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, omni-directional cameras, QuicktimeVR panoramas.
Introduction
Early inquiries

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

- 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.
The plenoptic function [Latin, plenum] was introduced formally in [2] in 1992. It describes all light information collected at a point in space-time. By fixing various parameters in the plenoptic function, one obtains different, more restrictive representations.

The plenoptic function is originally a 7D function,

$$f(V_x, V_y, V_z, \theta, \phi, \lambda, t)$$

where

- $V_x, V_y, V_z$ - viewpoint coords.
- $\theta, \phi$ - ray direction
- $\lambda$ - wavelength
- $t$ - time
Challenges: a first set

➢ Plenoptic image acquisition
   - Sensor design, calibration, synchronization
   - 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)
There is a continuum of methods spanning the IBR and CG fields

<table>
<thead>
<tr>
<th>Rendering with no geometry (&quot;true&quot; IBR)</th>
<th>Rendering with implicit geometry</th>
<th>Rendering with explicit geometry</th>
</tr>
</thead>
<tbody>
<tr>
<td>Light Fields</td>
<td>Lumigraph</td>
<td></td>
</tr>
<tr>
<td>Concentric mosaics</td>
<td>View interpolation</td>
<td>Layered-depth images</td>
</tr>
<tr>
<td>Mosaics, cylindrical, spherical panoramas</td>
<td></td>
<td>Texture-mapped models</td>
</tr>
<tr>
<td>View morphing</td>
<td></td>
<td>View-dependent texture</td>
</tr>
</tbody>
</table>

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)

*adapted from [3]
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
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]
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.
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

- 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 \( l(u,v,s,t) \) - Continuous 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]
  \[
i(u,v,s,t) = r(u,v,s,t) \ast [l(u,v,s,t) p(u,v,s,t)]
\]
- 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,
  \[
  L_s(\Omega_u, \Omega_v, \Omega_s, \Omega_t) = \sum_{m_1,m_2,l_1,l_2 \in Z} L(\Omega_u - \frac{2\pi m_1}{\Delta u}, \Omega_v - \frac{2\pi m_2}{\Delta v}, \Omega_s - \frac{2\pi l_1}{\Delta s}, \Omega_t - \frac{2\pi l_2}{\Delta t})
\]
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 \( \{a_1,a_2,\ldots,a_n\} \), 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

UCSB, Nov. 9, 2006
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
Light Field coding

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

* From Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG, JVT-T208, July 2006 [16]
Discretize surface light field onto triangular mesh, decompose vertex-based light field using PCA [6].

- Use active imaging method to get object geometry.

\[
f^{v_j}(r, s, \theta, \varphi) \approx \sum_{k=1}^{K} g^{v_j}_k(r, s) h^{v_j}_k(\theta, \varphi)
\]

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

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

### Light Field coding – Surface light field

- At vertex \(v_j\)
- \(r, s\) describe position on object
- \(\theta, \varphi\) describe irradiance orientation

- At vertex \(v_j\)
- \(r, s\) describe position on object
- \(\theta, \varphi\) describe irradiance orientation
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.

---

UCSB, Nov. 9, 2006
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)?
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]


Selected references [2]


