Camera-agnostic Format and Processing for Light-field Data

Mitra Damghanian, Paul Kerbiriou, Valter Drazic, Didier Doyen, Laurent Blondé

Technicolor Research & Innovation, Rennes, France

technicolor

Why?

Landscape:

- LF is an enabler for the next generation of 3D/AR/VR experiences
- AR/VR devices are working on incorporating LF capabilities
- All current and future LF-based applications need content
- LF capture systems are diverse

Bottleneck

Lack of unified representation, storage and processing formats for LF

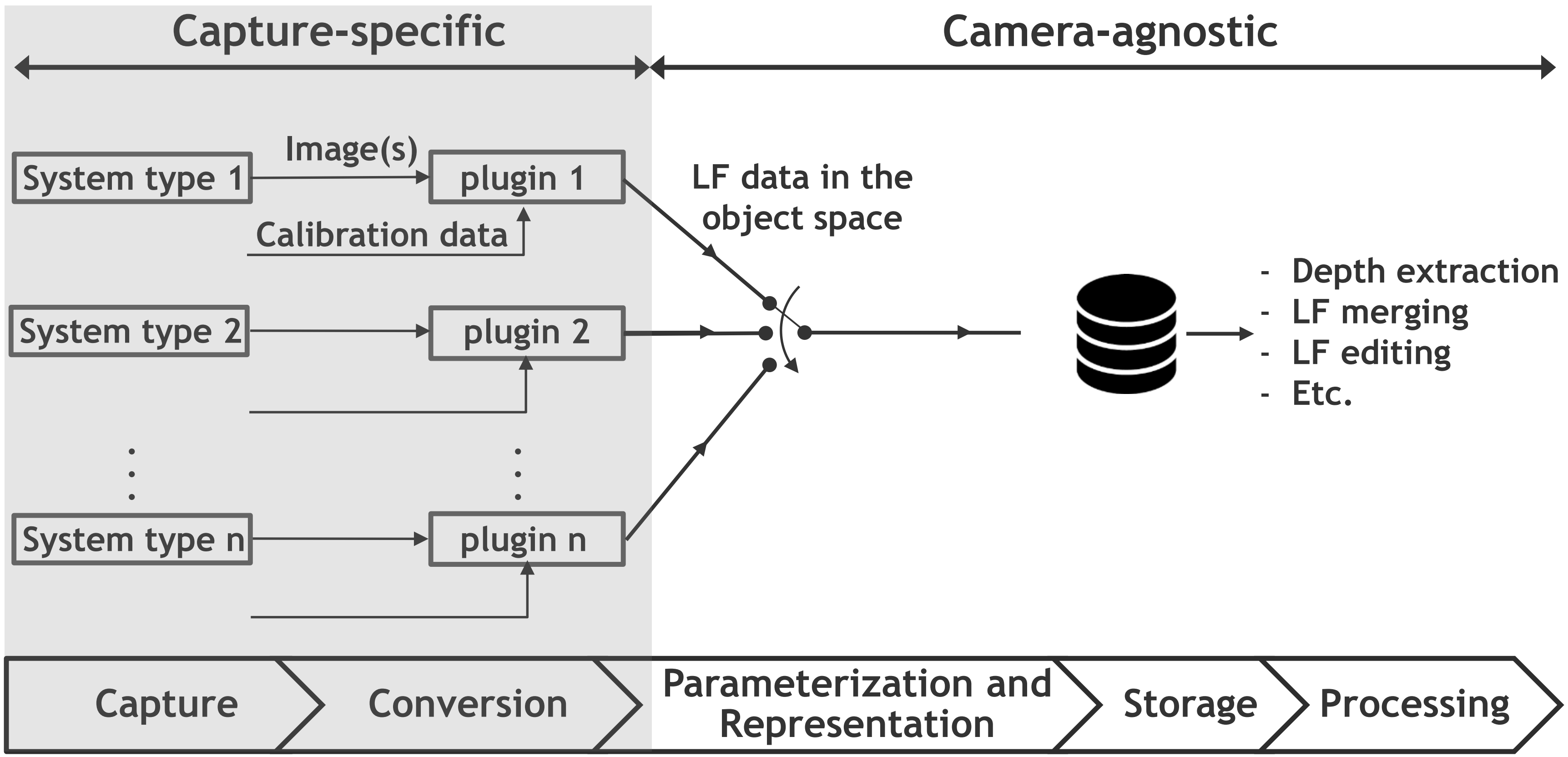
- Prevent cross-platform approaches
- Disperse investment/progress between algorithms
- Non-generic solutions (e.g. high quality results but only for one particular setup)
- Constrain standardization process

technicolor

Our contribution

- **Proposing** a camera-agnostic pipeline for storage and processing of LF data
- **Demonstrating the feasibility** of such pipeline

