How to achieve dense light field video compression?

Mehrdad Teratani (Associate Professor, Nagoya University)
Xin Jin (Graduate School at Shenzhen, Tsinghua University)
Dense light field

Muti-cam (multiview)  Plenoptic (lenslet)

2004-2008

2012 @ Osaka

NICT

2
End-to-end system for dense light field
Plenoptic Camera type 1.0

Plenoptic 1.0 (e.g., Lytro)

- Spatial resolution = number of microlens.
- Completely defocused relative to main lens image.

## GSST Plenoptic 1.0 Lenslet Data

<table>
<thead>
<tr>
<th>Lenslet Video Data</th>
<th>Resolution: 8656×6075</th>
<th>Color: 24 bits, PNG</th>
<th>Frame rate: 30 fps</th>
</tr>
</thead>
<tbody>
<tr>
<td>M44684: “Toys”</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M46258: “Teapots”</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M44684: “Trees”</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M46258: “Mini-garden”</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Plenoptic Camera 2.0

Single-Focused Plenoptic Camera

- resolution independent of microlenses
- spatial-angular resolution = free tradeoff point.
- Exactly focused on the main lens image.

Designed and made by Tsinghua University
GSST Plenoptic 2.0 Lenslet data

M46259: “Boys”

M46259: “Experimenting”

M49007: “cars”

Colored Lenslet Video Data

Resolution: 4088×3068
Color: 24 bits, BMP
Frame rate: 30 fps
Number of frames: 300

Multiview Video Data

Resolution of each view: 926×672
Views: 5×5
Number of frames: 300
Plenoptic Camera 2.0

Multi-Focused Plenoptic Camera

- Flexible (3 kind of ML)
- resolution independent of microlenses
- spatial-angular resolution = free tradeoff point.

"Raytrix," [https://www.raytrix.de/](https://www.raytrix.de/).
Nagoya University Lenslet data
Plenoptic 2.0

Colored Lenslet Video
Resolution: \textbf{2048x2048 pixels}
Color: 24 bits PNG, and YUV420
Frame rate: 30 fps
Number of frames: 300-400
Camera parameters (SDK output)
Raytrix R5-C- GigE-F2.4 (color)
Main Lens: LMVZ166HC (Kowa)

NagoyaFujita
Fixed Camera
Horizontal view
M47642, M49670

NagoyaOrigami
Fixed Camera
Top view
M47642, M49670

NagoyaDataLeading
Camera on turn table
Horizontal view
M47642, M49670

Tunnel_Train_2
Fixed Camera
Horizontal view
M41787
INRIA Lenslet data
Plenoptic 2.0

Colored Lenslet Video
Resolution: 3840 × 2160 pixels
Color: 24 bits, PNG
Frame rate: 30 fps
Number of frames: 300
Camera parameters (SDK output)

Raytrix R8

End-to-end system for dense light field

N18446: Exploration Experiments and Common Test Conditions for Dense Light Fields
Conversion from Lenslet to Multiview

- Plenoptic 1.0
  - Lenslet Video
  - Invertible conversion
    - Multiview Video
    - Non-Invertible conversion
    - Multiview Video
    - Introduce artifacts

- Plenoptic 2.0
  - Lenslet Video
  - Invertible conversion
    - Multiview Video
    - Non-Invertible conversion
    - Multiview Video
Plenoptic 1.0 Data Conversion Tool

Colored RAW Image
(7728 (H) x 5368 (V))

Raw Sensor Data (LL)
Non-invertible Subaperture Images (MV)
(625 (H) x 434 (V))

Devignetting & Demosaicing
Rotation and Scaling

Resampling
Slicing

Aligned Lenslet Image
(8656 (H) x 6075 (V))

Invertible Subaperture Images (MV)
(541 (H) x 434 (V))

Color Raw image
Color Lenslet image

P. David et al. MMSP2017
Plenoptic 1.0 Data Conversion Tool

M44684: New Test Sequences “Toys” and “Trees” Captured by a Light Field Camera @MPEG,Macao
Conversion from Lenslet to Multiview

