Compressive Light Field Photography

using Overcomplete Dictionaries and Optimized Projections

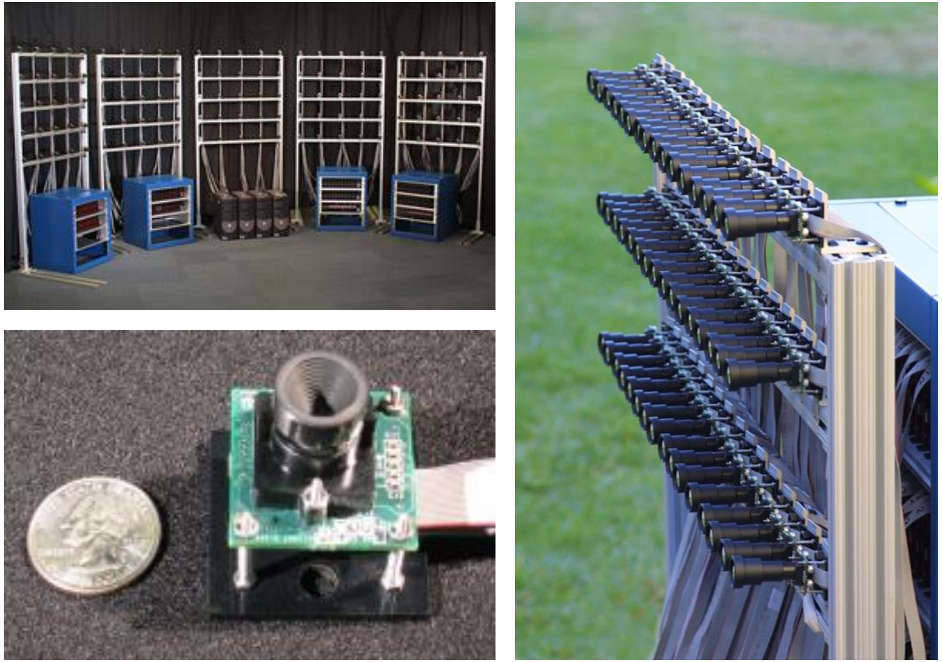
Kshitij Marwah[1] Gordon Wetzstein[1] Yosuke Bando[2,1] Ramesh Raskar[1]

[1]MIT Media Lab [2]Toshiba Corporation

Presenter: Chinghang Chen, Chenyang Li

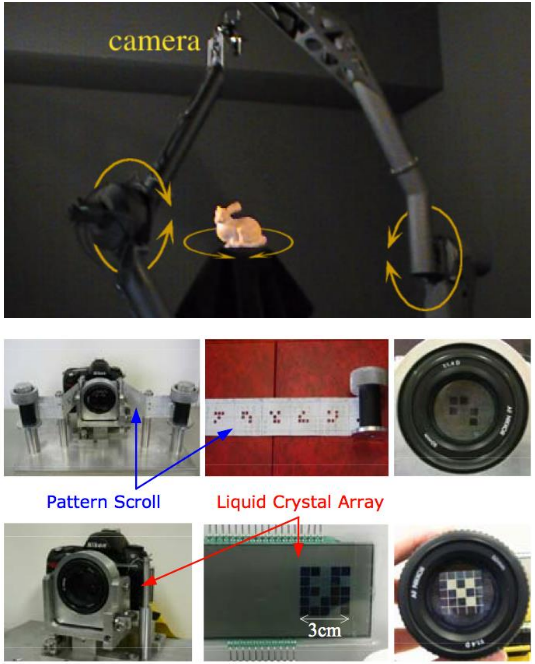
SIGGRAPH2013

How is it done today?