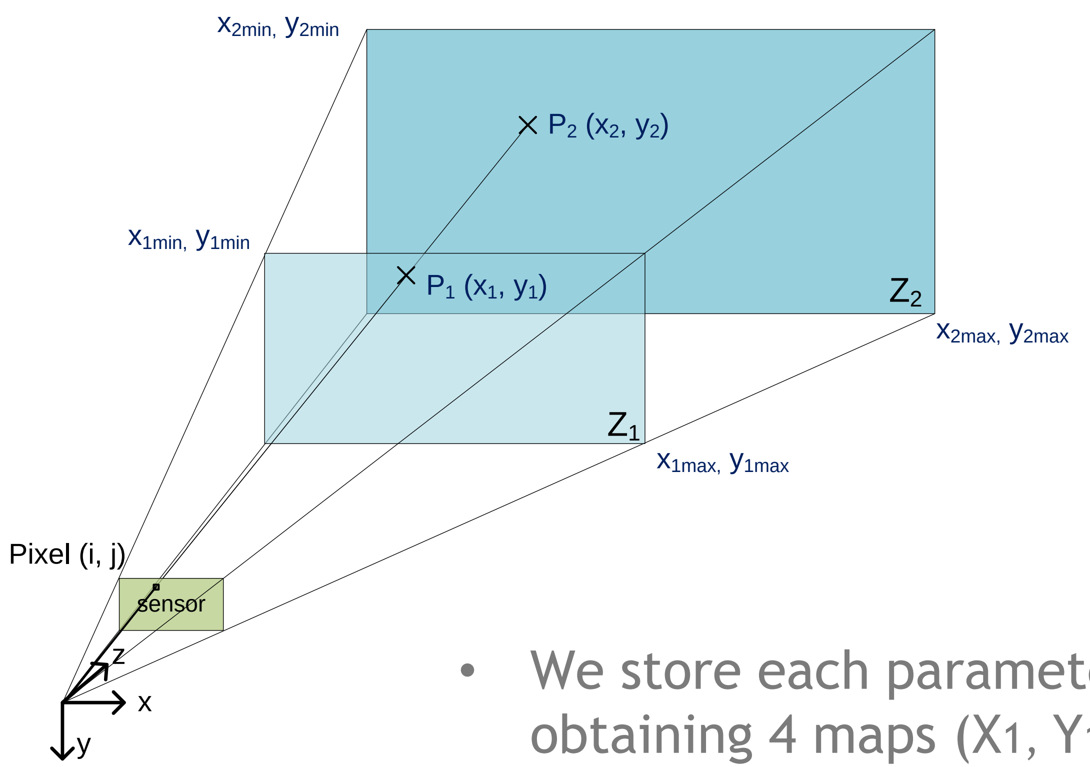
Camera-agnostic Pipeline for Light-field Data



technicolor

LFform - 2 planes parameterization

- LF-Form is a format to store description of 4D rays in a way compatible with sensor organisation.
- It allows to encode the 4D rays description with conventional codecs.



- We store each parameter (x_1, y_1, x_2, y_2) in a map, obtaining 4 maps (X_1, Y_1, X_2, Y_2)

technicolor

LFform – storage optimization for transmission

For each coordinate map

i. we compute the equation of the main plane that is, we estimate the parameters α, β, γ so that for we minimize the error $\left\| (\chi_{(u,v)} - \alpha\ u - \beta\ v\ - \gamma) \right\|$ where χ is X1, Y1, X2 or Y2 and u, v are the sensor coordinates of each pixel.

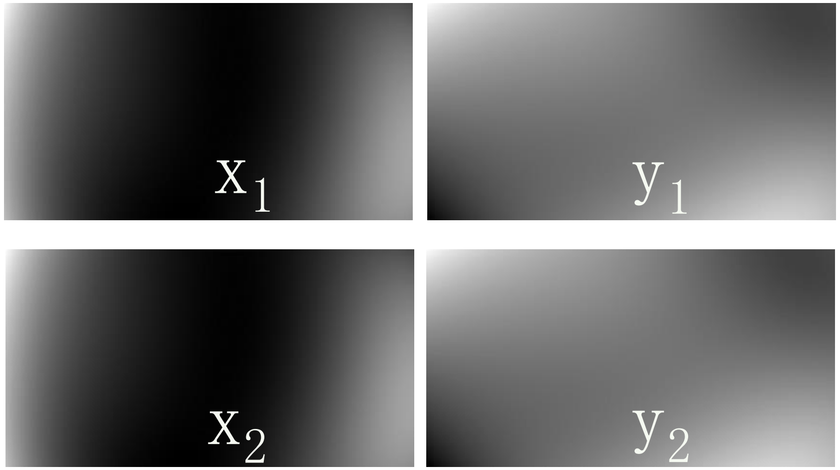
ii. Then, we can compute the residue (difference with the main plane) $\chi'_{u,v} = \chi_{(u,v)} - (\alpha\ u + \beta v +\ \gamma)$, resulting in a much lower range of amplitude.