Conversion from Lenslet to Multiview

Invertible conversion

Non-Invertible conversion

Plenoptic 1.0

Plenoptic 2.0

Lenslet Video

Multiview Video

Introduce artifacts

N18446: Exploration Experiments and Common Test Conditions for Dense Light Fields
Reference Lenslet content Convertor

$K(i, j) = \arg \min E_k(i, j)$

$$E'_k(i, j) = \sum_{s \in S} \sum_{t \in T} \sum_{(u, v) \in b(i, j)} \nabla^2 I_k^{(s, t)}(u, v)$$

$K(i, j)$ is the suitable patch size for the $(I,j)$-th microlens, $b(i,j)$ contains the pixel indices of patch border at the $(i,j)$-th microlens. $(s,t)$ and $(u,v)$ are view and pixel coordinates.
RLC: Plenoptic 2.0 to Multiview

Reference Lenslet content Convertor (RLC)
MPEG (2017 - 2019), IC3D2018
End-to-end system for dense light field

Capture
- Lenslet video
- Conversion from Lenslet to Multiview
- Multiview video

Data Format
- Lenslet video
- Encoder & Decoder
- Multiview video

Data Format
- Lenslet video
- Encoder & Decoder
- Multiview video

Display
- Lenslet video
- Conversion from Lenslet to Multiview
- Multiview video

N18446: Exploration Experiments and Common Test Conditions for Dense Light Fields
EE_MV: Multiview>Compression>Multiview

Dense Multiview Video

Reference images

Evaluation images for average PSNR

Lenslet video → Conversion from Lenslet to Multiview → Multiview video → Encoder & Decoder → Multiview video

<table>
<thead>
<tr>
<th>Plenoptic 1.0</th>
<th>Plenotic 2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>HTM16.2</td>
<td>HM-16.9_SCM_8.0</td>
</tr>
</tbody>
</table>

N18446: Exploration Experiments and Common Test Conditions for Dense Light Fields
EE_LL: Lenslet > Compression > Lenslet

<table>
<thead>
<tr>
<th>Plenoptic 1.0</th>
<th>Plenotic 2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>HTM16.2</td>
<td>HM-16.9_SCM_8.0</td>
</tr>
</tbody>
</table>
Compression of plenoptic 1.0 vs plenoptic 2.0

Plenoptic 1.0

<table>
<thead>
<tr>
<th></th>
<th>PSNR(dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>EE_LL</td>
<td></td>
</tr>
<tr>
<td>EE_MV</td>
<td></td>
</tr>
</tbody>
</table>

Plenoptic 2.0

<table>
<thead>
<tr>
<th></th>
<th>PSNR(dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>EE_LL</td>
<td></td>
</tr>
<tr>
<td>EE_MV</td>
<td></td>
</tr>
</tbody>
</table>

EE_LL > EE_MV

EE_MV > EE_LL
New coding tools for EE_LL

- M44685: Imaging Reshaping (IR)
- M46261: Boundary matching based prediction
Imaging Reshaping (IR)

Imaging Reshaping

Test conditions:
- Reference Software: HM-16.9SCM8.0
- Profile: HEVC Format Range Extension (RExt)
- All Intra Main
- Input Color Format: YUV4:4:4
- QP: 26, 32, 38, 44
- Evaluation: light field performance BD-bitrate

<table>
<thead>
<tr>
<th>Datasets</th>
<th>Image Name</th>
<th>IR vs. HEVC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lytro Illum</td>
<td>Ankylosaurus</td>
<td>-30.0%</td>
</tr>
<tr>
<td></td>
<td>Color_Chart</td>
<td>-15.6%</td>
</tr>
<tr>
<td></td>
<td>House &amp; Lake</td>
<td>-34.0%</td>
</tr>
<tr>
<td></td>
<td>Cards</td>
<td>-8.6%</td>
</tr>
<tr>
<td></td>
<td>Rubik’s Cube</td>
<td>-13.9%</td>
</tr>
<tr>
<td>Lytro 1.0</td>
<td>BSNMom</td>
<td>-23.1%</td>
</tr>
<tr>
<td></td>
<td>Cocktails</td>
<td>-19.8%</td>
</tr>
<tr>
<td></td>
<td>Dessert</td>
<td>-13.6%</td>
</tr>
<tr>
<td></td>
<td>Edelweiss</td>
<td>-4.1%</td>
</tr>
<tr>
<td></td>
<td>Flat_Toes</td>
<td>-20.6%</td>
</tr>
<tr>
<td>Average</td>
<td></td>
<td>-18.3%</td>
</tr>
</tbody>
</table>
Boundary matching based prediction

High correlations among the neighboring macropixels lead to collocated blocks being used to predict the current PU. The PU size can be 32x32 or 64x64.

