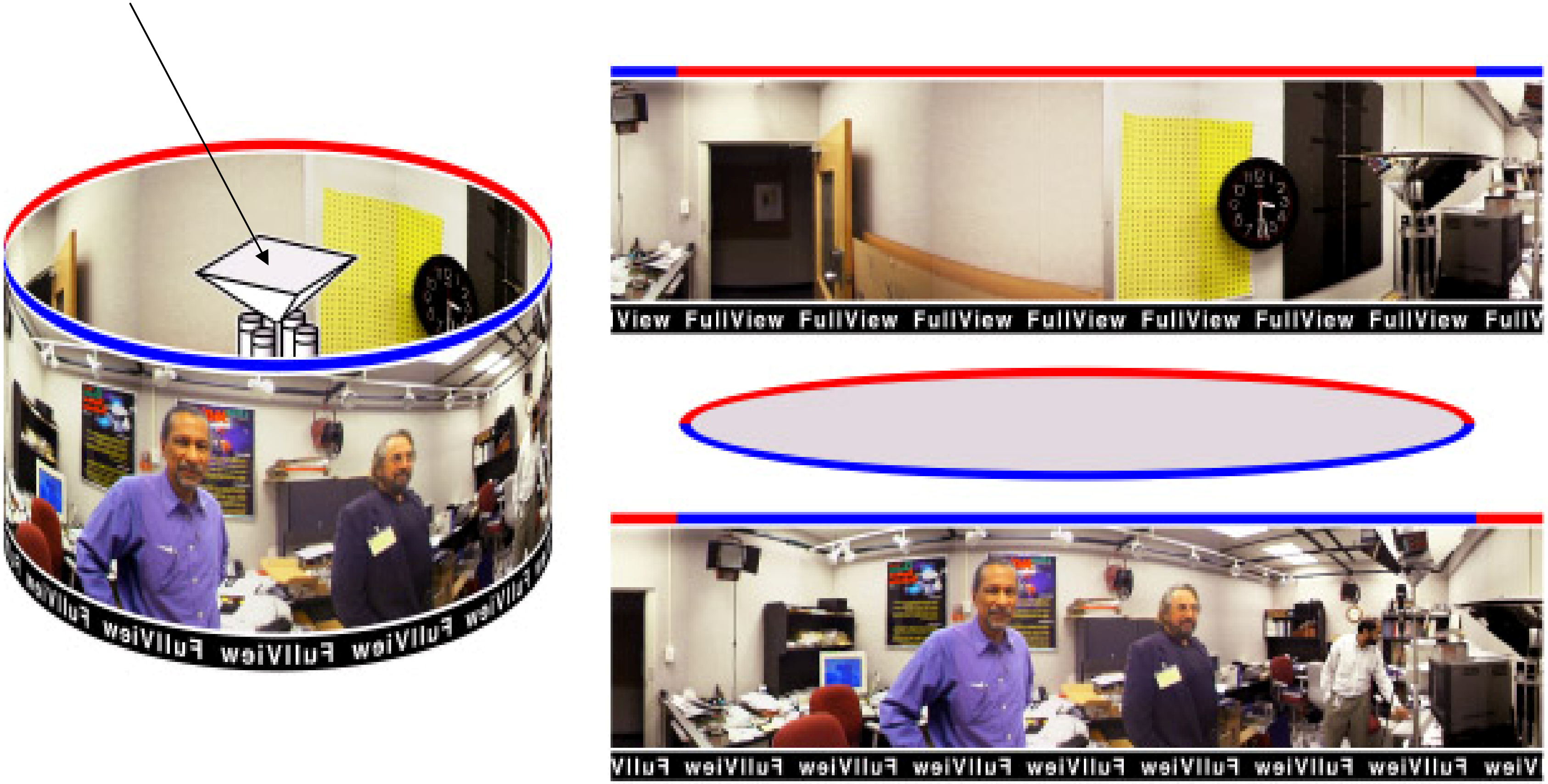
- Rendering from “omnidirectional” video camera (parabolic mirror+telecentric lens).
  - Generated arbitrary perspective views from video
  - Used either a monitor, or VR glasses + head tracker for “tele-presence” inside moving car

Omnidirectional video snapshot; captured using ParaMax Reality, a Sony video camera, a folded beach chair, and an Audi Quattro.

(DoCoMo Labs, 2002)

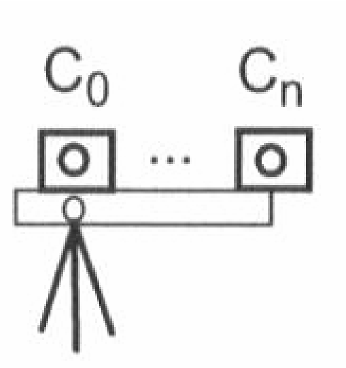
Cylindrical panorama acquisition and display

Note the catadioptric system used [9].



Concentric mosaics acquisition

- Rotate off-center camera(s), e.g., [10].



LADAR 3D acquisition

- Camera uses pulsed laser [11].
- Per-pixel scene depth determined by difference in ToA (time of arrival of pulses).

