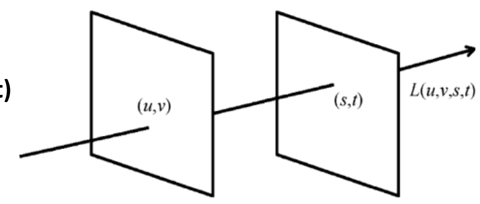
Graph-based Transforms for Predictive Light Field Compression based on Super-Pixels

Mira Rizkallah (1), Xin Su (2) Thomas Maugey (2) and Christine Guillemot (2)
(1) IRISA University of Rennes 1, Rennes, France
(2) INRIA Rennes Bretagne-Atlantique, Rennes, France

LIGHT FIELDS: APPLICATIONS AND CHALLENGES

◊ ***"Light Field"*** :

4D: Intersection with 2 planes **LF(u, v, s, t)**



Capture ONE photograph, Render After!

◊ **Functionalities:**

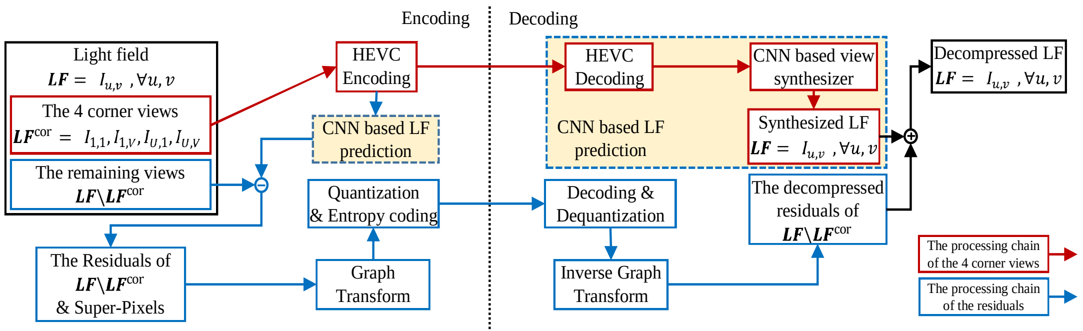
- **Refocusing :** Focusing at different regions of the scene.
- **Depth Estimation**: Estimating the depth of objects in the scene.
- **Extended Focus**: Simulating photographs with extended depth of field.

◊ **Challenge**:

- **Dense volumes of Data**
  Necessity of **Compression** for storage and transmission

Light Field Sub-Aperture images

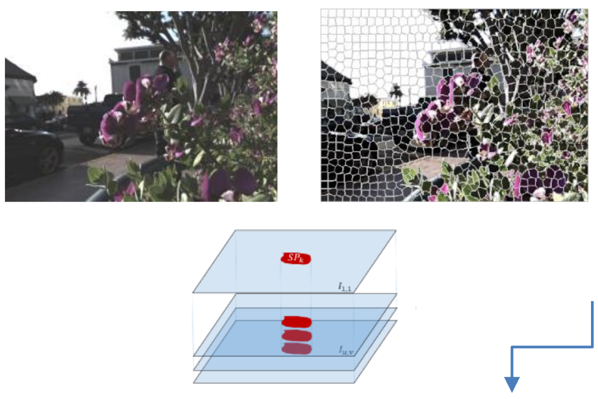
LIGHT FIELD PREDICTIVE CODING SCHEME



Prediction

Code 4 *corner views* using HEVC-Inter and use them to synthesize the whole light field using two convolutional neural networks (CNN). [1]

1. One CNN trained to model the disparity in the given light field
2. Another CNN to estimate the color of the synthesized views.