Camera Arrays

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



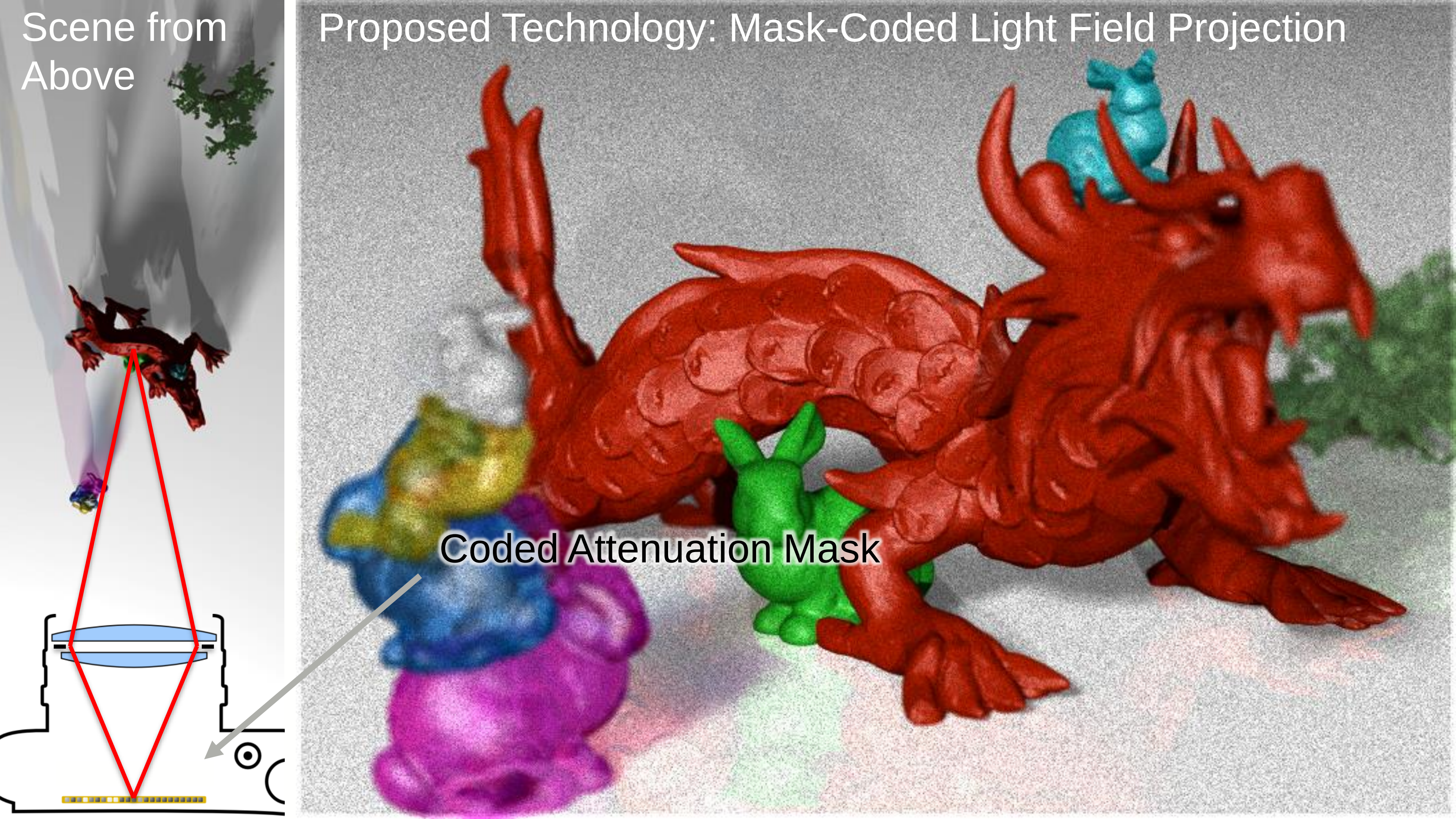
Sequential Acquisition

e.g., [Levoy and Hanrahan 1996],

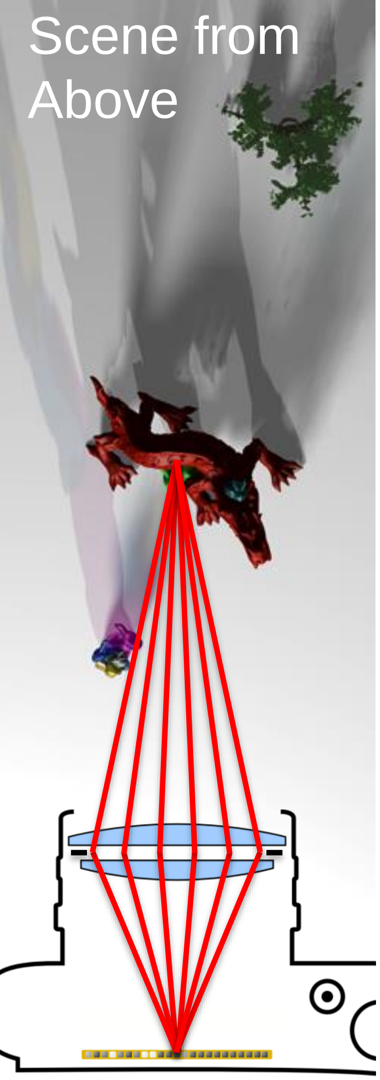
[Liang et al. 2008]

Problem & Assumption

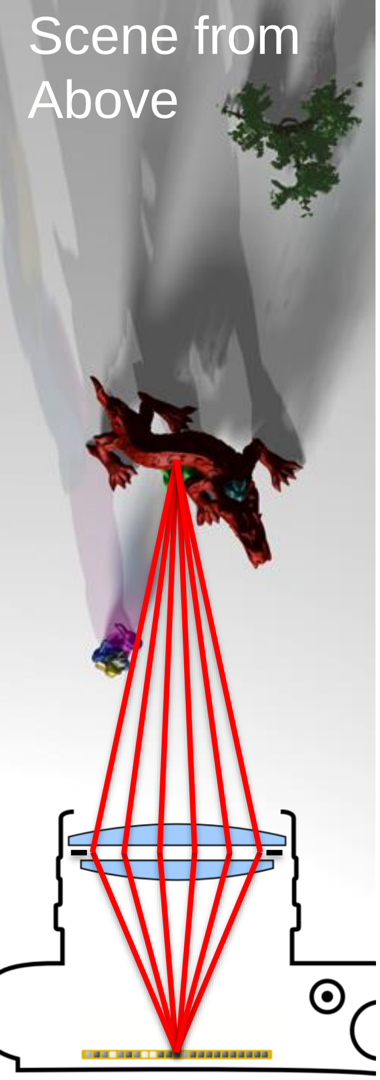
- Capture light field with one single camera by one snapshot without losing spatial resolution
- Natural light fields are sufficiently compressible in some basis or dictionary



Scene from Above
Proposed Technology: Mask-Coded Light Field Projection
Coded Attenuation Mask


Scene from
Above

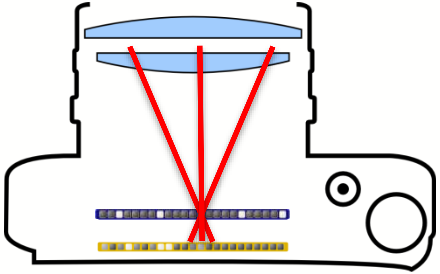

Light Field

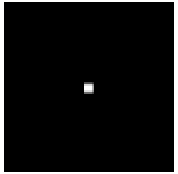

Scene from
Above


Light Field

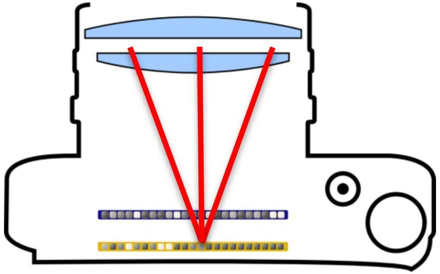
Previous Mask-Coded Light Field Projection

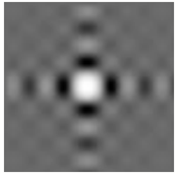
Parallax Barriers
[Ives 1903]

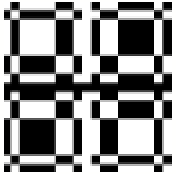




Sum of Sinusoids or MURA
[Veeraraghavan 2007]





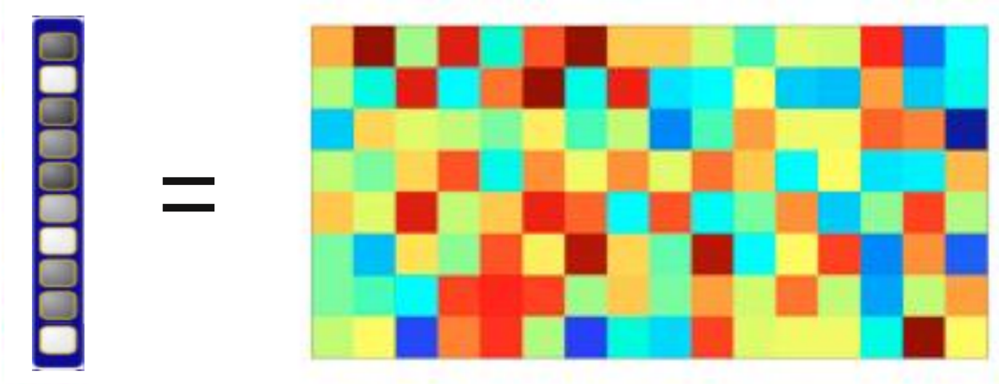


- Multiplexing + linear reconstruction
- Low resolution light fields similar to the lenslets design

“On Plenoptic Multiplexing and Reconstruction”, IJCV, Wetzstein et al. 2013

Compressive Sensing

$$\mathbf{i} = \Phi \mathbf{l}$$



Sparse Bases: DCT, Wavelets, etc.

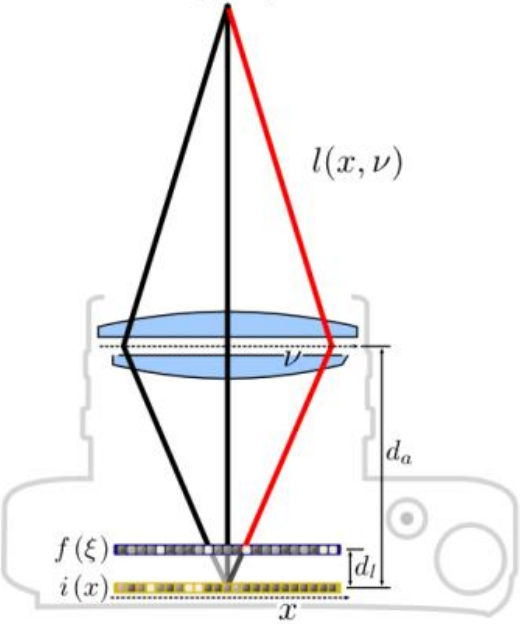
Light Field Capture

$$i\left(x\right) = \int_{\mathcal{V}} l\left(x, \nu\right) d\nu.$$

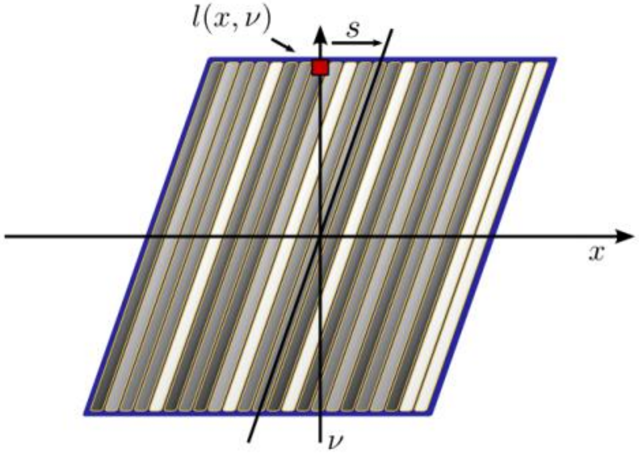
$$i(x) = \int_{\mathcal{V}} f\left(x + s(\nu - x)\right) l(x, \nu)\, d\nu$$