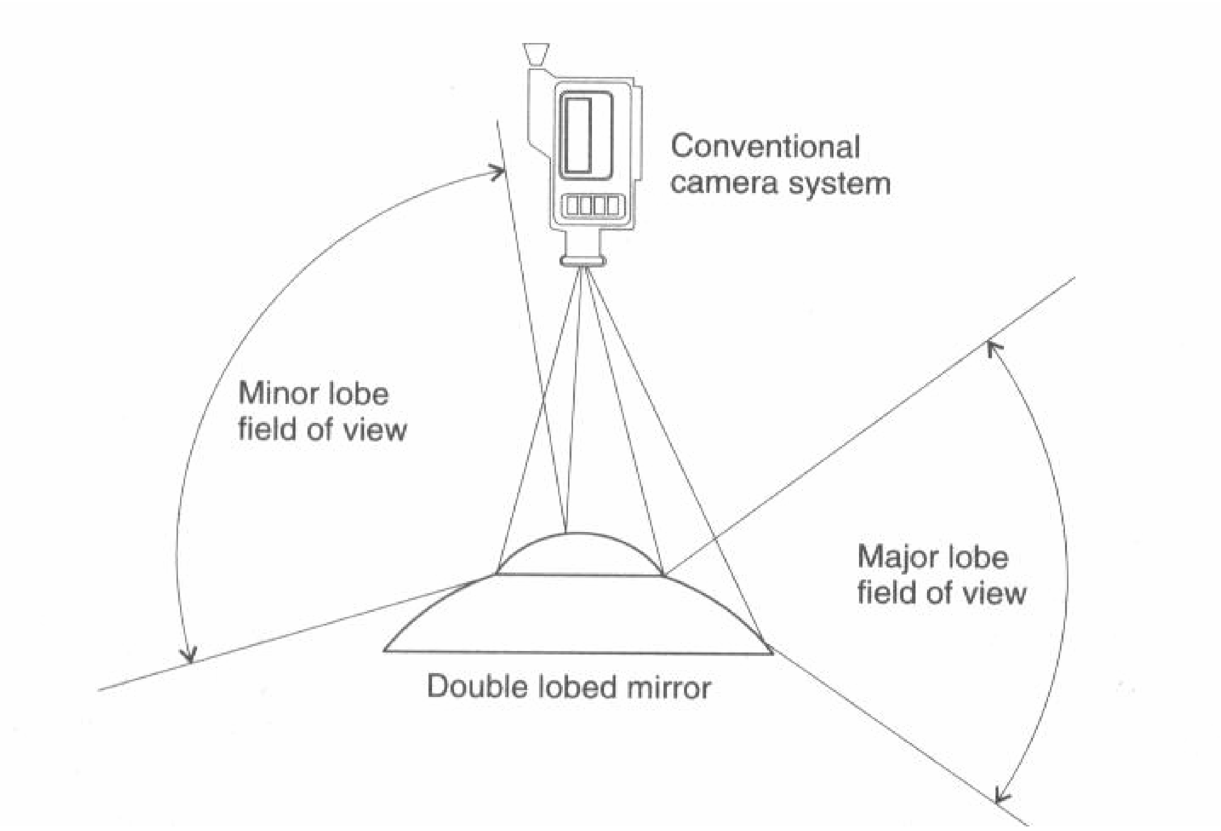
“At video frame rates (30Hz) their solid-state flash LADAR system is able to simultaneously measure the distance to every point in the scene by recording the time-of-flight of a laser pulse. At full speed the camera collects 500,000 range points per second using a 1.57um eye-safe laser that has been successfully tested at distances greater than 5km.The entire system is the size of a shoebox and weighs only 12 pounds.
It uses less than 60 watts of power and can be controlled from a laptop.” [11]

LADAR camera imagery

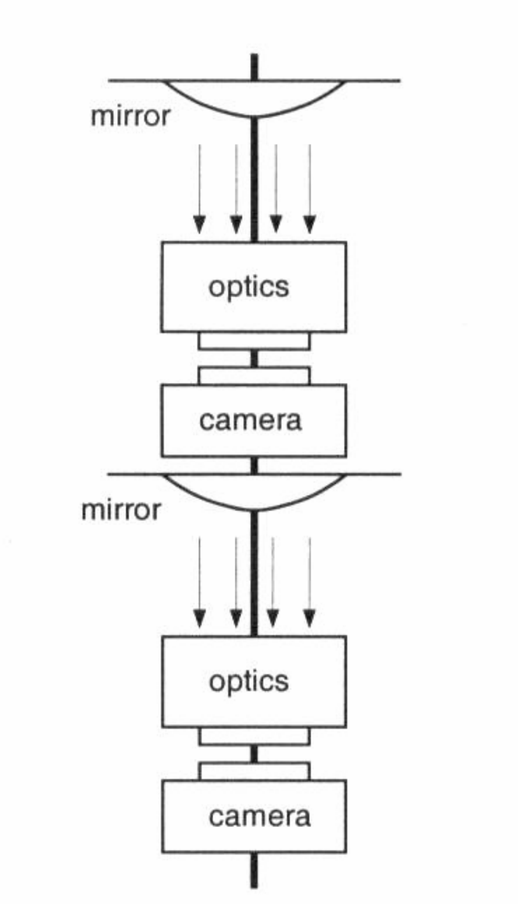
Stereo video

- E.g., interlaced left/right views.

