Light Field: A Quest for the Perfect Picture

“The Holodeck”

Dr. Panos Nasiopoulos
Hypervsn – holographic solutions

https://www.youtube.com/watch?v=4a0_Vb0e1ek
Capture and display what the human eye can see

Anaglyph 3D Glasses
Polarized Glasses
Shutter Glasses

Glasses
Capture and display what the human eye can see

Multiview displays

Image source: DIMENCO

Image source: Alloscopy

FOCUS & CONVERGENCE
Capture and display what the human eye can see

4K: improved resolution
HDR: High Dynamic Range
A quest for the perfect picture

CAPTURE INFORMATION THAT CAN HELPS US APPROXIMATE HUMAN PERCEPTION
We live in a Visual World

90% of information we process is visual
We live in a Visual World

90% of information we process is visual

50% of our brain actively processes visual content
We live in a Visual World

90% of information we process is visual

50% of our brain actively processes visual content

We process visuals 60,000x faster than text or audio
We live in a Visual World

90% of information we process is visual

50% of our brain actively processes visual content

We process visuals 60,000x faster than text or audio

We retain 80% of what we see
FIRST PHOTO TAKEN BY A CAMERA - 1826
FIRST COLOUR PHOTOGRAPH - 1861
FIRST DIGITAL PHOTOGRAPH - 1957

176 x 176
HOLODECK – 20??
WE SEE THE WORLD BECAUSE OF LIGHT
WE SEE THE WORLD BECAUSE OF LIGHT
WE SEE THE WORLD BECAUSE OF LIGHT
WE FOCUS ON CONVERGING RAYS
WE FOCUS ON CONVERGING RAYS
WE FOCUS ON CONVERGING RAYS
HOW TO RECREATE THE SCENE USING LIGHT FIELD?
WE CREATE INVERSE REFLECTION OF LIGHT
WE CREATE INVERSE REFLECTION OF LIGHT
WE CREATE INVERSE REFLECTION OF LIGHT
WE CREATE INVERSE REFLECTION OF LIGHT
WE CREATE INVERSE REFLECTION OF LIGHT
WE CREATE INVERSE REFLECTION OF LIGHT
WE CREATE INVERSE REFLECTION OF LIGHT
WE CREATE INVERSE REFLECTION OF LIGHT
WE CREATE INVERSE REFLECTION OF LIGHT
WE CREATE INVERSE REFLECTION OF LIGHT
WE CREATE INVERSE REFLECTION OF LIGHT
WE CREATE INVERSE REFLECTION OF LIGHT
WE CREATE INVERSE REFLECTION OF LIGHT
A picture is worth a 1000 words but …
A picture is worth a 1000 words but …

Sum of brightness
Ideally ....

Concept: every light ray from every point in space
Extremely complex
Difficult to achieve

**Plenoptic function**

\[ L(x, y, z, \theta, \phi, \lambda, t) \]

**5D:** \[ L(x, y, z, \theta, \phi) \]

**4D:** \[ L(x, y, s, t) \]

Light Field Technology

2006, Ren Ng
Current Light Field Capture Systems

Raytrix

Lytro

Microlens

2008
Who is involved?
Light Field Camera – Concept of Sub-Aperture Images

The same color means they come from the same perspective.
Light Field Camera – Concept of Sub-Aperture Images

↑ RAW light field image captured by sensors

9375  (15x625)

↑ Processed sub-aperture images

9375

↑ Single sub-aperture image

625 px

Massive Data

6510  (434)

434 px
Light Field Camera – Perspective Feature
Light Field Camera – Perspective Feature

15X15 different perspectives

Freely changing perspectives
Light Field Camera – Perspective Feature
Light Field Frame
Sub-Aperture Images
More details revealed
- occlusion removal
- depth plane extraction
Better Object Detection
Refocusing: Shift & Sum