$$s = d_l / d_a$$

Ray Optics



Light Field



Problem Formulation

$$\mathbf{i} = \mathbf{\Phi l}, \quad \mathbf{\Phi} = \left[\mathbf{\Phi}_1 \ \mathbf{\Phi}_2 \ \cdots \ \mathbf{\Phi}_{p_\nu^2}\right]$$

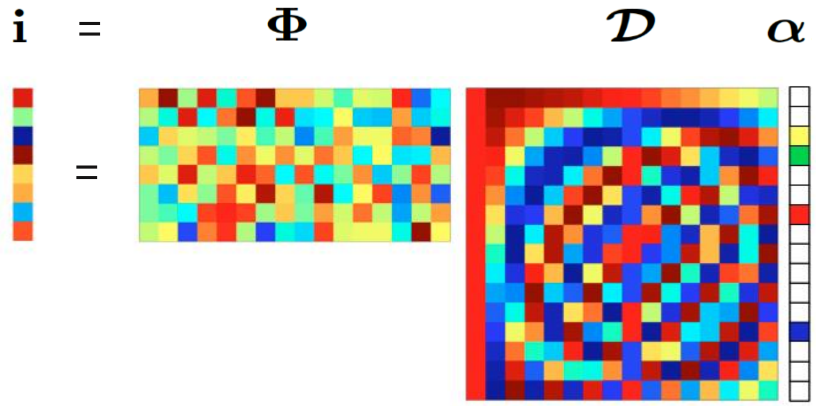
$\mathbf{i} \in \mathbb{R}^m$ the vectorized sensor image

$\mathbf{l} \in \mathbb{R}^n$ light field

$\mathbf{\Phi}_j \in \mathbb{R}^{m \times m}$

$$\mathbf{i} = \sum_j \mathbf{\Phi}_j \mathbf{l}_j$$

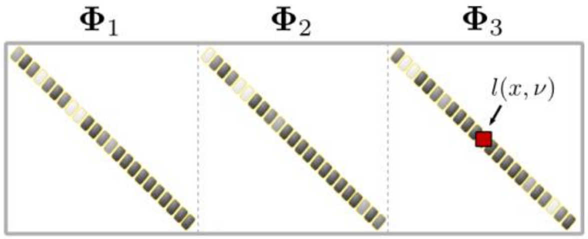
$$\mathbf{i} = \mathbf{\Phi l} = \mathbf{\Phi} \boldsymbol{\mathcal{D}} \boldsymbol{\alpha} \qquad \boldsymbol{\mathcal{D}} \in \mathbb{R}^{n \times d} \qquad \boldsymbol{\alpha} \in \mathbb{R}^d$$



Measurement Matrix

$$\underset{\{\alpha\}}{\text{minimize}} \quad \|\boldsymbol{\alpha}\|_0$$
$$\text{subject to} \quad \|\mathbf{i} - \mathbf{\Phi}\boldsymbol{\mathcal{D}}\boldsymbol{\alpha}\|_2 \leq \epsilon$$

$$\underset{\{\alpha\}}{\text{minimize}} \ \|\mathbf{i} - \mathbf{\Phi}\boldsymbol{\mathcal{D}}\boldsymbol{\alpha}\|_2 + \lambda \|\boldsymbol{\alpha}\|_0$$



Φ

Solve for Alpha

- Greedy Methods

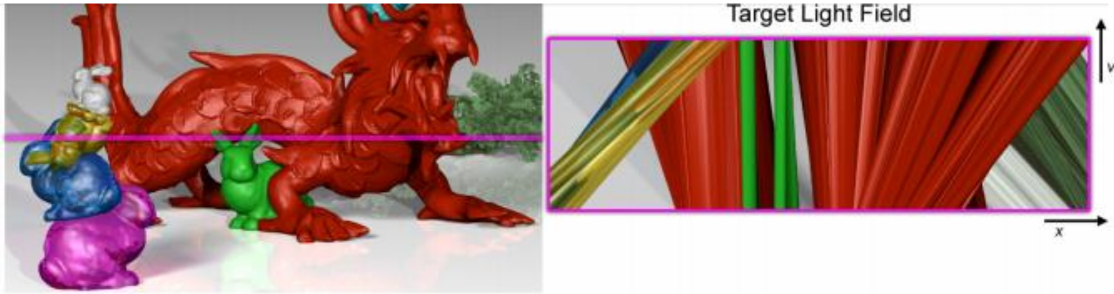
Orthogonal Matching Pursuit (OMP)