Wide-angle stereo



a)



b)

Plenoptic signal processing

Plenoptic sampling [1]

- Light Fields over-sample to counter aliasing.
  - More intensive acquisition
  - More storage
  - More redundancy (which can be exploited in coding)

- Lumigraphs use approximate geometry to improve rendering performance.
  - But geometry is hard to get for real scenes

- Fundamental problem: plenoptic sampling
  - Interplay of factors: scene depth and texture, number of sample images, rendering resolution

- How many samples and how much depth and texture information are needed to reconstruct an anti-aliased, continuous representation of the plenoptic function for a given resolution?

Plenoptic sampling [2]

- Let $l(u,v,s,t)$ - Continous light field

  $p(u,v,s,t)$ - Sampling function (e.g., rectangular sampling lattice)

  $r(u,v,s,t)$ - Combined filtering and interpolation low-pass filter

  $i(u,v,s,t)$ - Reconstructed light field
- In the spatial domain [12]

$$i(u,v,s,t) = r(u,v,s,t) \circledast [l(u,v,s,t)\, p(u,v,s,t)]$$

- For example, for a rectangular sampling lattice

$$l_s(u,v,s,t) = l(u,v,s,t) \sum_{n_1,n_2,k_1,k_2 \in Z} \delta(u - n_1 \Delta u)\delta(v - n_2 \Delta v)\delta(s - k_1 \Delta s)\delta(t - k_2 \Delta t)$$

- In the frequency domain,

$$L_s(\Omega_u,\Omega_v,\Omega_s,\Omega_t) = \sum_{m_1,m_2,l_1,l_2 \in Z} L(\Omega_u - \frac{2\pi m_1}{\Delta u}, \Omega_v - \frac{2\pi m_2}{\Delta v}, \Omega_s - \frac{2\pi l_1}{\Delta s}, \Omega_u - \frac{2\pi l_2}{\Delta t})$$

Plenoptic sampling [3]

- Find *r(u,v,s,t)* for anti-aliased light field reconstruction.
- Minimum plenoptic sampling rate [12] is a function of:
  - Minimum and maximum depth in the scene (regardless of depth variation between bounds)
  - Highest frequency of the light field signal, determined by the scene texture distribution
  - Resolution of the sampling camera
  - Resolution of the rendering (rendering at higher resolution is wasteful)

View entropy

- What constitutes a good view of a scene?
  - No consensus, difficult to define
  - Has something to do with the amount of information about the scene.
  - Could use information theory concepts: “viewpoint entropy”

- Possible basic elements of a viewpoint quality function:
  - Number of faces of objects seen
  - Size of projected area of faces

- Shannon’s entropy: $H(X) = -\sum_{i=1}^{n} p_i \log p_i,$

  where X takes values from source alphabet {a1,a2,...,an}, and $p_i = P\{X = a_i\}$

- Viewpoint entropy [13], for a sphere of directions centered at viewpoint:

$$I = -\sum_{i=1}^{n} \frac{A_i}{A_t} \log \frac{A_i}{A_t},$$

where n is the number of faces in the scene,

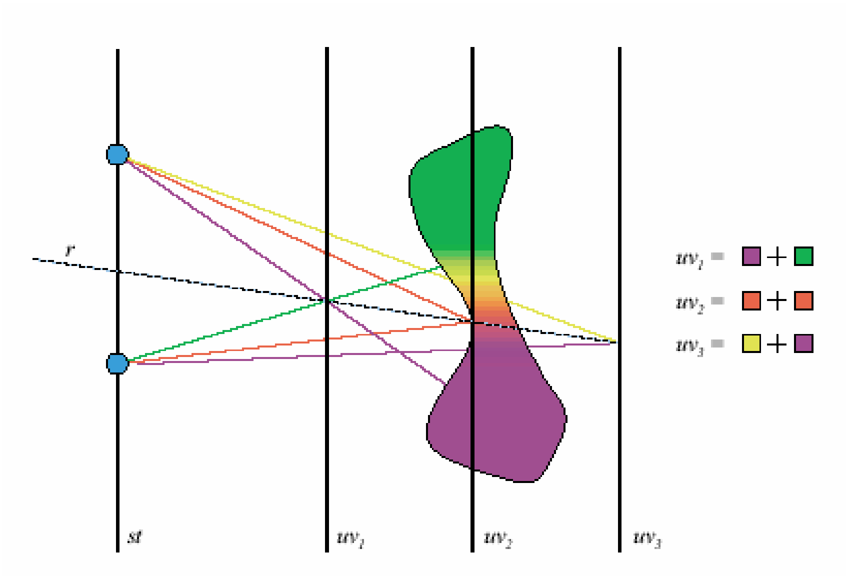
A_i is the projected area of face *i* over the sphere

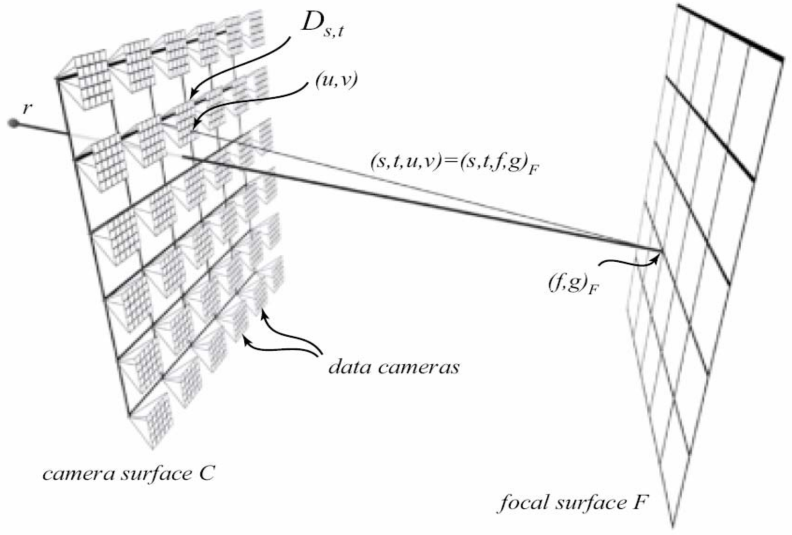
A_t is the total area of the sphere, A_0 corresponds to background