Identify the depth layer in all the views
Refocusing: Shift & Sum

Shift the views to match the base view

Base view
Focused on the “Tire”
Focused on the “Small Flower”
Focused on the “Flowers”
Focused on the “Leafs”
Focused on the “other tire”
Microlens depth limitation
Current Light Field Capture Systems

Camera Array

Raytrix

Microlens

Lytro

© Stanford
REFOCUSING – DOES NOT WORK FOR CAMERA ARRAY CONTENT
REFOCUSING – DOES NOT WORK FOR CAMERA ARRAY CONTENT
Due to the longer baseline
Shifting only aligns to the sharpest part

- Occluded objects/new details creates ghosting
REFOCUSING & PERSPECTIVE VIEWING IN ENTERTAINMENT

Applications:
- Live sports
- Interactive Movies
- Games

Pros:
- Every Sub-aperture image is captured from slightly different vantage point
- Better depth estimation than 3D
- Many focus points
- Better Object separation/recognition
ENTERTAINMENT – LIGHT FIELD
LIGHT FIELD IN AUGMENTED REALITY

Changing the AR/VR Landscape

- Accurate Object Recognition
- Accurate Depth Estimation
- Accurate Overlay

Magic Leap has raised more than $2.7 billion in funding in 2 years (AT&T is one of the investors)
LIGHT FIELD IN AUGMENTED REALITY

- Existing AR: Challenging precise overlay of synthetic augmentations on real-life content

Segmentation, object identification, depth estimation have been extremely difficult challenges for real-time applications.
Combine eye trackers with light field AR

Eye trackers can check where the viewer looks (heat map)

Light field properties are used to identify object of interest and focus only on that plane and location

Visual information from light field camera is analyzed to identify area

Vase (30 BC)
LIGHT FIELD IN AUGMENTED REALITY

Medical
LIGHT FIELD IN AUGMENTED REALITY
Cameras
- Detect traffic lights, read road signs, keep track of the position of other vehicles, look for pedestrians and obstacles.

Lidar
- Bounces pulses of light off the surroundings. They are analyses to identify lane markings and the edges of roads.

Radar sensors
- Monitor position of other vehicles nearby. Already used in adaptive cruise control systems.
For now, autonomous vehicles/driving are limited to some places, isolated streets, specific controlled environment (depth information may suffice)

Why? Limited Visual Information
LIGHT FIELD IN AUTONOMOUS DRIVING

But the long term goal, which is 5 years and beyond, will expand these to a much broader application which will need our visual solutions to step up…
But the long term goal, which is 5 years and beyond, will expand these to a much broader application which will need our visual solutions to step up…

Capture
- light intensity and directional information,
- focus and depth properties, and
- visual cues.
The rich, immersive video information provided by Light Field will help AI “act” as close as possible to a human.
LIGHT FIELD IN DIGITAL HEALTH

Accurate Less Intrusive Endoscopy

A regular surgical camera

View from the camera
Accurate Less Intrusive Endoscopy

Light Field Display

Light Field Camera
5G enables real-time remote surgery

© TransEnterix
LIGHT FIELD IN DIGITAL HEALTH

Non-Intrusive Monitoring
LIGHT FIELD IN DIGITAL HEALTH

Labelling

Collection ➔ Extraction ➔ Processing ➔ Training

Walking ➔ Getting Back Up ➔ Sitting on the mat ➔ Fallen

Manually labelled using labelr2.py
LIGHT FIELD IN DIGITAL HEALTH

Training - LRCN

Long Term Recurrent Convolutional Network
- Use a CNN to extract key features from frames
- Feed into LSTM to gain temporal information
Light Field information will improve this accuracy, since richer visual information will be processed by the system.