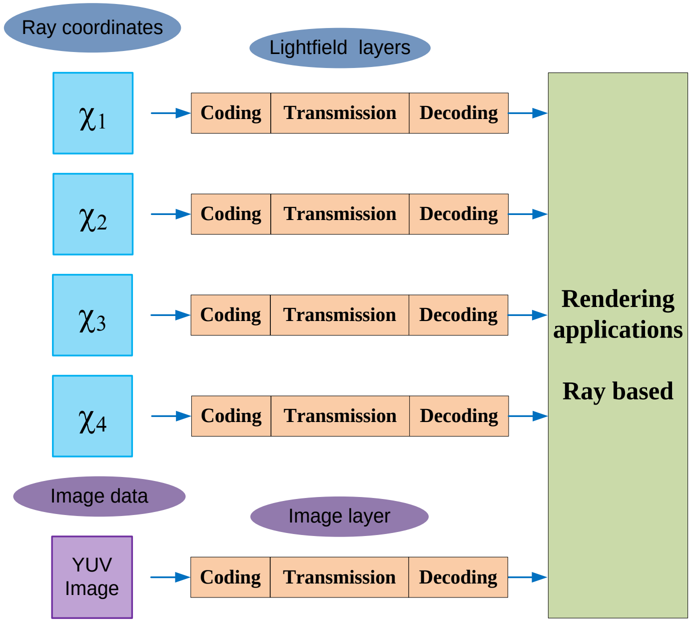
iii. The last step consist in reduction and quantization of the residue $\chi''_{u,v} = (2^N - 1) \dfrac{\chi'_{u,v} - \min(\chi'_{u,v})}{\max(\chi'_{u,v}) - \min(\chi'_{u,v})}$ where N is a number of bits chosen for an equilibrium between minimization of errors versus the capacity of the storage/transmission channel we intend to use.

All transformation parameters (α, β, γ, $\max(\chi'_{u,v})$, $\min(\chi'_{u,v})$, N) are kept as metadata that will be used to restore the 4D parameters after transmission and/or storage.

technicolor

LFform - Transmission

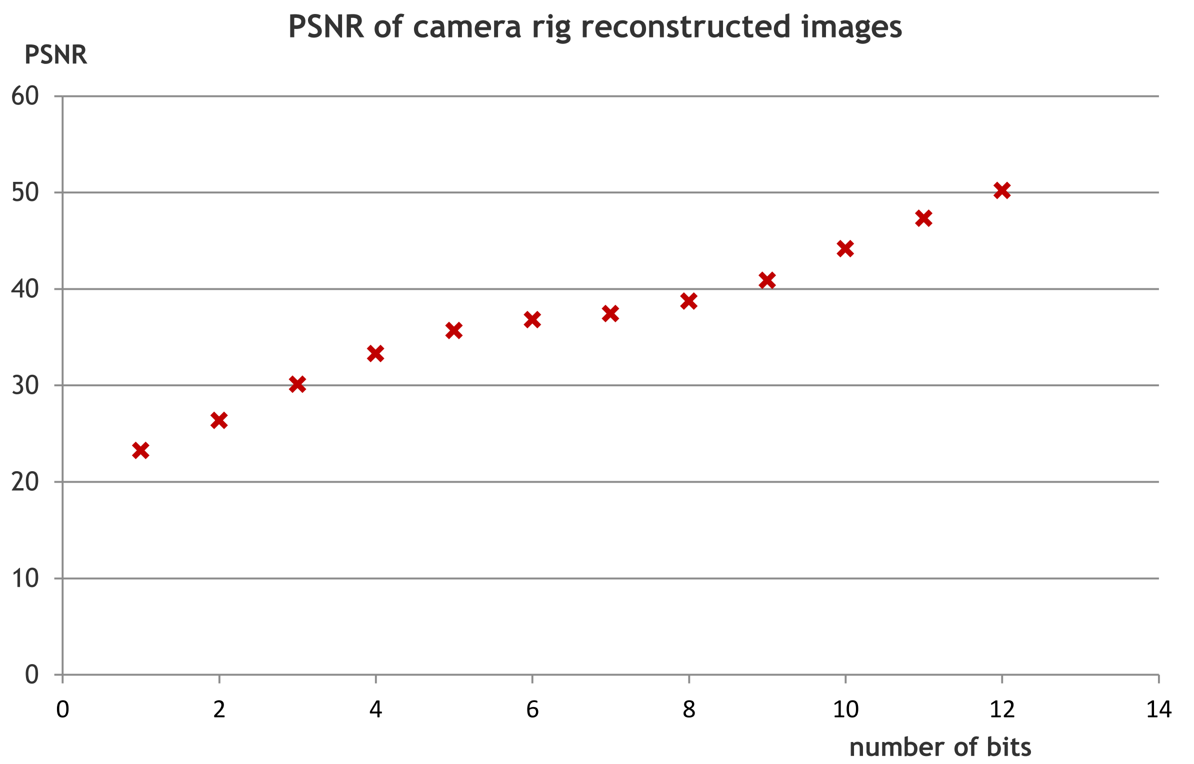




Example of the four maps obtained for a camera after information reduction and quantization on 8 bits

technicolor

LFform – Evaluation of transformation impact



PSNR of refocused images using components maps quantized with different bit depth.