Dynamically-reparameterized Light Fields [1]

- Interactive, variable depth-of-field and variable focus.
  - In original Light Field focal plane is fixed
  - A Lumigraph uses depth correction to improve rendering

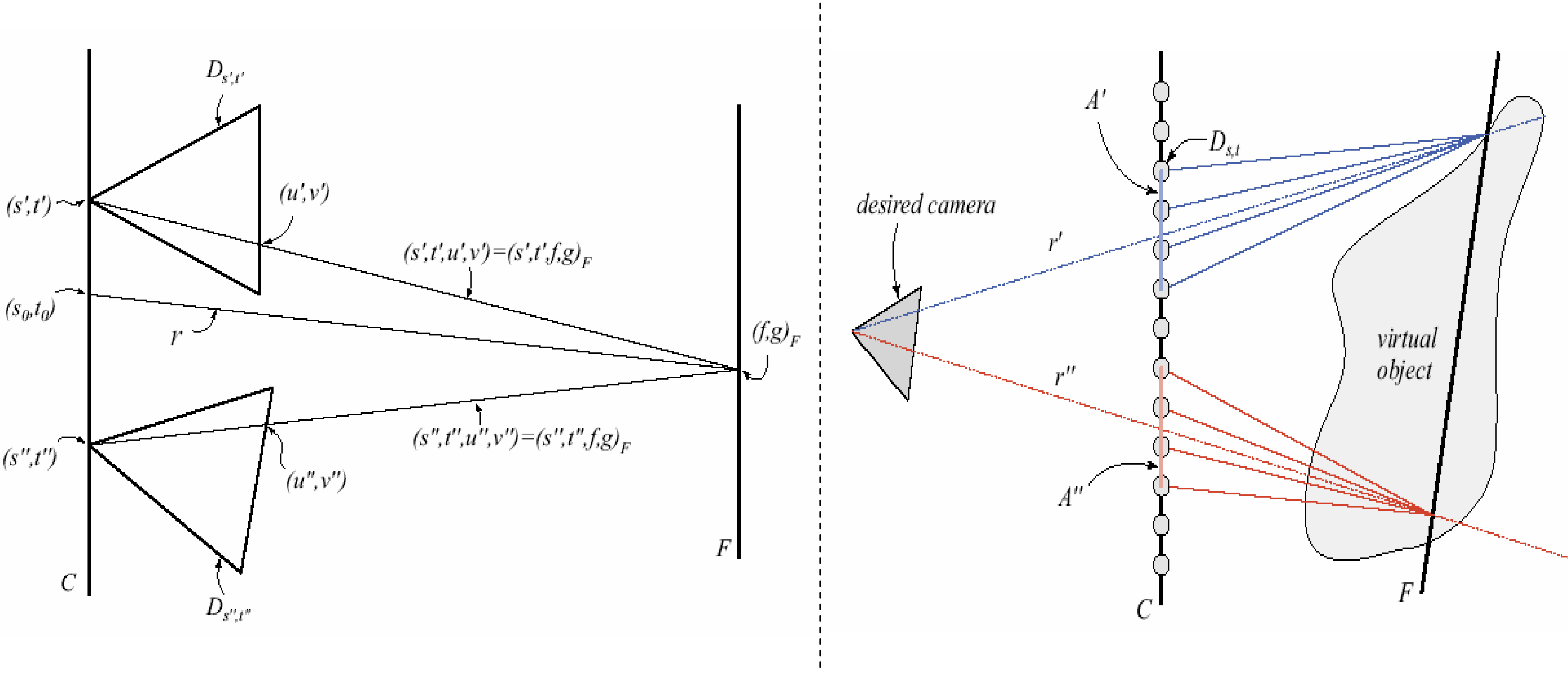
- Parameterization [14] using a virtual camera surface and focal surface
  - Focal plane can be “swept” through the scene to bring in focus various portions ➔ render by re-sampling the light field accordingly
  - Can have two or more distinct regions that are in focus simultaneously





Dynamically-reparameterized Light Fields [2]