This type of visual information will allow us to accurately track behaviour and determine changes that may be associated with various medical conditions, for early diagnosis and prevention.
DIGITAL HEALTH & COMMUNICATIONS - Challenges

- Bandwidth
- Real-time streaming
- Security
Compression
Practical Light Field Camera arrays
Compression IBP frames

Intra prediction

Motion compensation

Previous frame only

Trade-off between overhead and residuals
LIGHT FIELD – MASSIVE DATA: COMPRESSION

Conventional Video

Light Field Video

Inter-view Correlation

Temporal Correlation
LIGHT FIELD – Existing Compression Methods

LF-MVC
(Wang Et al. 2016)
Is Top Left the Best Place to get started?

**LF-MVC**
(Wang Et al. 2016)

Encode the central frame as I frame ✓
Again only vertical or horizontal references ✗
38.18% BD-Rate of LF-MVC (5 x 5 Views) ✓
Does not scale well ✗

**Khoury’s Method**
(Khoury Et al. 2018)
LIGHT FIELD – Existing Compression Methods

Full Scheme
(Avramelos et al. 2019)

- Maximizes B frames so compression efficient ✓
- Again I frame at a corner ✗
- Diagonal references ✗
- Predicting frames from far ✗
- Performs 24% better than LF-MVC and 15% worse than Khoury’s method
- Does not scale well ✗
SSIM heatmap for Chess Pieces (5x5)
Empirical results over 9 test sequences

<table>
<thead>
<tr>
<th>Avg of all views</th>
<th>.982</th>
<th>.985</th>
<th>.985</th>
<th>.984</th>
<th>.976</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>.983</td>
<td>.985</td>
<td>.986</td>
<td>.985</td>
<td>.981</td>
</tr>
<tr>
<td></td>
<td>.984</td>
<td>.99</td>
<td><strong>.991</strong></td>
<td>.989</td>
<td>.985</td>
</tr>
<tr>
<td></td>
<td>.984</td>
<td>.987</td>
<td>.99</td>
<td>.986</td>
<td>.984</td>
</tr>
<tr>
<td></td>
<td>.98</td>
<td>.981</td>
<td>.984</td>
<td>.982</td>
<td>.984</td>
</tr>
</tbody>
</table>
SSIM based Compression Method
Universal Pseudo-sequence (UPS) SSIM based Compression Method

Unlike most of the PSB prediction structures and coding orders discussed above, the proposed Universal Pseudo-sequence (UPS) based structure takes full advantage of both horizontal and vertical correlation among the views.
SSIM based Compression Method

- Has highest similarity with its neighbors
- SSIM of all the views
- Highest one is I-frame

- SSIM calculation makes smallest possible P frames
- To maintain our structure, next level of P frames are predicted

- B frames with vertical and horizontal references

- Leverage bidirectional correlation
- Scalable to number of views
- Scalable to view arrangements

5x5 views
SSIM based Compression Method

Can be extended to any number of views

Feasible for most camera array structures

4x4 views

5x6 views

Circular

Hexagonal

Rectangular
Performance – Microlens content (5x5)

<table>
<thead>
<tr>
<th>Resolution</th>
<th>1920x1080</th>
</tr>
</thead>
<tbody>
<tr>
<td>FPS</td>
<td>30; 300 frames</td>
</tr>
<tr>
<td>Baseline</td>
<td>4mm</td>
</tr>
</tbody>
</table>

Raytrix R8
Performance – Animated content (3x3)

Resolution | 1024x436
FPS          | 20; 100 frames
Performance – Camera Array content (3x5) and (4x4)

Resolution | 1920x1080
FPS         | 30; 300 frames
Baseline    | 6 to 10 cm
Light Field Video Compression for Random-access efficiency
Light Field – Random Access matters