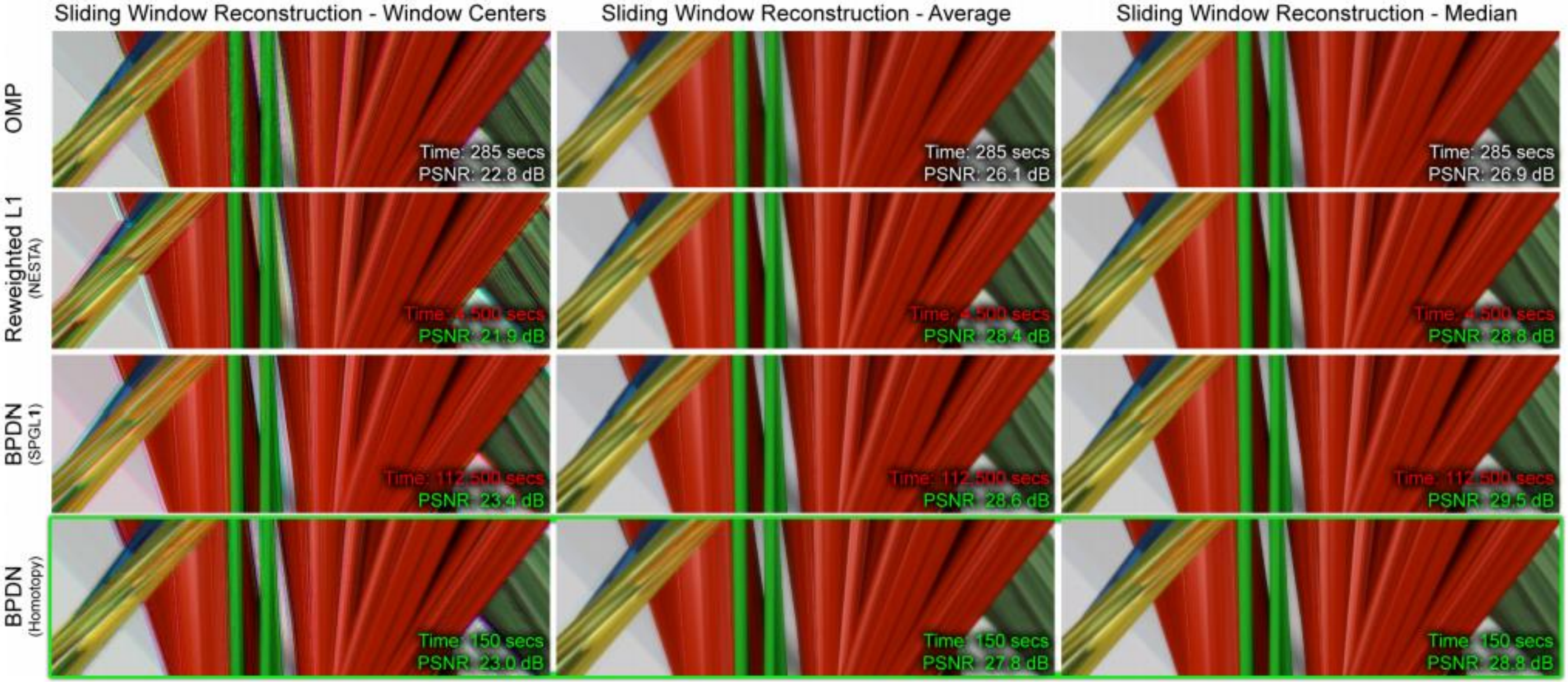
- Convex Relaxation Methods

Basis Pursuit Denoise (BPDN)

window of patch size $p_x \times p_x$, $p_x^2 = m$

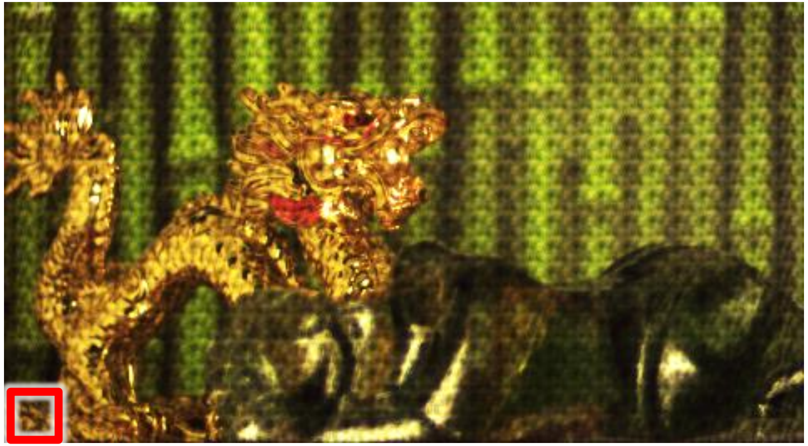
A sparse set of 4D light field atoms, each of size $p_x \times p_x \times p_\nu \times p_\nu$, $p_x^2 p_\nu^2 = n$, is then reconstructed for each sensor pixel.



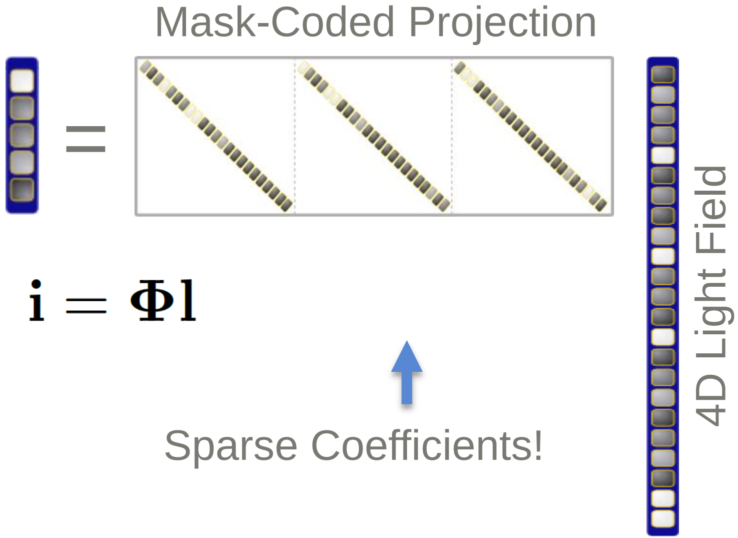


Compressive Light Field Reconstruction

Captured 2D Image



4D Reconstruction

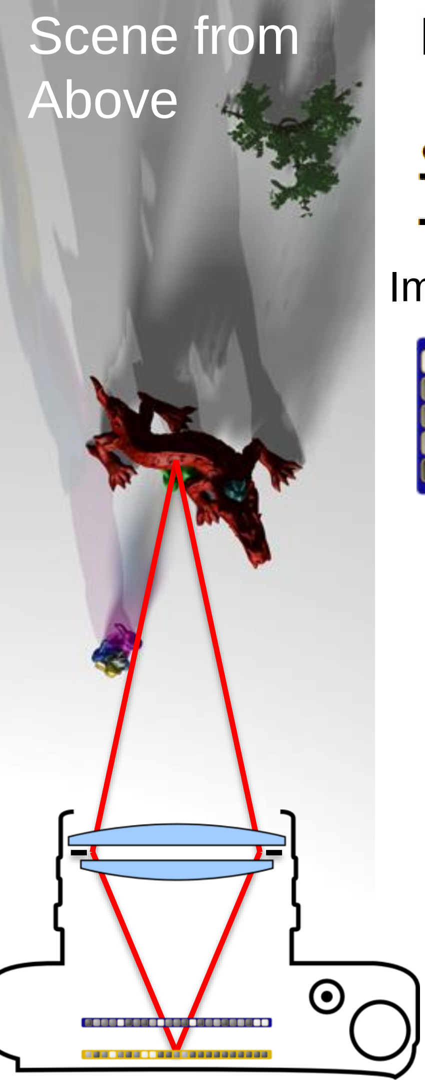


$$\mathbf{i} = \mathbf{\Phi l}$$

Sparse Coefficients!

Basis Pursuit Denoise:

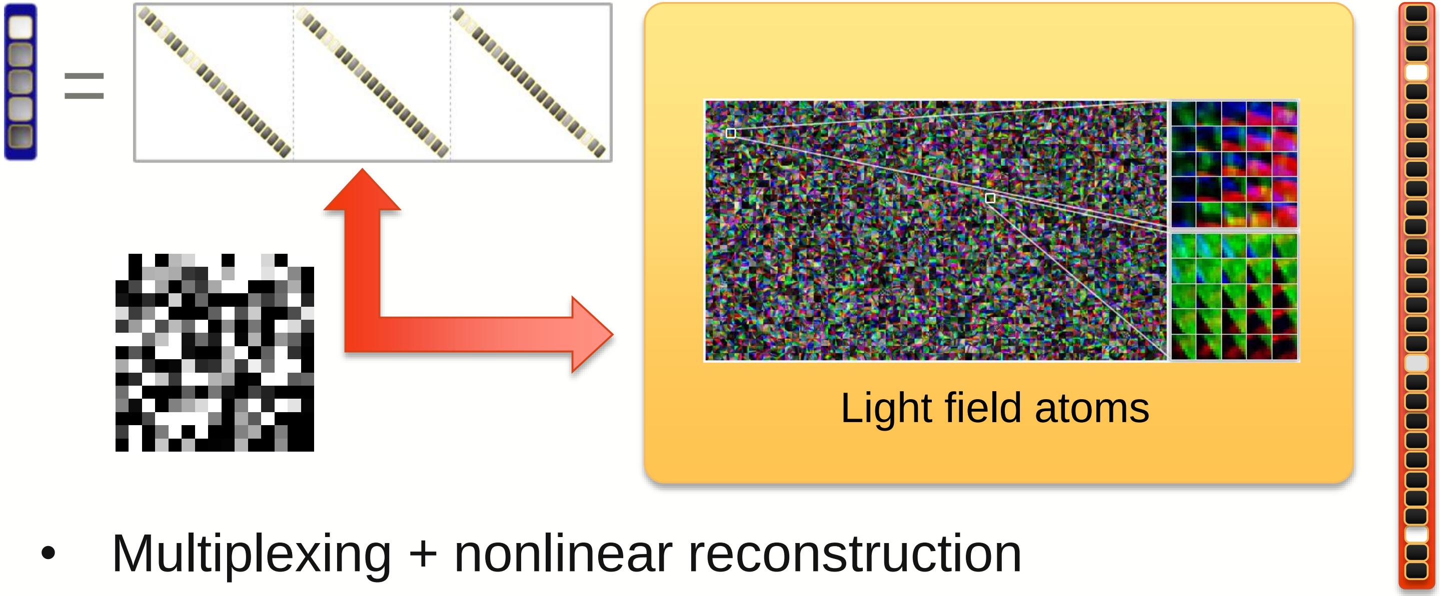
$$\underset{\{\boldsymbol{\alpha}\}}{\text{minimize}} \quad \|\boldsymbol{\alpha}\|_1$$
$$\text{subject to} \quad \|\mathbf{i} - \mathbf{\Phi}\boldsymbol{\mathcal{D}}\boldsymbol{\alpha}\|_2 \leq \epsilon$$



Proposed Technology: Optimized Mask w.r.t. the Dictionary

$$\mathbf{i} = \Phi \mathcal{D} \alpha$$

Image | Coded projection | Dictionary | Coefficients



- Multiplexing + nonlinear reconstruction
- Higher spatial resolution

Compressive Light Field Representation

$$\mathbf{l} = \mathcal{D}\,\alpha \quad \text{such that } \alpha \text{ is } \textit{sparse}$$

Light field vector — Basis matrix — Coefficient vector

Only a few non-zero coefficients

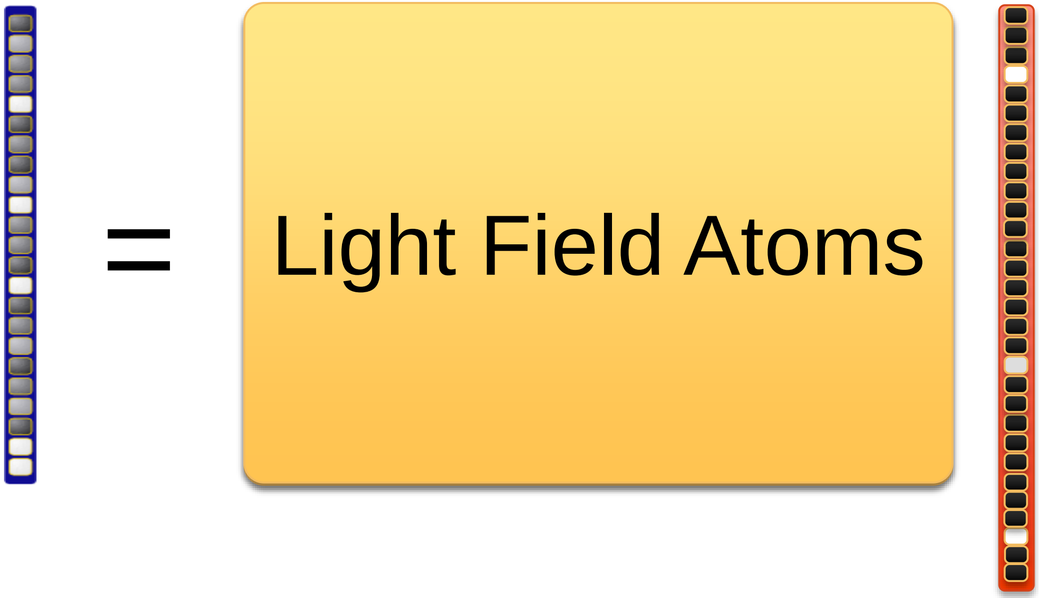
Similar to DCT in JPEG

But DCT is not enough
for light field reconstruction

Compressive Light Field Representation

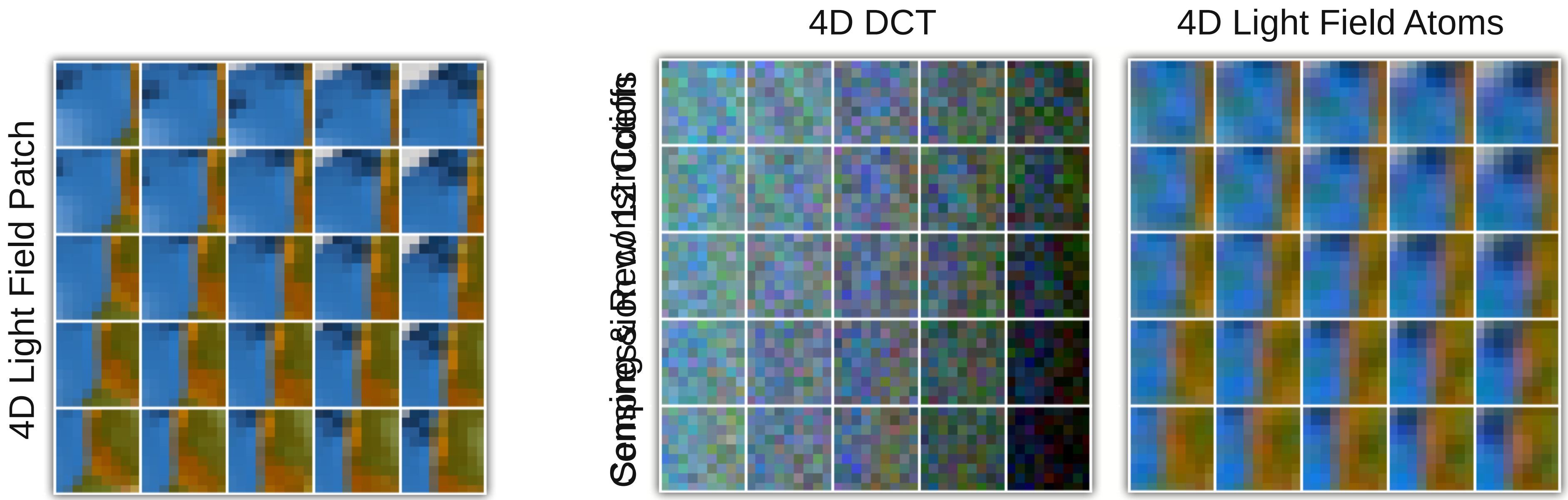
$\mathbf{l} = \mathcal{D} \; \alpha$ s.t. α is *sparse*