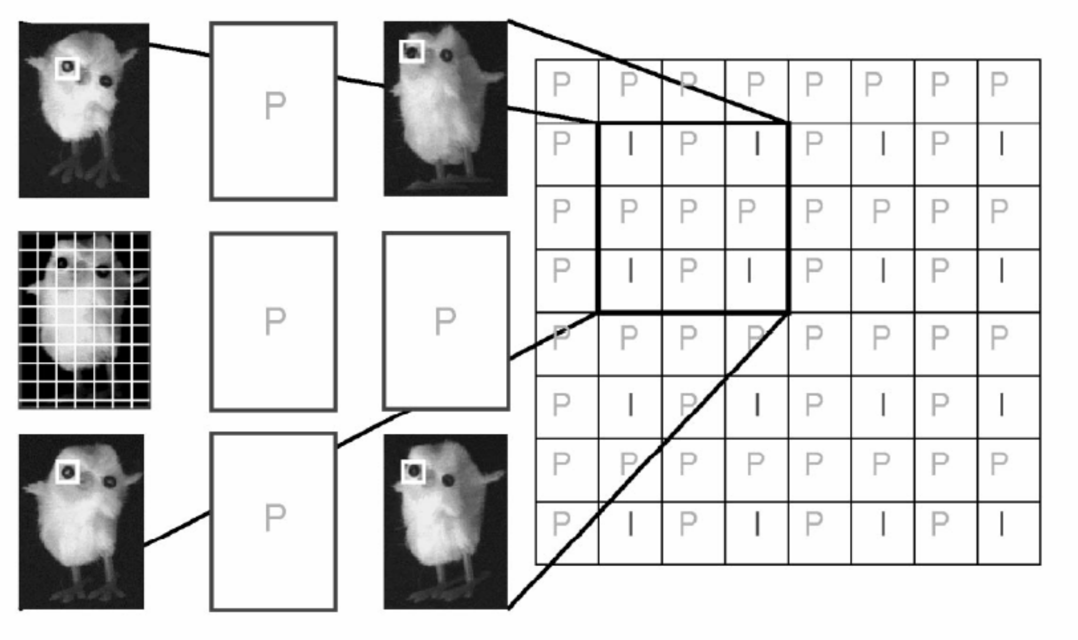
- Synthetic aperture



*from [13]

Light Field coding – Early methods

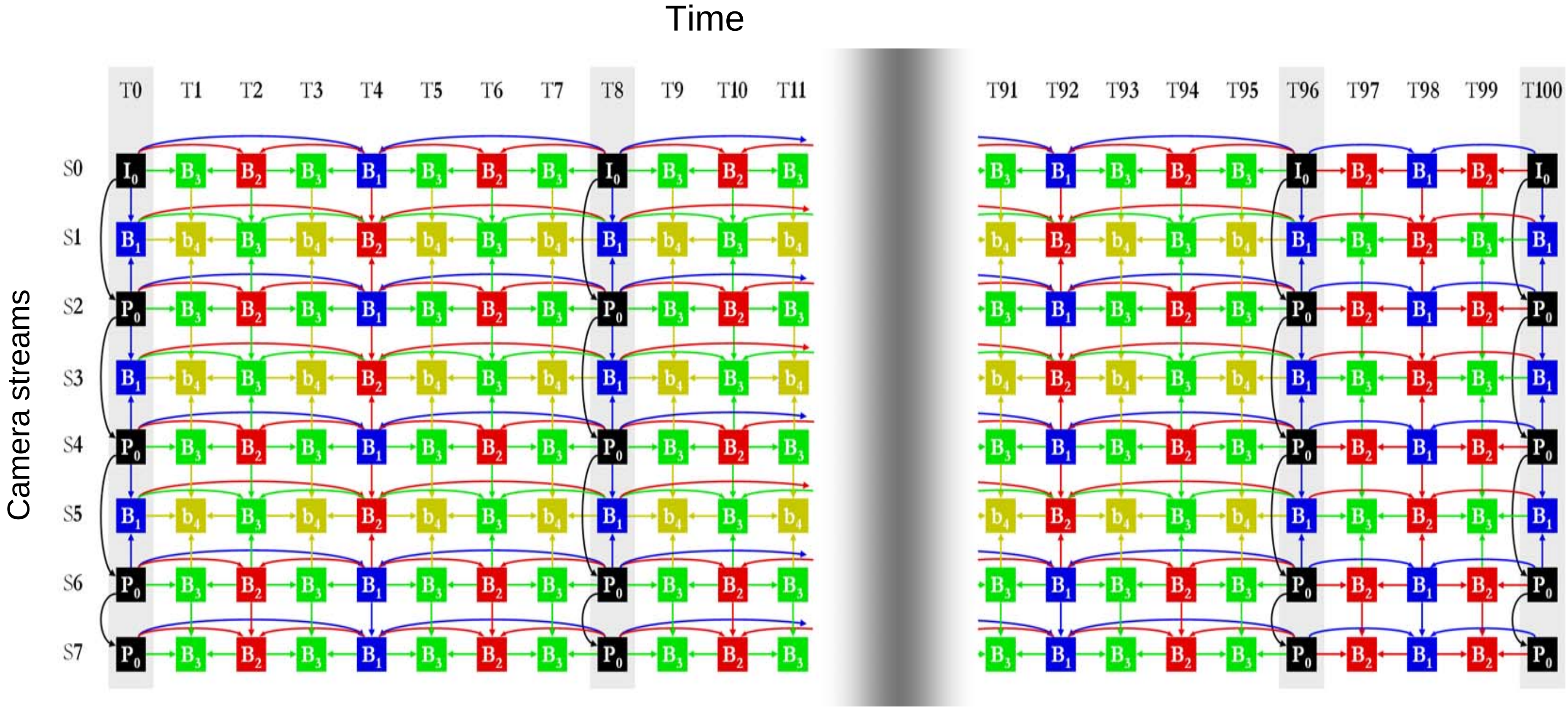
- JPEG coding of each image in the image plane of a light field slab.
- Vector quantization + Lempel Ziv entropy coding (gzip) [4]
- Spatial Intra (I) and Predicted (P) pictures [15]
  - Disparity compensation



P	P	P	P	P	P	P	P
P	I	P	I	P	I	P	I
P	P	P	P	P	P	P	P
P	I	P	I	P	I	P	I
P	P	P	P	P	P	P	P
P	I	P	I	P	I	P	I
P	P	P	P	P	P	P	P
P	I	P	I	P	I	P	I

Light Field coding

- Inter-view prediction structure based on AVC, using hierarchical B pictures.