Weights determination:

\[
D = \left\lfloor \frac{N}{h} \right\rfloor \times h
\]

\[
S = \begin{cases} 
\frac{w}{2}, & \left\lfloor \frac{N}{h} \right\rfloor \in O \\
0, & \left\lfloor \frac{N}{h} \right\rfloor \in E 
\end{cases}
\]

Weighted prediction:

\[
y' = w_0 x_0 + w_1 x_1 + w_2 x_2 + w_3 x_3
\]

PU size is 16x16, 4x4, or 8x8.
Test Conditions
• Test sequences: Teapots, Mini-garden
• Total frames: 60 frames
• Resolution: 8656(H) \times 6074(V)
• Anchor: HTM-HM mode

• Compared methods: IBC, BMP, IBC+BMP
• Configuration: all intra, random access
• QP: 24, 30, 36, 42 and 48
• Evaluation: RD curve and BD-Bitrate
## Compression Performance on Plenoptic 1.0

### BD-Bitrate result for QP 30, 36, 42 and 48 under the all Intra configuration

<table>
<thead>
<tr>
<th></th>
<th>IBC vs. HTM</th>
<th>BMP vs. HTM</th>
<th>IBC+BMP vs. HTM</th>
<th>IBC+BMP vs. IBC</th>
<th>Codec version</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Toys</strong></td>
<td>-16.44%</td>
<td>-21.17%</td>
<td>-23.33%</td>
<td>-8.20%</td>
<td>HM15.0+RExt</td>
</tr>
<tr>
<td><strong>Trees</strong></td>
<td>-3.69%</td>
<td>-4.14%</td>
<td>-5.36%</td>
<td>-1.74%</td>
<td>8.0+SCM-2.0</td>
</tr>
<tr>
<td><strong>Avg.</strong></td>
<td>-10.07%</td>
<td>-12.66%</td>
<td>-14.35%</td>
<td>-4.97%</td>
<td></td>
</tr>
<tr>
<td><strong>Teapots</strong></td>
<td>-22.28%</td>
<td>-24.78%</td>
<td>-32.55%</td>
<td>-13.21%</td>
<td>HM-</td>
</tr>
<tr>
<td><strong>Mini-garden</strong></td>
<td>-30.30%</td>
<td>-31.98%</td>
<td>-38.30%</td>
<td>-11.54%</td>
<td>16.9_SCN_8.0</td>
</tr>
<tr>
<td><strong>Avg.</strong></td>
<td>-26.29%</td>
<td>-28.38%</td>
<td>-35.43%</td>
<td>-12.38%</td>
<td></td>
</tr>
<tr>
<td><strong>Avg. (All)</strong></td>
<td><strong>-18.18%</strong></td>
<td><strong>-20.52%</strong></td>
<td><strong>-24.89%</strong></td>
<td><strong>-8.67%</strong></td>
<td></td>
</tr>
</tbody>
</table>