technicolor

Camera-agnostic Processing, Depth Extraction

Depth extraction algorithm using color proximity

- Dividing the space into volume elements: (x, y, z)
- Defining the set of rays intersecting each volume element: $V(x, y, z) = \{r_1, \dots, r_n\}$

- Computing color proximity for each two rays *i*, *j* in each voxel: $\sqrt{(R_i - R_j)^2 + (G_i - G_j)^2 + (B_i - B_j)^2}$

- For each ray in each voxel, computing the number of rays in its color vicinity, c_i with regards to a threshold ρ:

$$c_i = \left|\left\{r_i \in \mathrm{V(x,y,z)}: \sqrt{(R_i - R_j)^2 + (G_i - G_j)^2 + (B_i - B_j)^2} \le \rho \right\}\right|$$

- Defining a local fitness function: $f(x, y, z) = max(c_i)$
- Defining a fitness function *F* over a window of size *(2n + 1) x (2n + 1)* and centered at *(x, y)*

$$F(x, y, z, n) = \sum_{i,j=-n,\dots,n} f(x + i, y + j, z)$$

- The depth value d is then assigned for each pixel position *(x,y)* by finding the depth with maximum fitness *F* along the *z* dimension:

$$d(x, y) = z|_{F(x,y,z,n) = max\,(F(x,y,z,n))}$$

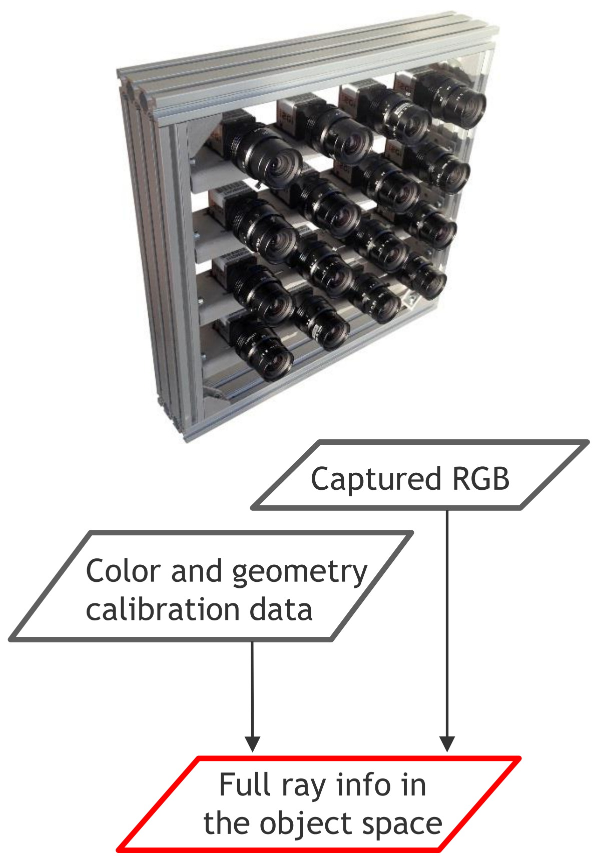
- Define a confidence measure *C* for each pixel (x,y) the z value of the first and third maxima of *F* along *Z* axis at pixel position (x,y):

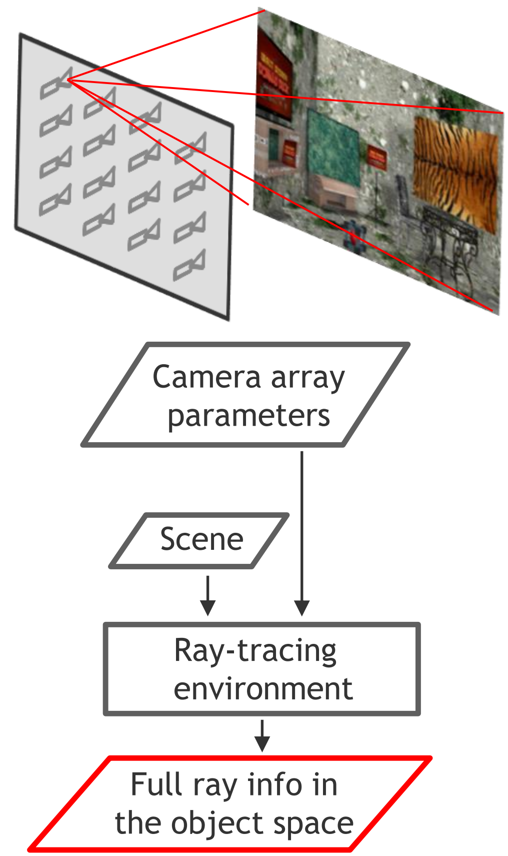
$$C(x, y) = \frac{z|_{max_3\,(F(x,y,z,n))} - z|_{max_1\,(F(x,y,z,n))}}{z|_{max_3\,(F(x,y,z,n))}}$$

- Finally, filtering the depth image using the confidence map. For pixel *(x_i, y_i)* with confidence below a defined threshold, we replace *d(x_i, y_i)* with the depth of the neighboring pixel (*i-1* to *i+1*) which has the highest confidence.