	T0	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T91	T92	T93	T94	T95	T96	T97	T98	T99	T100
S0	I_0	B_3	B_2	B_3	B_1	B_3	B_2	B_3	I_0	B_3	B_2	B_3	B_3	B_1	B_3	B_2	B_3	I_0	B_2	B_1	B_2	I_0
S1	B_1	b_4	B_3	b_4	B_2	b_4	B_3	b_4	B_1	b_4	B_3	b_4	b_4	B_2	b_4	B_3	b_4	B_1	B_3	B_2	B_3	B_1
S2	P_0	B_3	B_2	B_3	B_1	B_3	B_2	B_3	P_0	B_3	B_2	B_3	B_3	B_1	B_3	B_2	B_3	P_0	B_2	B_1	B_2	P_0
S3	B_1	b_4	B_3	b_4	B_2	b_4	B_3	b_4	B_1	b_4	B_3	b_4	b_4	B_2	b_4	B_3	b_4	B_1	B_3	B_2	B_3	B_1
S4	P_0	B_3	B_2	B_3	B_1	B_3	B_2	B_3	P_0	B_3	B_2	B_3	B_3	B_1	B_3	B_2	B_3	P_0	B_2	B_1	B_2	P_0
S5	B_1	b_4	B_3	b_4	B_2	b_4	B_3	b_4	B_1	b_4	B_3	b_4	b_4	B_2	b_4	B_3	b_4	B_1	B_3	B_2	B_3	B_1
S6	P_0	B_3	B_2	B_3	B_1	B_3	B_2	B_3	P_0	B_3	B_2	B_3	B_3	B_1	B_3	B_2	B_3	P_0	B_2	B_1	B_2	P_0
S7	P_0	B_3	B_2	B_3	B_1	B_3	B_2	B_3	P_0	B_3	B_2	B_3	B_3	B_1	B_3	B_2	B_3	P_0	B_2	B_1	B_2	P_0

* From Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG, JVT-T208, July 2006 [16]

Light Field coding– Surface light field

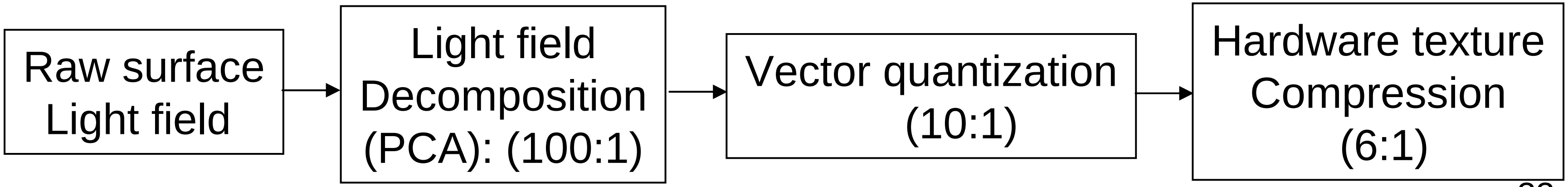
- Discretize surface light field onto triangular mesh, decompose vertex-based light field using PCA [6].
  - Use active imaging method to get object geometry.

$$f^{v_j}(r,s,\theta,\varphi) \approx \sum_{k=1}^{K} \overset{\text{Surface map}}{g_k^{v_j}(r,s)}\, \overset{\text{View map}}{h_k^{v_j}(\theta,\varphi)}$$

- At vertex v_j
- r, s describe position on object
- θ ,φ describe irradiance orientation

$$F^{v_j} = \begin{bmatrix} f^{v_j}(r_1,s_1,\theta_1,\varphi_1) & \cdots & f^{v_j}(r_1,s_1,\theta_N,\varphi_N) \\ \vdots & \ddots & \vdots \\ f^{v_j}(r_M,s_M,\theta_1,\varphi_1) & \cdots & f^{v_j}(r_M,s_M,\theta_N,\varphi_N) \end{bmatrix}$$

$$\tilde{F}^{v_j} = \sum_{k=1}^{K} u_k v_k^T$$



Raw surface Light field

Light field Decomposition (PCA): (100:1)

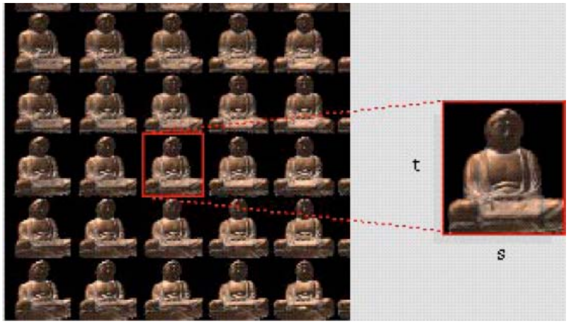
Vector quantization (10:1)

