Compressive Light Field Photography using Overcomplete Dictionaries and Optimized Projections

Kshitij Marwah¹  Gordon Wetzstein¹  Yosuke Bando²,¹  Ramesh Raskar¹

¹MIT Media Lab  ²Toshiba Corporation

Presenter: Chinghang Chen, Chenyang Li
How is it done today?

Camera Arrays
e.g., [Wilburn et al. 2002,2005]

Sequential Acquisition
e.g., [Levoy and Hanrahan 1996],
[Liang et al. 2008]
Problem & Assumption

• Capture light field with one single camera by one snapshot without losing spatial resolution
• Natural light fields are sufficiently compressible in some basis or dictionary
Scene from Above

Proposed Technology: Mask-Coded Light Field Projection

Coded Attenuation Mask
Scene from Above

Previous Mask-Coded Light Field Projection

Parallax Barriers
[Ives 1903]

Sum of Sinusoids or MURA
[Veeraraghavan 2007]

- Multiplexing + linear reconstruction
- Low resolution light fields similar to the lenslets design

Compressive Sensing

$$i = \Phi l$$

Sparse Bases: DCT, Wavelets, etc.
Light Field Capture

\[ i(x) = \int l(x, \nu) \, d\nu. \]

\[ i(x) = \int f(x + s(\nu - x)) l(x, \nu) \, d\nu \]

Ray Optics

\[ s = \frac{d_l}{d_a} \]

Light Field
Problem Formulation

\[ i = \Phi l, \quad \Phi = \begin{bmatrix} \Phi_1 & \Phi_2 & \cdots & \Phi_{p^2} \end{bmatrix} \]

\[ i \in \mathbb{R}^m \quad \text{the vectorized sensor image} \]

\[ l \in \mathbb{R}^n \quad \text{light field} \]

\[ \Phi_j \in \mathbb{R}^{m \times m} \]

\[ i = \sum_j \Phi_j l_j \]

\[ i = \Phi l = \Phi D \alpha \quad D \in \mathbb{R}^{n \times d} \quad \alpha \in \mathbb{R}^d \]

\[
\begin{align*}
\text{minimize} & \quad \|\alpha\|_0 \\
\text{subject to} & \quad \|i - \Phi D \alpha\|_2 \leq \epsilon
\end{align*}
\]

\[
\begin{align*}
\text{minimize} & \quad \|i - \Phi D \alpha\|_2 + \lambda \|\alpha\|_0
\end{align*}
\]
Solve for Alpha

- Greedy Methods
  Orthogonal Matching Pursuit (OMP)
- Convex Relaxation Methods
  Basis Pursuit Denoise (BPDN)
A sparse set of 4D light field atoms, each of size $p_x \times p_x \times p_\nu \times p_\nu$, $p_x^2 p_\nu^2 = n$, is then reconstructed for each sensor pixel.
Compressive Light Field Reconstruction

Captured 2D Image

4D Reconstruction

Mask-Coded Projection

\[ i = \Phi l \]

Sparse Coefficients!

Basis Pursuit Denoise:

\[
\begin{align*}
\text{minimize} & & \|\alpha\|_1 \\
\text{subject to} & & \|i - \Phi D\alpha\|_2 \leq \epsilon
\end{align*}
\]
Proposed Technology: Optimized Mask w.r.t. the Dictionary

\[ i = \Phi \]

- \( i \) = Image
- \( \Phi \) = Coded projection
- \( D \) = Dictionary
- \( \alpha \) = Coefficients

- Multiplexing + nonlinear reconstruction
- Higher spatial resolution
Compressive Light Field Representation

\[
\mathbf{l} = \mathbf{D} \mathbf{\alpha} \quad \text{such that } \mathbf{\alpha} \text{ is sparse}
\]

- \( \mathbf{l} \): Light field vector
- \( \mathbf{D} \): Basis matrix
- \( \mathbf{\alpha} \): Coefficient vector

Only a few non-zero coefficients

Similar to DCT in JPEG
But DCT is not enough for light field reconstruction
Compressive Light Field Representation

\[ \mathbf{1} = \mathbf{D} \alpha \quad \text{s.t.} \quad \alpha \text{ is sparse} \]

Light field vector

Dictionary

Coefficient vector

Can lead to fewer non-zero coefficients

Light Field Atoms
Compressibility

4D Light Field Patch

4D DCT

4D Light Field Atoms
Compressibility Evaluation

Light field atoms have better compression performance than other standard bases.
Sample 1,800,000 random 4D patches from training light fields, use coreset of 50000 patches.

\[ l_i = D \alpha_i \quad \text{s.t.} \quad \alpha_i \text{ is sparse} \]
minimize $\{\mathcal{D,A}\}$ subject to $\forall i, \|A_i\|_0 \leq k$
Light Field “Atoms” in Dictionary

Light fields can be represented by only a few of these atoms

5,000 atoms, each 9x9 pixels and 5x5 views
Optical Preservation of Light Field Info

\[ i = \Phi l D \alpha \]

Image \( \Phi \) Coded projection \( l \) Light field \( D \) Dictionary \( \alpha \) Coefficient vector

We need to be able to distinguish atoms from their projections
Scene from Above

Proposed Technology: Mask-Coded Light Field Projection

- random and optimized optical codes
- multiplexing & nonlinear reconstruction
Mask Pattern Optimization

\[
\begin{align*}
\text{minimize} \quad & \| I - G^T G \|_F \\
\text{subject to} \quad & 0 \leq f_i \leq 1, \quad \forall i \\
& \sum_i f_i / m \geq \tau
\end{align*}
\]

\[G = \Phi D\]

\[i = \Phi D \alpha\]

Image \quad Coded projection \quad Dictionary \quad Coefficient vector
Prototype Setup with a Variable Mask

- Imaging Lens
- Polarizing Beamsplitter
- LCoS
- Virtual Sensor
- SLR Camera
- Image sensor
- Camera
Diffuse Scene

Coded 2D Projection

Reconstructed 4D Light Field
Diffuse Scene

Coded 2D Projection → Reconstructed 4D Light Field
Refocus
Glossy Scene with Refraction

Coded 2D Projection

Reconstructed Light Field
5x5 viewpoints
Light field represented by 5 most significant coefficients only

4D DCT

4D Light Field Atoms

PSNR 24.5 dB

PSNR 27.2 dB
Additional Applications – Denoising

Noisy 4D Light Field

Denoised 4D Light Field
Pros:
No spatial resolution loss, and one snapshot will do.
The dictionary is able to recover occluded scene, sharp edge, or complex lighting condition such as refraction.

Cons:
Dictionary is expensive to train, and the atoms are adapted to training data. (depth range, aperture diameter, scene structures)
The reconstruction complexity.
Light transmission loss.
Solution to important issues
Should talk more on the limitation, depth of field, or angular resolution
The hardware implementation in this paper did not address artifacts such as angle-dependent color and intensity nonlinearities.
1.5
Thank you

- Q&A