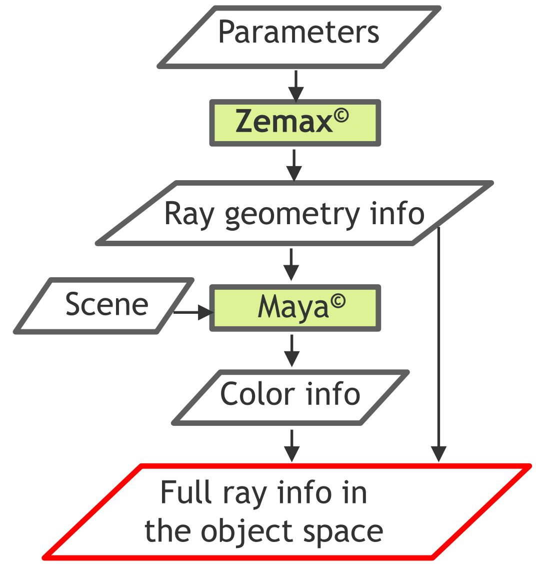
technicolor

Acquisition Systems





Parameter	$T3$	$T4$
Pixel size (µm)	1.4	6.8
Pixel per micro-lens	10 × 10	73 × 73
Micro-lens count	392 × 328	100 × 76
Micro-lens pitch (µm)	13.89	496.6
Micro-lens focal (µm)	25	1500
Sensor to micro-lens (mm)	25	1600
Main lens focal (mm)	35	80
Micro-lens to main lens (mm)	35	104



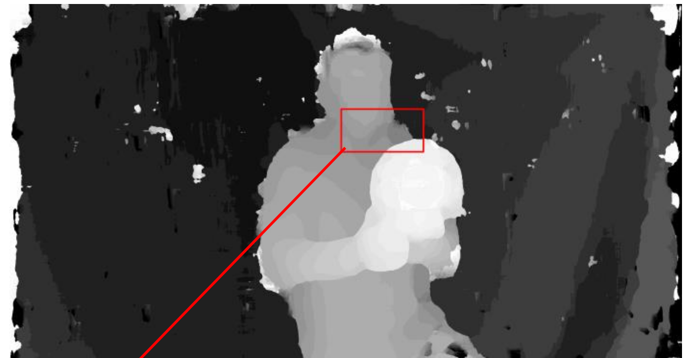
Camera array *T1*	Camera array *T2*	Focused Plenoptic *T3*	Lytro-like Plenoptic *T4*
Physical	CG		

- ✓ Diverse setups (sampling)
- ✓ No prior depth extraction algorithm that works for all

technicolor

Experimental Results, Physical Camera-rig

Scene



Extracted depth map

Depth details

Absolute difference of depth results
(from LF before and after coding)

Algorithm parameters, T1

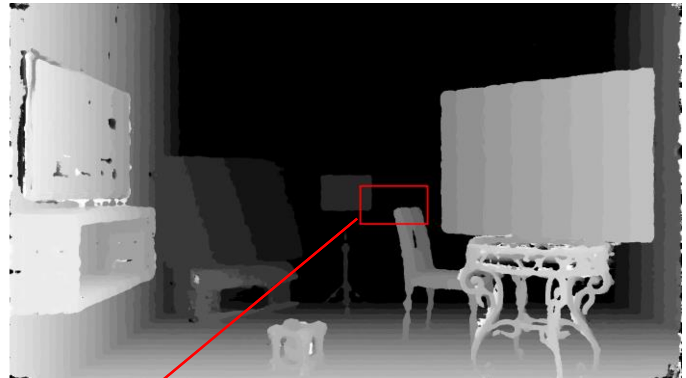
z_{min} (m)	2.25
z_{max} (m)	4.50
Depth planes #	30
$x \times y$	2048 × 1088
ρ	24
n (pixel)	19

ρ threshold for color proximity
n size of the neighborhood

technicolor

Experimental Results, CG Camera-rig

Scene



Extracted depth map

Depth details

Absolute difference of depth results
(from LF before and after coding)