Super-pixel segmentation

- Segmentation of the *central view* using SLIC [2]
- Propagation to other views without changing the position and size of the segmentation masks

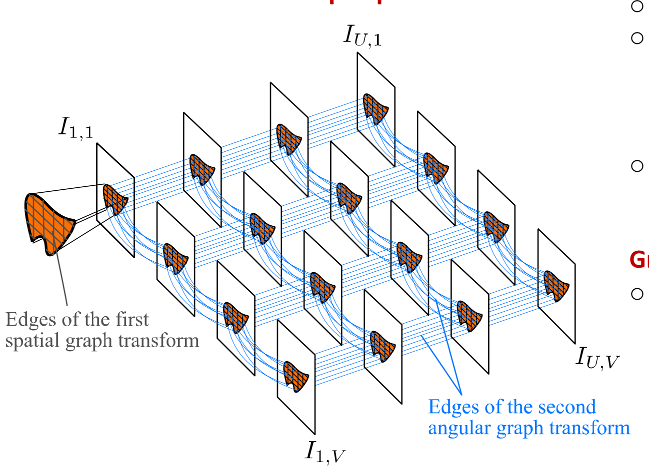
Residues coding

Purpose of the study

- Having signals to code(residues) and signal supports (Super-pixels) , Construct Local separable Graphs and use Graph Transforms to capture the correlations in both spatial and angular dimensions.

GRAPH DEFINITION

For each super-pixel



The Graph

- The vertices are pixels in all the views.
- Edges connect neighboring pixels inside the super-pixel in a view and corresponding pixels across neighboring views
- The residues obtained after prediction are the signal residing on the vertices of the graph.

Graph transform

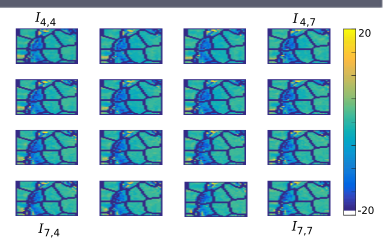
- To explore the correlations and compact the energy of the residual signal, we first **perform local super-pixel based spatial GT** followed by **local angular GT**.

GRAPH BASED TRANSFORM AND CODING

Spatial graph transform (1st)

Given the residues luminance values in one view v of the light field and a segmentation map M, the k^{th} superpixel can be represented by a signal $f_k^v \in \mathbb{R}^{N_k}$

We construct a 4-nearest-neighbor graph to capture correlations between the signal values.



To define the transform, we compute the Laplacian matrix and its eigenvectors:

$$\mathbf{L} = \mathbf{D} - \mathbf{A} \qquad \mathbf{L} = \mathbf{U}^{\mathbf{T}}\mathbf{\Lambda}\mathbf{U}$$

$$\boxed{\hat{f}_k^v = \mathbf{U} f_k^v}$$

Angular graph transform (2nd)

For a specific band number l and superpixel k, the band signal is defined as

$b_k^l = \{\hat{f}_k^v(l), \quad v = 1 : N_v\}$

Notations:

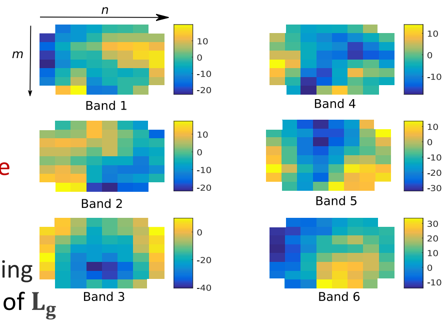
$\mathbf{A}$: Adjacency matrix ($\mathbf{A}_{ij} = 1$ only if there is an edge connecting pixels i and j)

$\mathbf{D}$: Degree matrix (diagonal matrix with $d_{ii} = \sum_{j \in V} A(i,j)$)

$\mathbf{U}$: Set of eigenvectors in a matrix form

uGT $\mathbf{L_v} = \mathbf{D_v} - \mathbf{A_v} \qquad \mathbf{L_v} = \mathbf{U_v^T}\mathbf{\Gamma}\mathbf{U_v}$

$$\boxed{\hat{b}_k^l = \mathbf{U_v} b_k^l}$$

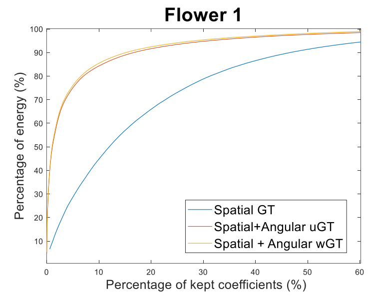


wGT : To explore the various correlation patterns in the different frequency bands, divide them into 64 groups. For each group g, learn a Laplacian matrix $\mathbf{L_g}$ [3] using observations of all superpixels.The band signals belonging to this group are thus projected onto the eigenvectors of $\mathbf{L_g}$