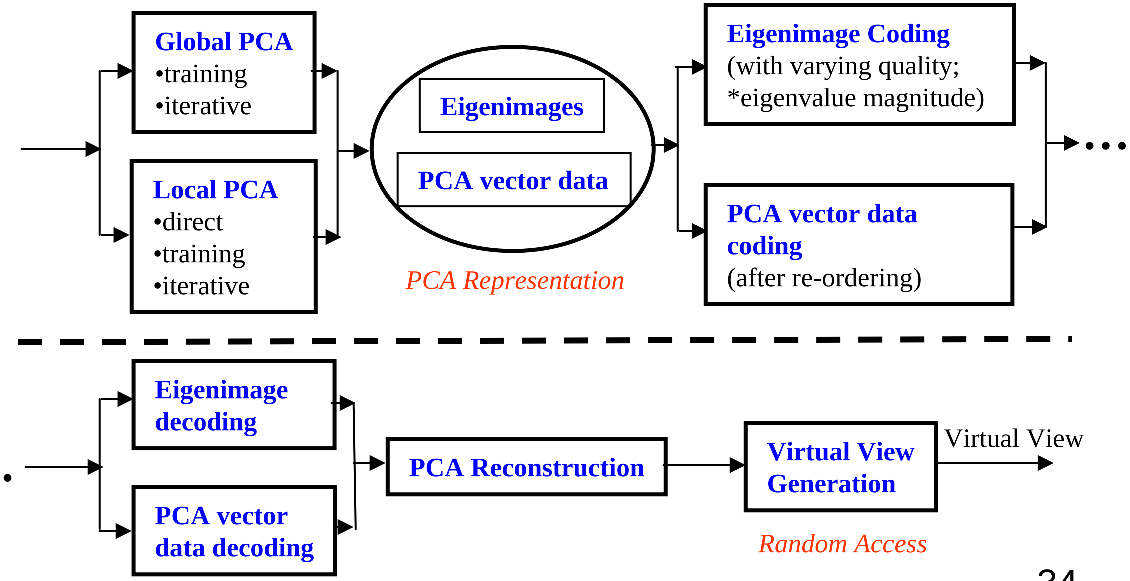
Hardware texture Compression (6:1)

Light Field representation and coding

- Coding and rendering light fields:
  - Need for high compression suggests use of predictive coding (temporal- and disparity-wise), thereby increasing inter-picture dependency
  - Need for random access suggests the use of intra-coding techniques

- Compromise by exploiting statistical inter-view redundancy: use Principal Component Analysis-based approaches [17].
  - Code subspace description (eigenimages) and transformed images.





Global PCA
•training
•iterative

Local PCA
•direct
•training
•iterative

Eigenimages

PCA vector data

PCA Representation

Eigenimage Coding
(with varying quality;
*eigenvalue magnitude)

PCA vector data coding
(after re-ordering)

…

…

Eigenimage decoding

PCA vector data decoding

PCA Reconstruction

Virtual View Generation

Virtual View

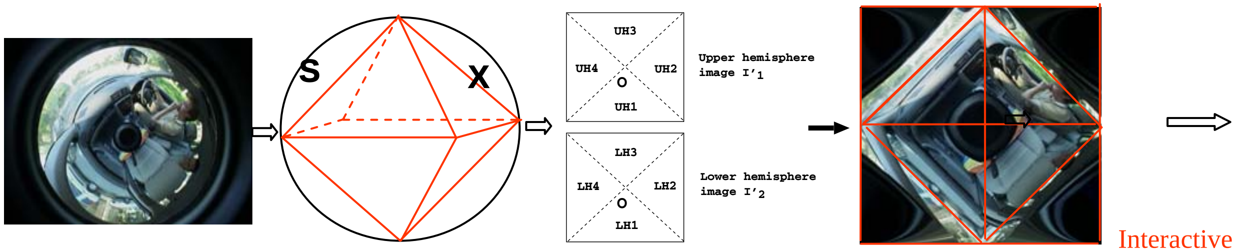
Random Access

Omnidirectional representation and coding

- There are many types of omnidirectional acquisition systems
    - Introduce an intermediate, universal representation to which all raw images can be mapped [18].
        - e.g., for facilitating coding using standard video codecs (MPEG)

- Universal format
    - Project to virtual sphere centered at SCOP of catadioptric system
    - Project onto faces of inscribed polyhedron: octahedron [3DAV[18]], dodecahedron [3DAV]
    - Project and pack polyhedron faces in the plane