Algorithm parameters, T2

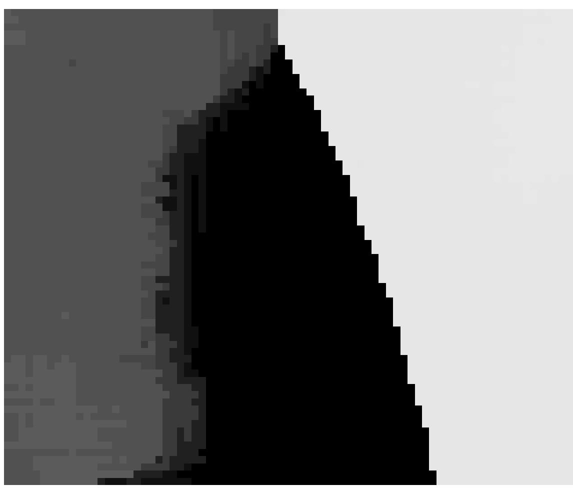
z_{min} (m)	1.20
z_{max} (m)	3.00
Depth planes #	30
$x \times y$	1920 × 1080
ρ	4
n (pixel)	7

ρ threshold for color proximity
n size of the neighborhood

Experimental Results, CG Lytro-like Plenoptic

Scene

Extracted depth map



Depth details

Absolute difference of depth results
(from LF before and after coding)

Algorithm parameters, T3

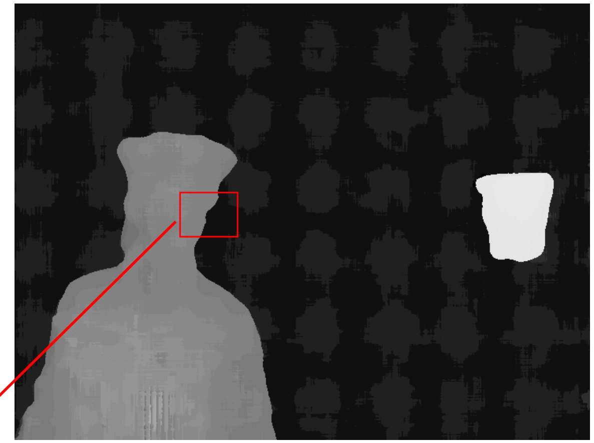
z_{min} (m)	0.80
z_{max} (m)	3.30
Depth planes #	60
$x \times y$	392 × 328
ρ	2
n (pixel)	9

ρ threshold for color proximity
n size of the neighborhood

technicolor

Experimental Results, CG Focused Plenoptic

Scene



Extracted depth map

Depth details

Absolute difference of depth results
(from LF before and after coding)

Algorithm parameters, T4

z_{min} (m)	0.80
z_{max} (m)	3.30
Depth planes #	60
$x \times y$	730×554
ρ	4
n (pixel)	9

ρ threshold for color proximity
n size of the neighborhood

technicolor

Experimental results, camera-agnostic

- ✓ No post-processing treatment
- ✓ No explicit knowledge of the acquisition system

Summary

1. Various acquisition systems:
   - Physical camera array (in-house 4X4 camera-rig)
   - CG light-field (Lytro-like plenoptic camera)
   - CG light-field (focused plenoptic camera)
   - CG camera array
2. Parameterization and representation of LF in the object space
3. Storing and restoring geometry and color info (LFpack and LFform)
4. Depth extraction from LF data in the object space
5. Experimental results

✓ Feasibility of a camera-agnostic pipeline for LF data

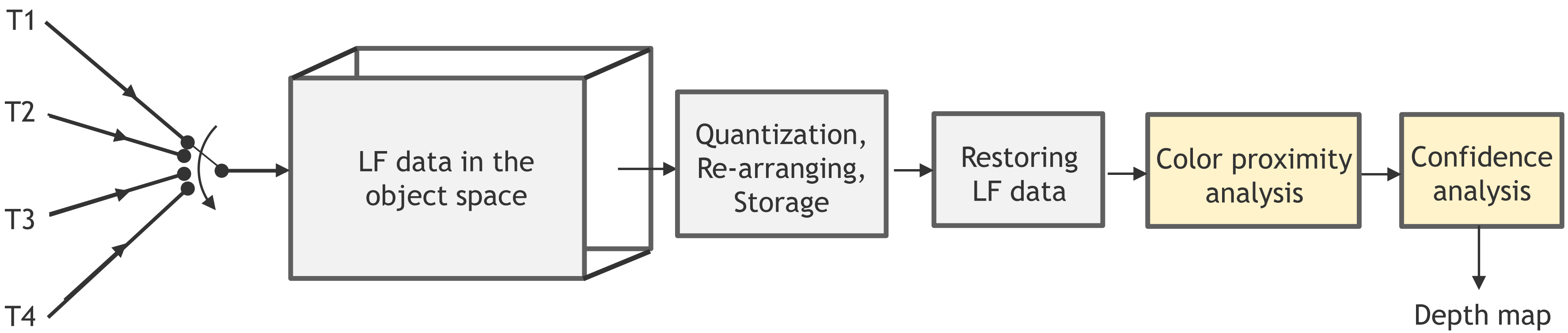
technicolor

Main Future Tracks

- Camera-agnostic compression format of LF data
- A toolbox for camera-agnostic light-field editing and processing