### BD-Bitrate result for QP 30, 36, 42 and 48 under the random access configuration

<table>
<thead>
<tr>
<th></th>
<th>IBC vs. HTM</th>
<th>BMP vs. HTM</th>
<th>IBC+BMP vs. HTM</th>
<th>IBC+BMP vs. IBC</th>
<th>Codec version</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Toys</strong></td>
<td>-16.62%</td>
<td>-19.25%</td>
<td>-22.36%</td>
<td>-6.84%</td>
<td>HM-15.0+RExt-</td>
</tr>
<tr>
<td><strong>Trees</strong></td>
<td>-35.58%</td>
<td>-1.50%</td>
<td>-36.03%</td>
<td>-0.72%</td>
<td>8.0+SCM-2.0</td>
</tr>
<tr>
<td><strong>Avg.</strong></td>
<td><strong>-26.10%</strong></td>
<td><strong>-10.38%</strong></td>
<td><strong>-29.20%</strong></td>
<td><strong>-3.78%</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Teapots</strong></td>
<td>-21.76%</td>
<td>-11.93%</td>
<td>-27.16%</td>
<td>-6.78%</td>
<td>HM-</td>
</tr>
<tr>
<td><strong>Mini-garden</strong></td>
<td>-26.52%</td>
<td>-18.20%</td>
<td>-32.26%</td>
<td>-7.72%</td>
<td>16.9_SCN_8.0</td>
</tr>
<tr>
<td><strong>Avg.</strong></td>
<td>-24.14%</td>
<td>-15.07%</td>
<td>-29.71%</td>
<td>-7.25%</td>
<td></td>
</tr>
<tr>
<td><strong>Avg. (All)</strong></td>
<td><strong>-25.12%</strong></td>
<td><strong>-12.72%</strong></td>
<td><strong>-29.45%</strong></td>
<td><strong>-5.52%</strong></td>
<td></td>
</tr>
</tbody>
</table>
Compression Performance on Plenoptic 2.0

Test Conditions:

- Test sequences: Tunnel_Train_2, Chess-Pieces, Boxer-IrishMan-Gladiator and ChessPieces-MovingCamera.
- Total frames: 100 frames
- Resolution: R5:2048(H) × 2048(V), R8:3840(H) × 2160(V).

- Anchor: HTM-HM mode
- Compared methods: IBC, BMP, IBC+BMP
- Configuration: all intra, random access
- QP: 28, 33, 38 and 46
- Evaluation: BD-Bitrate and RD curve

![Tunnel_Train_2(Random)](image)

![Boxer-IrishMan-Gladiator(Random)](image)

M47314: Boundary matching based prediction for lenslet video compression® MPEG, Geneva
# Compression Performance on Plenoptic 2.0

## BD-Bitrate result for QP 28,33,38 and 46 under all intra configuration

<table>
<thead>
<tr>
<th>Codec version</th>
<th>Tunnel_Train_2</th>
<th>Chess-Pieces</th>
<th>Boxer-IrishMan-Gladiator</th>
<th>ChessPieces-MovingCamera</th>
<th>Avg. (All)</th>
</tr>
</thead>
<tbody>
<tr>
<td>IBC vs. HTM</td>
<td>-16.15%</td>
<td>-42.99%</td>
<td>-37.84%</td>
<td>-24.81%</td>
<td><strong>-30.45%</strong></td>
</tr>
<tr>
<td>BMP vs. HTM</td>
<td>-3.61%</td>
<td>-20.77%</td>
<td>-15.95%</td>
<td>-22.83%</td>
<td><strong>-15.79%</strong></td>
</tr>
<tr>
<td>IBC+BMP vs. HTM</td>
<td>-16.54%</td>
<td>-43.42%</td>
<td>-38.55%</td>
<td>-29.57%</td>
<td><strong>-32.02%</strong></td>
</tr>
<tr>
<td>IBC+BMP vs. IBC</td>
<td>-0.52%</td>
<td>-0.85%</td>
<td>-1.17%</td>
<td>-6.54%</td>
<td><strong>-2.27%</strong></td>
</tr>
</tbody>
</table>

## BD-Bitrate result for QP 28,33,38 and 46 under random access configuration

<table>
<thead>
<tr>
<th>Codec version</th>
<th>Tunnel_Train_2</th>
<th>Chess-Pieces</th>
<th>Boxer-IrishMan-Gladiator</th>
<th>ChessPieces-MovingCamera</th>
<th>Avg. (All)</th>
</tr>
</thead>
<tbody>
<tr>
<td>IBC vs. HTM</td>
<td>-13.66%</td>
<td>-30.22%</td>
<td>-27.45%</td>
<td>-17.29%</td>
<td><strong>-22.16%</strong></td>
</tr>
<tr>
<td>BMP vs. HTM</td>
<td>-2.01%</td>
<td>-13.89%</td>
<td>-10.02%</td>
<td>-13.60%</td>
<td><strong>-9.88%</strong></td>
</tr>
<tr>
<td>IBC+BMP vs. HTM</td>
<td>-13.63%</td>
<td>-30.67%</td>
<td>-29.24%</td>
<td>-19.22%</td>
<td><strong>-23.19%</strong></td>
</tr>
<tr>
<td>IBC+BMP vs. IBC</td>
<td>0.03%</td>
<td>-0.63%</td>
<td>-2.16%</td>
<td>-2.41%</td>
<td><strong>-1.29%</strong></td>
</tr>
</tbody>
</table>