Khoury’s Method

Worst Frame to Decode: 6

LF-MVC

Worst Frame to Decode: 11

Full Scheme

Worst Frame to Decode: 18
Proposed Methods

Reduced worst case random-access from 6 to 4

Diagonal Reference Based Prediction structure

Random-access reduces to 2 from 6

SSIM Assisted Diagonal Reference Based Prediction structure

Highest SSIM

Increases compression efficiency by 15%

Average random-access increases from 2.56 to 2.72

Reduced worst case random-access from 6 to 4
**Random Access performance**

<table>
<thead>
<tr>
<th>Method</th>
<th>Random-access Complexity</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Average</td>
<td>Worst-case</td>
</tr>
<tr>
<td>LF-MVC</td>
<td>4.92</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>Khoury’s</td>
<td>3.2</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>Full scheme</td>
<td>6.08</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td><strong>Diagonal reference</strong></td>
<td><strong>2.56</strong></td>
<td><strong>4</strong></td>
<td></td>
</tr>
<tr>
<td>SSIM assisted RAE</td>
<td>2.72</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>UPS</td>
<td>3.84</td>
<td>8</td>
<td></td>
</tr>
</tbody>
</table>

Best Random-access
Best Random-access to compression trade-off
Refocusing
The dense camera array consists of 25 camera modules fixed on a stainless steel stand.
The video content is captured on the University campus, with the scene including walking people, bicycles, vehicles and buildings as background.
The frames from each camera are not perfectly aligned.
Our Light Field Camera array – Outdoors content

Color differences exist in the frames captured by different camera.

Camera1  Camera2
Different types of geometry distortion for single camera and the camera array.

Our Light Field Camera array – Outdoors content
Problems - Geometry Distortion (2)

Different types of geometry distortion for single camera and the camera array.

This type of distortion is described by **extrinsic** parameters.

Rotation and translation between each two camera pairs.
Different types of geometry distortion for single camera and the camera array.

Images captured by 5 cameras in the SAME row. Lens distortion is already fixed.

Multiple camera distortion

Rotation and translation between cameras

This type of distortion is described by **extrinsic** parameters.
Performance – Visualization

Visual comparison before and after calibration and correction
Quantitative comparison before and after calibration/correction

The bitrate vs. PSNR at four different QP levels (25, 28, 30, and 33)

On average, for the same bitrate the objective visual quality (PSNR) increases by 0.519 dB, while for the same quality the bitrate savings are 30.275%.
Light Field View Synthesis
Plenoptic Camera - challenges

+ High *Angular* Resolution

- Low *Spatial* Resolution
Huge amount of data to transmit
Network for View Synthesis

Sparsely Sampled SA-LF Images

Spatial Feature Extraction

Angular Feature Extraction

Macro-pixel LF Images

Angular Features

Angular Features

Conv 1x1

ReLU

+

Up-sampling

Conv 1x1

Sub-pixel

Macro-pixel LF Images

Densely Sampled SA-LF Images
LF Data Representation

3) Epipolar Image Plane (EPI)
Improving Up-sampled Quality

Low-resolution Sup-aperture Images → Bi-cubic Up-sampling → Low-resolution Up-sampled Sup-aperture Images → EPI representation of up-sampled LR images showing discontinuity in disparity → EPI representation of HR images showing continuity in disparity → High-resolution Up-sampled Sup-aperture Images
Improving Up-sampled Quality

Low-resolution Up-sampled Sup-aperture Images → EPI representation of up-sampled LR images showing discontinuity in disparity → EPI representation of HR images showing continuity in disparity → High-resolution Up-sampled Sup-aperture Images
Results
Subjective Results

Side by side evaluation
based on Recommendation BT.500-13 DSIS
Number of subjects: 18
Number of sequences: 10
Subjects asked which method they prefer
Industry & Market
Market
Magic Leap

- Multiplanar AR HMD
- Dual plane per eye
- LCOS SLM
- Waveguide optics

(Head Mounted Display)
Escape the Screen

We're creating a world where content escapes the screen and merges with reality.
Light Field Extended Reality (XR) eyewear which will provide medical professionals lifelike augmented imagery.
CONCLUSION

VISUAL ACQUISITION WILL HAVE MOST DISRUPTIVE IMPACT ON MANY TRENDS & MARKETS
Thank you!

Contact:
panos@ece.ubc.ca
http://dml.ece.ubc.ca