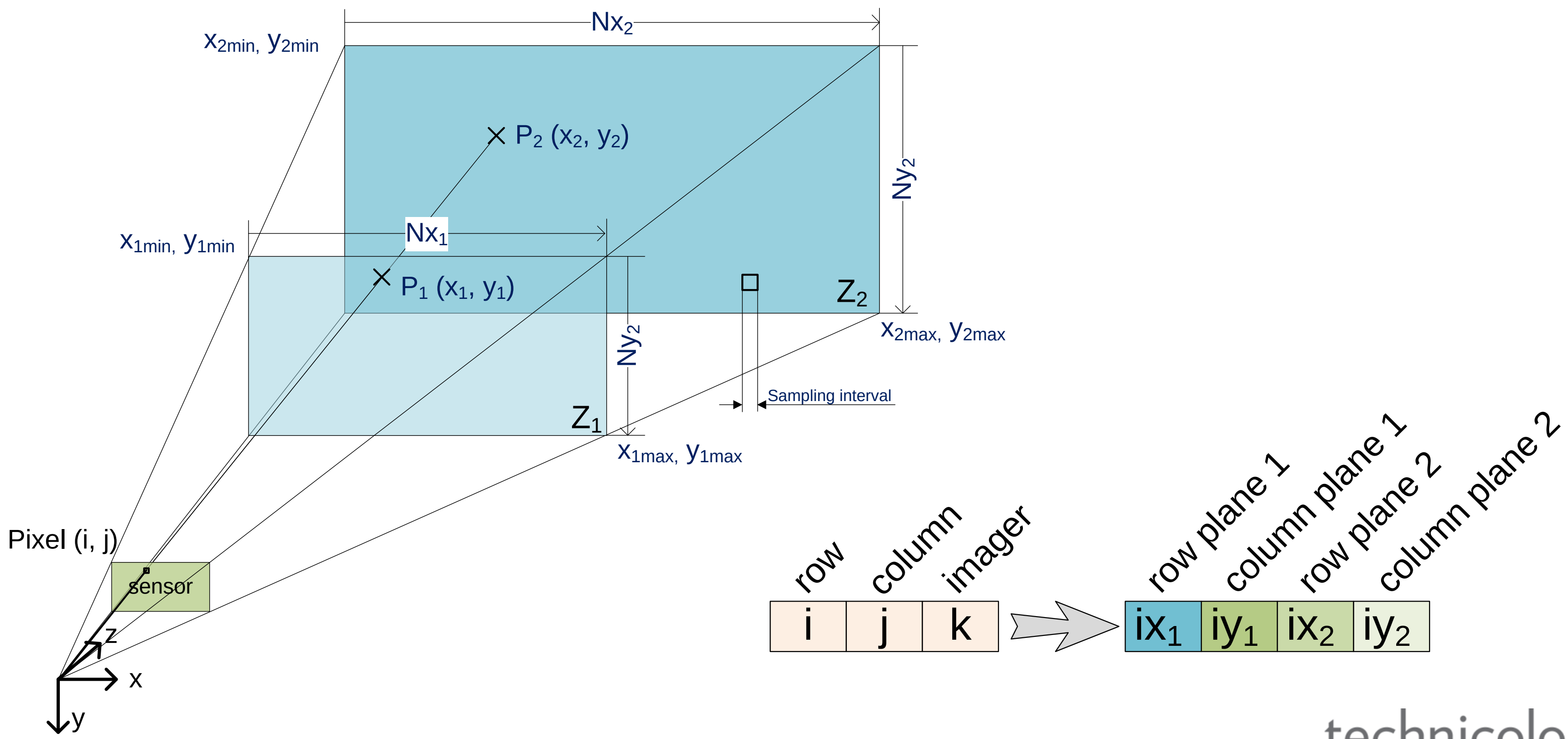
technicolor

Applying the Camera-agnostic Pipeline

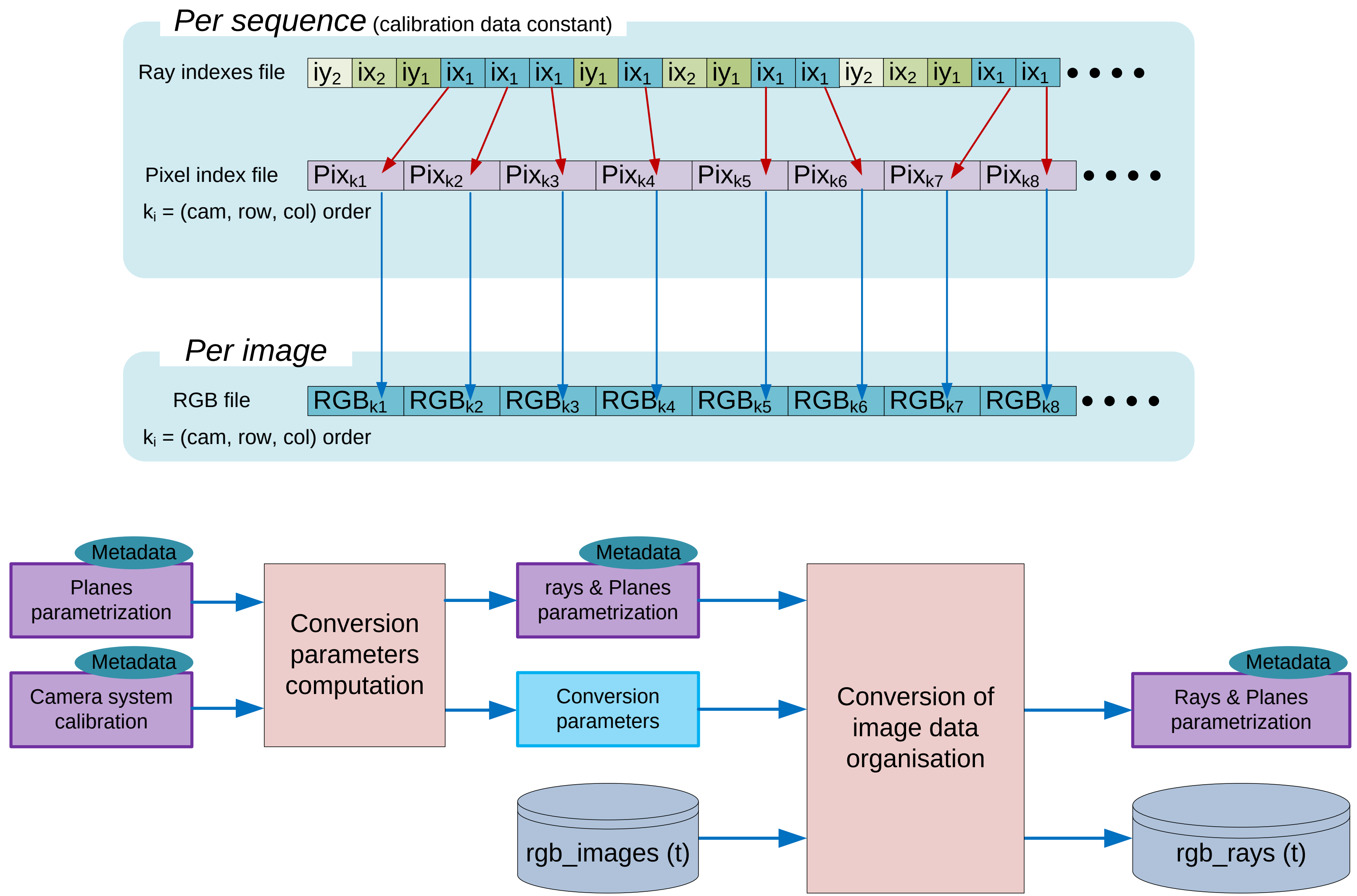


LFpack - 2 planes parameterization and quantization

- LFpack is a format to store description of 4D rays in a compact way
- Need to reorder the RGB values according to the ray ordered suite



LFpack – Pixel reordering

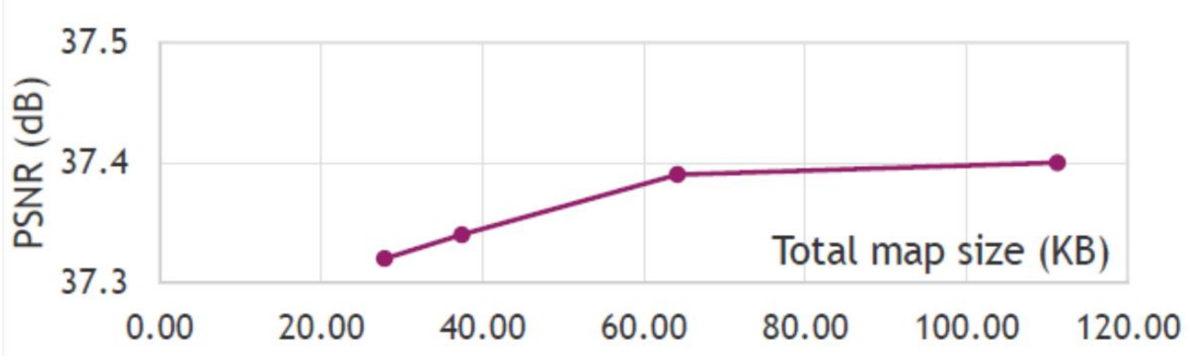


LFpack – Storage size and impact on refocused image

Inputs	
Total number of pixels (rays)	35 651 584
Size (4 B per coordinates)	570 425 344
Outputs	
Number of 4D cells	
$(N_{x1}, N_{y1}, N_{x2}, N_{y2})$	(944, 572, 3695, 2106)
Number of indices	73 369 782
File size (4 B per index)	293 479 128
Size after huffman coding (bytes)	71 356 799

Example of storage size for 35 million rays (multicamera-rig).

PSNR of the refocused image after “coding/decoding” maps.



technicolor