## Compression Performance on Plenoptic 2.0

<table>
<thead>
<tr>
<th>Codec version</th>
<th>Tunnel_Train_2</th>
<th>Chess-Pieces</th>
<th>Boxer-IrishMan-Gladiator</th>
<th>ChessPieces-MovingCamera</th>
<th>Avg. (All)</th>
</tr>
</thead>
<tbody>
<tr>
<td>IBC vs. HTM</td>
<td>-13.66%</td>
<td>-30.22%</td>
<td>-27.45%</td>
<td>-17.29%</td>
<td><strong>-22.16%</strong></td>
</tr>
<tr>
<td>BMP vs. HTM</td>
<td>-2.01%</td>
<td>-13.89%</td>
<td>-10.02%</td>
<td>-13.60%</td>
<td><strong>-9.88%</strong></td>
</tr>
<tr>
<td>IBC+BMP vs. HTM</td>
<td>-13.63%</td>
<td>-30.67%</td>
<td>-29.24%</td>
<td>-19.22%</td>
<td><strong>-23.19%</strong></td>
</tr>
<tr>
<td>IBC+BMP vs. IBC</td>
<td>0.03%</td>
<td>-0.63%</td>
<td>-2.16%</td>
<td>-2.41%</td>
<td><strong>-1.29%</strong></td>
</tr>
</tbody>
</table>
Light Field Capture to Display System

Coded Aperture Camera

Acquired images

Layered patterns

Layer type 3D display

ViewPLUS ProFUSION 25
Light Field Representation based on Weighted Sum of Binary Patterns

\[ B_n(x - d_1 s, y - d_1 t) \]
\[ B_n(x - d_2 s, y - d_2 t) \]
\[ B_n(x - d_3 s, y - d_3 t) \]

\[ r_n(s, t) \]
\[ d_1 \text{ shift} \]
\[ d_2 \text{ shift} \]
\[ d_3 \text{ shift} \]

\[ L(s, t, x, y) \]
End-to-end system for dense light field

N18446: Exploration Experiments and Common Test Conditions for Dense Light Fields
3D TV Based on Spatial Imaging

Requirements for 3D TV

- No special glasses
- Full parallax
- Natural 3D image

Spatial Imaging

3D Display

Object reconstructed as spatial image

Reconstructed light rays

Observer

3D image based on Integral Photography

Diffuser

10 mm
Distance from lens array to 3D image

40 mm
Integral photography

**Basic principle**

**Recording**
- Object
- Lens array
- Elemental images
- Light rays from object
- Recording media

**Reconstructing**
- Reconstructed light rays
- Elemental images
- Reconstructed 3D image
- Lens array
- Recording media

M. G. Lippmann (1908)
Integral 3D TV

Basic configuration

Capture

Object → Lens array → High-definition camera → Elemental image

Display

High-definition display → Lens array → 3D image → Viewer

Features
Real-time capture and display of moving 3D images
Real objects (not computer graphics) are captured and displayed
Full-parallax images

Problem
Integral 3D system requires huge number of pixels
Problem with Integral 3D system

- Integral 3D system requires huge number of pixels

\[ M = N \times E \]

- \( M \): Total pixels in elemental images
- \( E \): Number of pixels in one elemental image
- \( N \): Number of elemental images (lenses)
Concluding Remark

- Lenslet data
  - integral display, 3D modeling, refocusing, multiview rendering.
  - Multimedia, medical applications

- Compression efficiency of lenslet data
  - Utilize the structure of lenslet data for inter/intra predictions
  - Novel image transform and entropy coding methods
  - Utilize machine learning tools

- New compression method for lenslet
  - leads to improvement over the existing standards.
Acknowledgement