Contributions

- Encoder-decoder network for disparity regression and reflection separation in light fields
- Supervised and unsupervised training
- 3D convolutions to compute features over the vertical and horizontal stack
- Synthetic light field data with ground truth intrinsic components and disparity

Light field structure

Defined on 4D ray space \( R = \Pi \times \Omega \), which parametrizes rays \( r = (x, y, s, t) \) by their intersection coordinates with two planes \( \Pi \) and \( \Omega \) [5]. Intersection with the focal plane \( \Pi \) gives view point coordinates \((s, t)\), while the image plane \( \Pi \) denotes image coordinates \((x, y)\).

Dichromatic model

The input light field \( L \) can be additively separated into its diffuse and specular components \( D \) and \( S \) [6, 10].

Diffuse component and the disparity correspond to the same projections of the same 3D points and share the same pattern.

Specular component follows the specular flow [9], which depends on the local surface geometry and viewpoint change.

Data

- Synthetic data
- Light field benchmark [2]
- Real world data

Generated based on Blender plugin provided by [2]

- 171 scenes
- 321 textures
- 109 environmental maps
- 36 pre-built scenes with objects from Chocofur and The British Museum
- Randomized position, amount of specularity, texture and color

Network architecture

- Paths:
  - One encoding
  - Four decoding
- Six groups of three residual blocks each
- Disparity is a 2D decoder
- Reconstructed light field, diffuse and specular decoders are 3D

- Block types:
  - Depth and resolution the same
  - Reduced resolution, while increased feature depth
- Network operates on the horizontal and vertical EPI stacks in parallel
- The feature output is joined on the bottom layer
- Compression rate is 8.3%

---

This work was supported by the ERC Starting Grant “Light Field Imaging and Analysis” (LIA 336978, FP7-2014).
Presented at the International Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, June 2018.
Network inputs

- crosshair-shaped subset of 17 views
- patch size $9 \times 48 \times 48$
- $\approx 160,000$ patches

Comparisons on the synthetic light fields

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MSE $\times 100$:</td>
<td>5.9</td>
<td>23.1</td>
<td>30.0</td>
<td>35.5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

reflection separation

<table>
<thead>
<tr>
<th>diffuse</th>
<th>specular</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>LMSE $\times 100$:</td>
<td>0.15</td>
</tr>
<tr>
<td>GMSE $\times 100$:</td>
<td>0.12</td>
</tr>
<tr>
<td>SSIM $\times 100$:</td>
<td>0.12</td>
</tr>
</tbody>
</table>

Results for the real-world and synthetic light fields

Visual and quantitative evaluations

<table>
<thead>
<tr>
<th>input disparity diffuse specular</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lytro Illum plenoptic camera</td>
</tr>
<tr>
<td>koala</td>
</tr>
<tr>
<td>flowers, [8]</td>
</tr>
<tr>
<td>buddha</td>
</tr>
<tr>
<td>input disparity diffuse specular</td>
</tr>
<tr>
<td>various sources</td>
</tr>
<tr>
<td>Cotton [2]</td>
</tr>
<tr>
<td>HCI [13]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>input disparity diffuse specular</th>
</tr>
</thead>
<tbody>
<tr>
<td>ground truth Set 1</td>
</tr>
<tr>
<td>ground truth Set 2</td>
</tr>
<tr>
<td>ground truth Set 3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>input disparity diffuse specular</th>
</tr>
</thead>
<tbody>
<tr>
<td>LMSE $\times 100$: 0.11</td>
</tr>
<tr>
<td>LMSE $\times 100$: 0.19</td>
</tr>
<tr>
<td>LMSE $\times 100$: 0.11</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>input disparity diffuse specular</th>
</tr>
</thead>
<tbody>
<tr>
<td>LMSE $\times 100$: 0.14</td>
</tr>
<tr>
<td>LMSE $\times 100$: 0.3</td>
</tr>
<tr>
<td>LMSE $\times 100$: 0.18</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>input disparity diffuse specular</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE $\times 100$: 18.42</td>
</tr>
<tr>
<td>MSE $\times 100$: 5.82</td>
</tr>
<tr>
<td>MSE $\times 100$: 3.54</td>
</tr>
</tbody>
</table>

References