Transform coefficients coding : Simple quantization and entropy coding

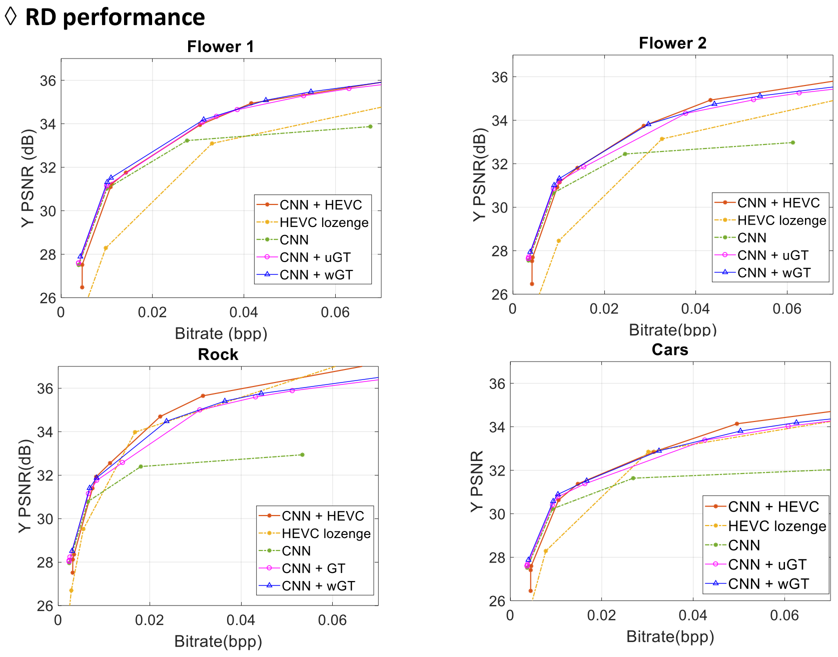
EXPERIMENTS AND ANALYSIS

◊ **Energy compaction**



- **Higher Energy Compaction observed with the angular transform** compared with only applying the spatial transform, with a slight improvement for the wGT.
- **Utility of exploring interview correlations** between residues in different views and **adapting graph weights.**

◊ **RD performance**



	CNN+uGT *vs* CNN	CNN+uGT *vs* HEVC lozenge	CNN+uGT *vs* CNN+HEVC	CNN+wGT *vs* CNN+uGT
Car	0.6	0.9	0.3	0.1
Flower 1	0.3	1.7	0.2	0.1
Flower 2	0.4	1.6	0.3	0.2
Rock	-0.1	0.7	-0.1	0.3

Table 1: Bjontegaard comparison (ΔPSNR(dB)) at low bitrate (< 0.04 bpp)

Take-home Messages

- Graph transforms are suitable tools for exploiting spatial and angular correlations in light field data. With a simple transform coding scheme, we can attain the performance of complex HEVC-based coding.
- Future work is dedicated for dealing with disocclusions and building more consistent super-rays across the views, to take better advantage of the graph transform .

[1] N. K. Kalantari, T.-C. Wang, and R. Ramamoorthi. Learning-based view synthesis for light field cameras. *ACM Transactions on Graphics (Proceedings of SIGGRAPH Asia 2016)*, 35(6), 2016.

[2] R. Achanta, A. Shaji, Kevin K. Smith, A. Lucchi, P. Fua, and S. Susstrunk. SLIC Superpixels Compared to State-of-the-Art Superpixel Methods. *IEEE Trans. Pattern Anal. Mach. Intell.*

[3] H. E. Egilmez, E. Pavez, and A. Ortega. Graph learning from data under laplacian and structural constraints. *IEEE Journal of Selected Topics in Signal Processing*, 11(6):825–841, Sept 2017.

ERC advanced grant CLIM