Light field vector — Dictionary — Coefficient vector

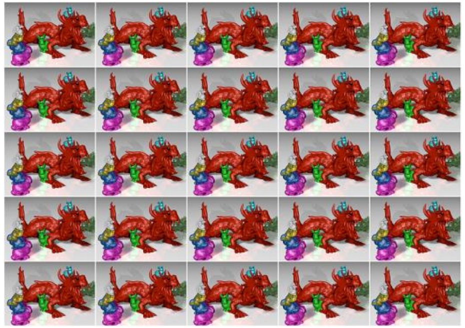


Can lead to fewer non-zero coefficients

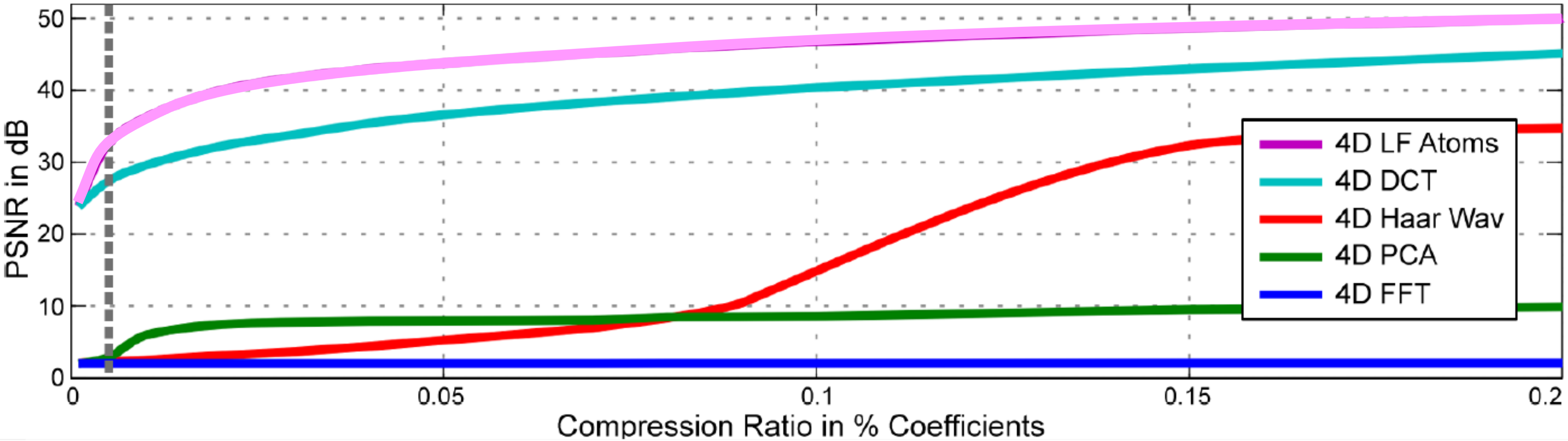
Compressibility



Compressibility Evaluation



Light field atoms have better compression performance than other standard bases



Dictionary Learning

$$\mathbf{l}_i = \mathcal{D} \, \alpha_i \quad \text{s.t.} \quad \alpha_i \text{ is } \textit{sparse}$$

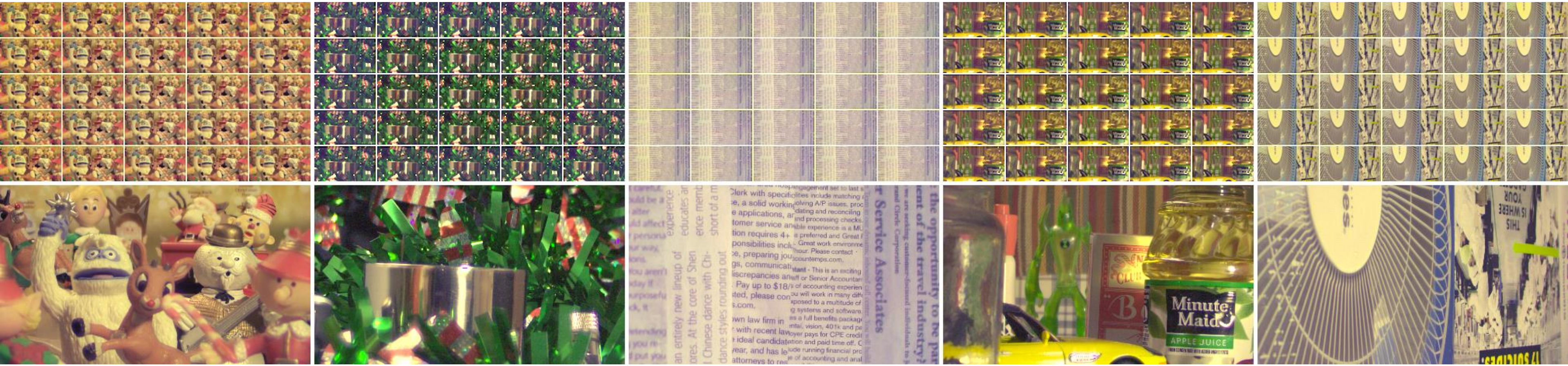
Training light field — Dictionary — Coefficient vector — for all i

Sample 1,800,000 random 4D patches from training light fields, use coreset of 50000 patches

Dictionary Learning

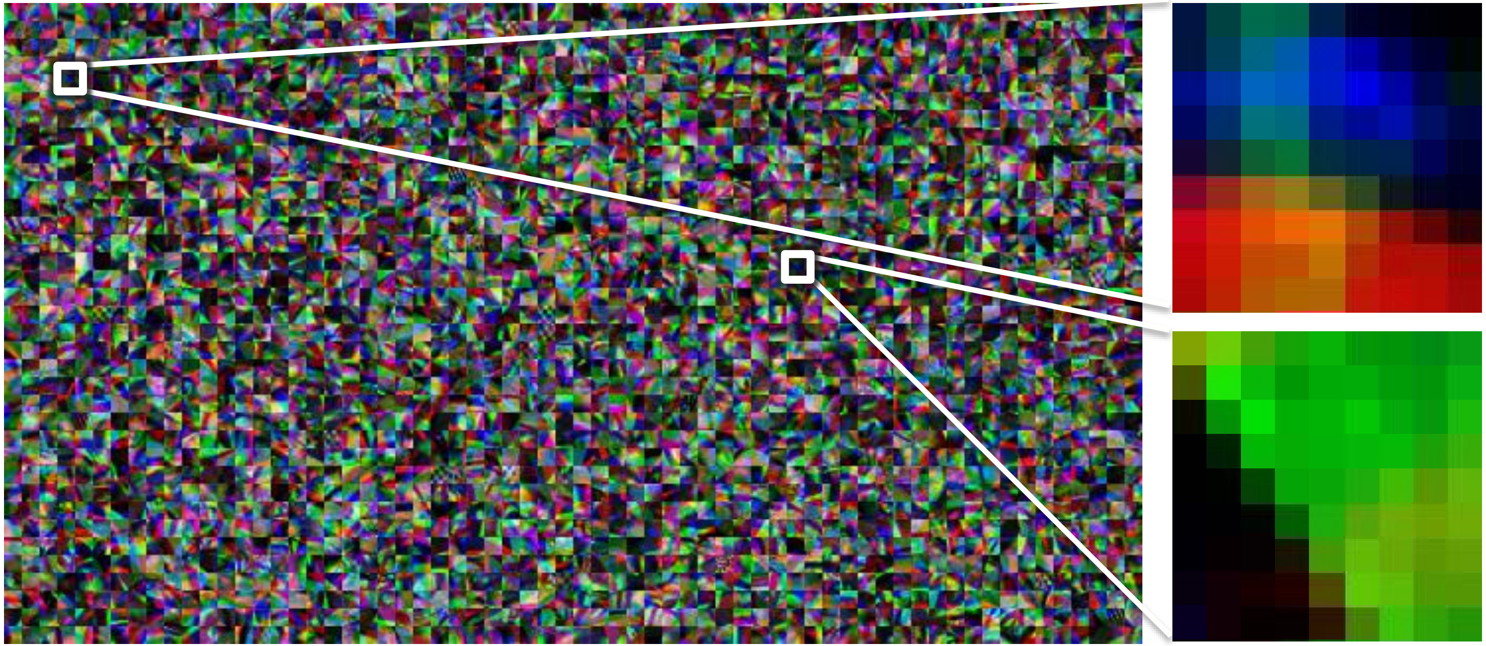
$$\underset{\{\mathcal{D},\mathcal{A}\}}{\text{minimize}} \quad \|\mathbf{L} - \mathcal{D}\mathcal{A}\|_F$$