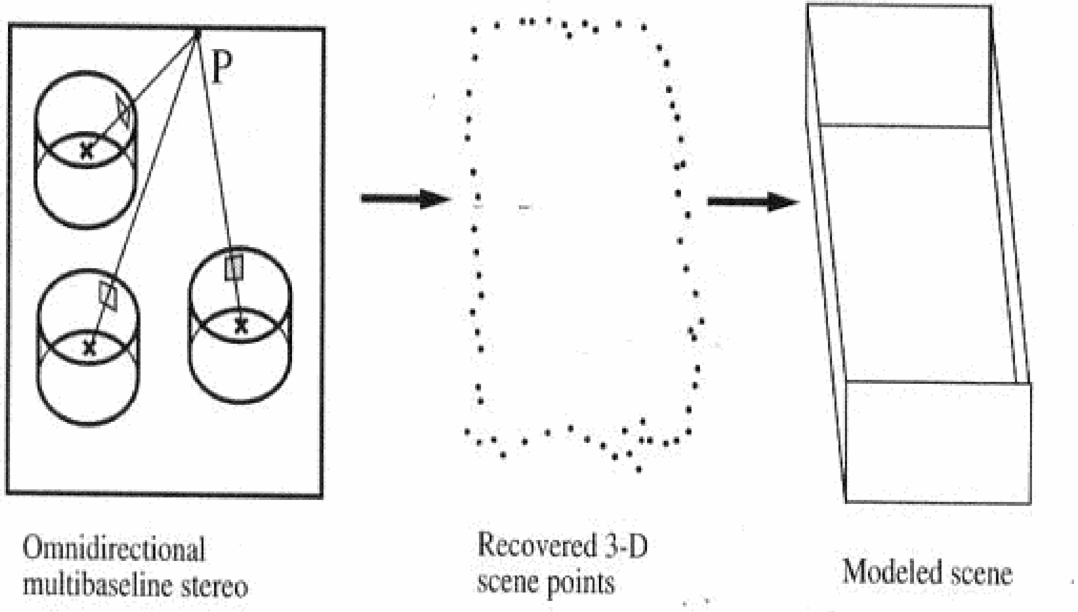
- What is the best representation for coding such images (e.g., interesting questions on prediction in omnidirectional video)?



Interactive
view generation
(texture mapping)

3D reconstruction using omnidirectional images

- Omnidirectional systems can be used for 3D reconstruction [19].
  - Advantage: generate 3D data for a wide field of view (FOV)--no need to do multiple depth maps merging as in narrow FOV (error prone).



- At each camera location, capture a cylindrical panorama.
- Use stereo to extract 3D structure of scene.

Summary

- Plenoptic acquisition, processing, coding, transmission, and display is a rich research area:
  - Inter-disciplinary: e.g., computer vision, optics, signal processing, computer graphics
  - Many challenges remain; new techniques await discovery

- Increased exposure in international standards (e.g., MPEG 3DAV, MVC), technical conferences.

- Number of feasible applications is increasing:
  - 3DTV
  - Light Field photography (consumer ?)
  - Panoramic viewing
  - Tele-presence
  - Active cameras (e.g., LADAR).

- Participate!

Selected references [1]

[1] A. Gershun, "The Light Field", *J. Math and Phys.,* vol.18, pp. 51-151, Moscow, 1939.

[2] E.H. Adelson and J.R. Bergen, "The plenoptic function and the elements of early vision", *Computation Models of Visual Processing*, pp.3-20, MIT Press 1991.

[3] H.-Y. Shum and S.B. Kang, "A review of image-based rendering techniques", *SPIE Visual Communications and Image Processing (VCIP),* pp. 2-13, 2000.

[4] M. Levoy and P. Hanrahan, "Light Field rendering", *Proc. ACM SIGGRAPH*, pp. 31-42, ACM Press,1996.

[5] S.J. Gortler et al., "The Lumigraph", *Proc. ACM SIGGRAPH*, pp.43-54, ACM Press,1996.

[6] W.-C. Chen, J.-Y. Bouguet, M.H. Chu, and R. Grzeszczuk, "Light Field mapping: efficient representation and hardware rendering of surface light fields", *ACM SIGGRAPH* 2002.

[7] Stanford Computer Graphics Laboratory, Plenoptic camera.

[8] E.H. Adelson and J.Y.A. Wang, "Single Lens Stereo with Plenoptic Camera", *PAMI*, vol. 14, No. 2, pp. 99-106, 1992.

[9] FullView cameras, Bell Labs, 1995.

[10] H.-Y. Shum and L.-W. He, "Rendering with concentric mosaics", *ACM SIGGRAPH* 1999.

[11] LADAR camera, *Advanced Scientific Concepts*,advancedscientificconcepts.com.

[12] J.-X. Chai, S.-C. Chan, H.-Y.Shum, and X. Tong, "Plenoptic sampling", *Conference on Computer Graphics and Interactive Techniques*, 2000.

Selected references [2]

[13] P.-P. Vasquez, M. Feixas, M.Sbert, and W. Heidrich, “Viewpoint selection using viewpoint entropy”, *Proceedings of the Vision, Modeling, and Visualization (VMV) Conference,* 2001.

[14] A. Isaksen, L. McMillan, and S.J. Gortler, “Dynamically reparameterized light fields”, *ACM SIGGRAPH 2000*, pp. 297-306, 2000.

[15] M. Magnor and B. Girod, “Adaptive block-based light field coding." *Proc. 3rd International Workshop on Synthetic and Natural Hybrid Coding and Three-Dimensional Imaging IWSNHC3DI '99*, pp. 140-143, September 1999.

[16] Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG, *Document JVT-T208*, July 2006.

[17] D. Lelescu and F. Bossen, “Representation and coding of light field data”, *Graphical Models*, vol. 66, pp. 203-225, July 2004.

[18] D. Lelescu and F. Bossen, “Representation of panoramic and omnidirectional images”, *Document M9273*, ISO/IEC JTC 1/SC 29/WG 11, Awaji, Japan, 2002.

[19] S.B. Kang and R. Szeliski, “3-D Scene data recovery using omnidirectional multibaseline stereo”, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1996.

[20] D. Lelescu and F. Bossen, “Additional requirements for 3D AV coding”, *Document M8482*, ISO/IEC JTC 1/SC 29/WG 11, Klagenfurt, Austria, 2002.