$$\text{subject to} \quad \forall i, \|\mathcal{A}_i\|_0 \leq k$$



Light Field "Atoms" in Dictionary

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

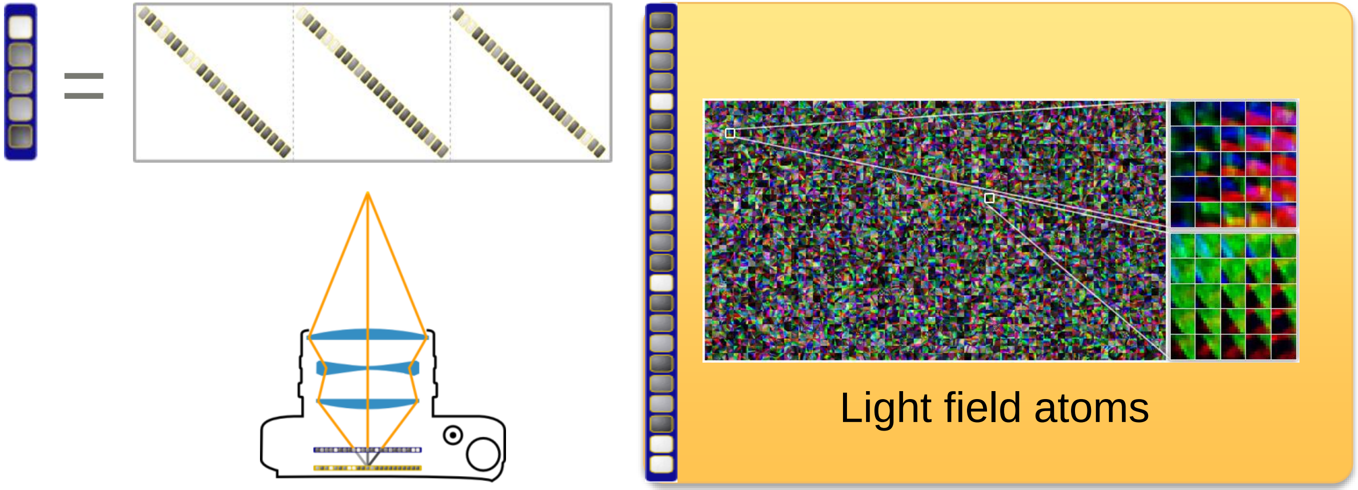


5,000 atoms, each 9x9 pixels and 5x5 views

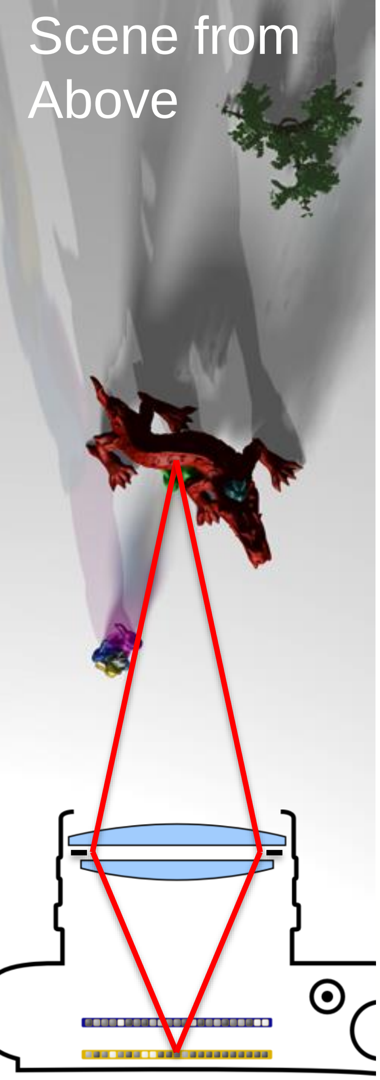
Optical Preservation of Light Field Info

$$\mathbf{i} = \boldsymbol{\Phi} \; \mathbf{l} \; \mathcal{D} \; \boldsymbol{\alpha}$$

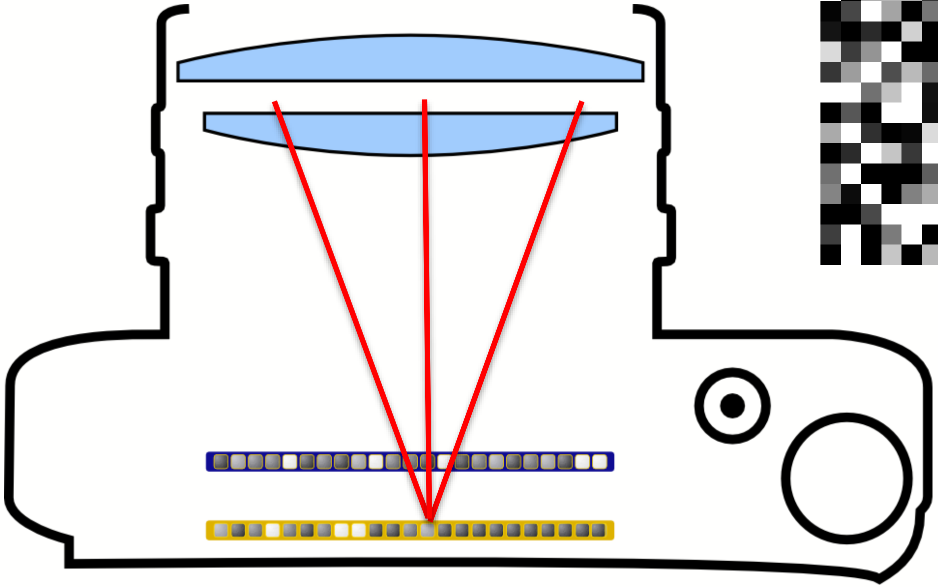
Image | Coded projection | Light field | Dictionary | Coefficient vector

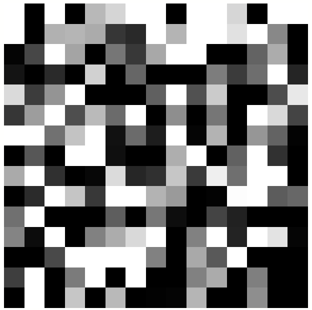


We need to be able to distinguish atoms from their projections



Proposed Technology: Mask-Coded Light Field Projection





- random and optimized optical codes
- multiplexing & nonlinear reconstruction

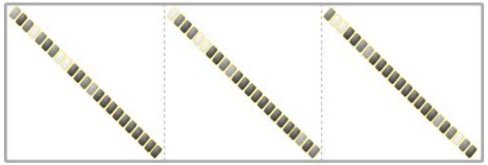
Mask Pattern Optimization

$$\begin{aligned} \underset{\{\mathbf{f}\}}{\text{minimize}} \quad & \left\| \mathbf{I} - \mathbf{G}^T \mathbf{G} \right\|_F \\ \text{subject to} \quad & 0 \leq f_i \leq 1, \ \forall i \\ & \textstyle\sum_i f_i / m \geq \tau \end{aligned}$$

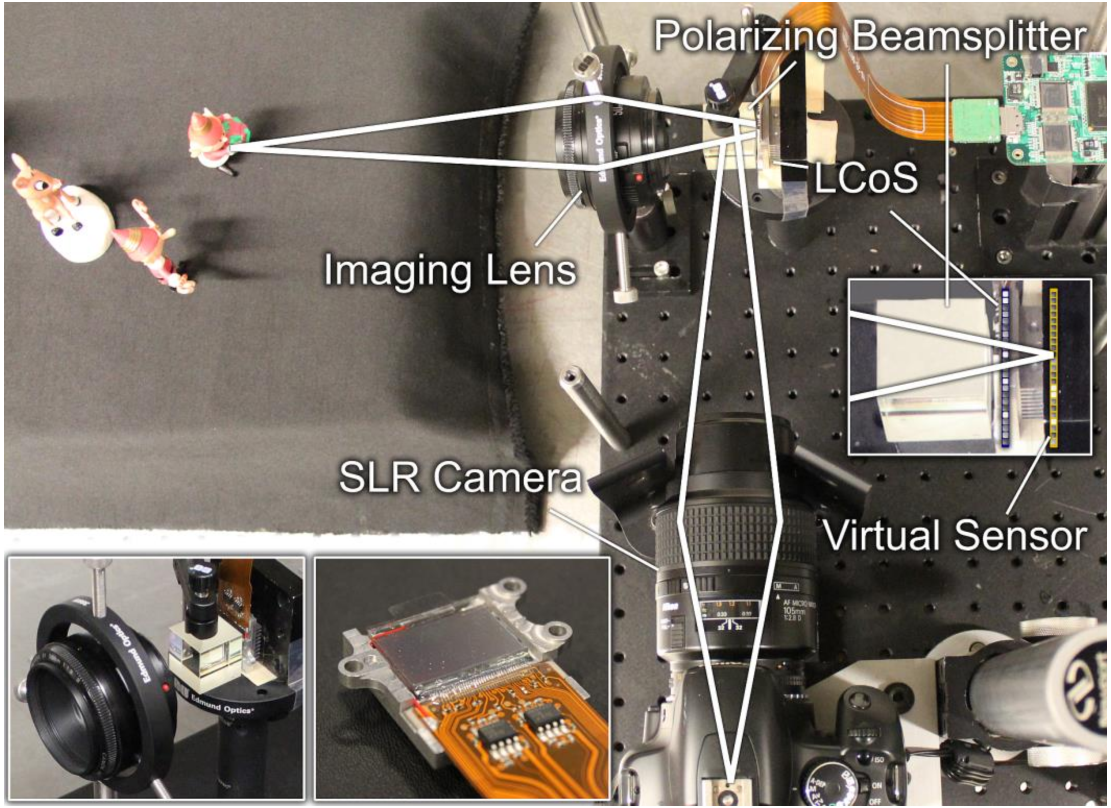
$$\mathrm{G} = \Phi \mathcal{D}$$

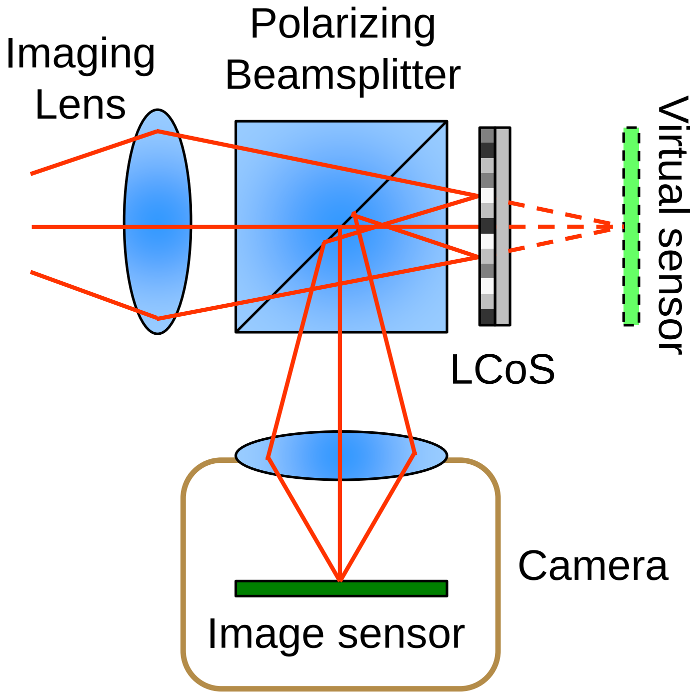
$$\mathbf{i} = \quad \mathbf{\Phi} \quad \mathcal{D} \quad \alpha$$

Image | Coded projection | Dictionary | Coefficient vector



Prototype Setup with a Variable Mask





Diffuse Scene

Coded 2D Projection

Reconstructed 4D Light Field

Diffuse Scene

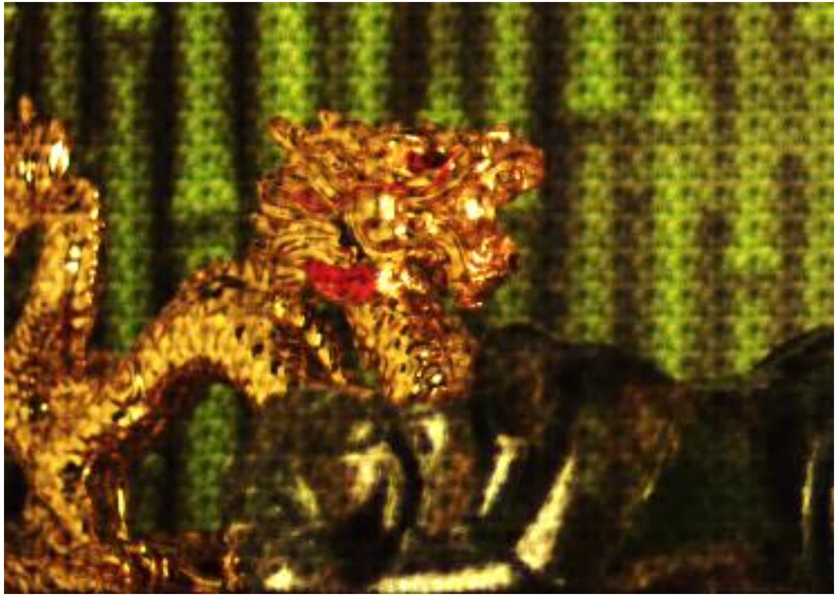
Coded 2D Projection

Reconstructed 4D Light Field

Refocus

Real Focus

Glossy Scene with Refraction



Coded 2D Projection

Reconstructed Light Field
5x5 viewpoints

Animated Scenes

Additional Applications – Compression

Light field represented by **5** most significant coefficients only

4D DCT | 4D Light Field Atoms | 4D DCT | 4D Atoms

Additional Applications – Denoising

Approach Summary

- 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.

Paper Summary

- 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