Chiara GALDI[a], Lara YOUNES[b], Christine GUILLEMOT[b], Jean-Luc DUGELAY[a]

[a] Digital Security Department
EURECOM, Sophia Antipolis, FRANCE
{chiara.galdi, jean-luc.dugelay}@eurecom.fr

[b]Inria Rennes - Bretagne Atlantique
Campus universitaire de Beaulieu, Rennes, FRANCE
{lara.younes, christine,guillemot}@inria.fr

EURECOM Sophia Antipolis
Inria inventeurs du monde numérique

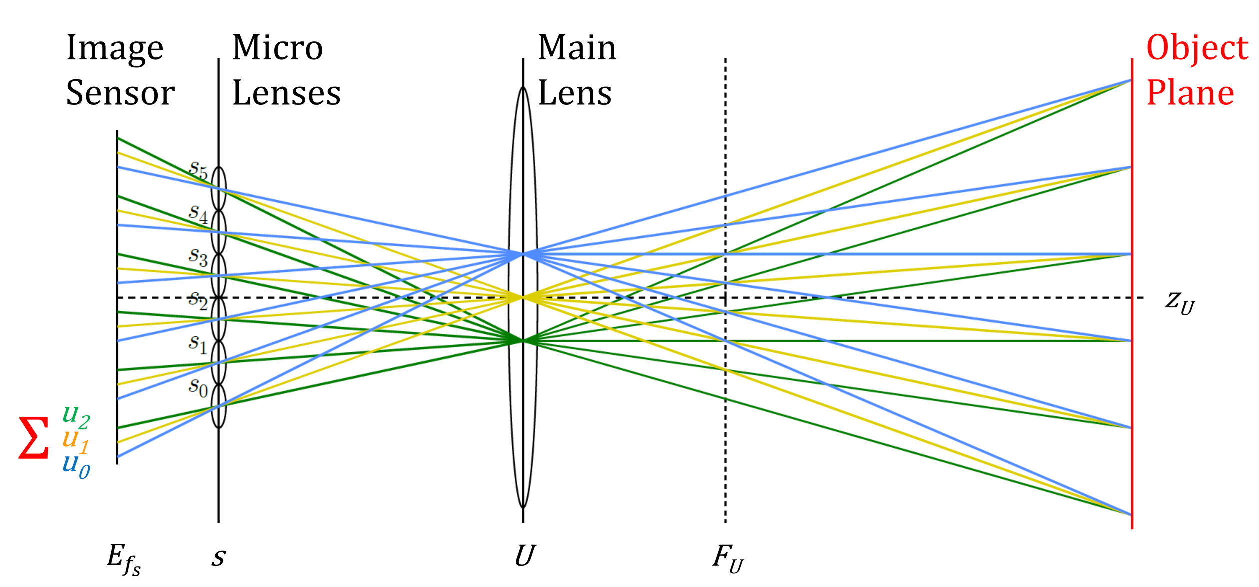
A new framework for optimal facial landmark localization on light-field images

Light fields

The term "**plenoptic**" comes from the Latin words plenus ("full") + optic. The plenoptic function is the 5-dimensional function representing the intensity of the light observed from every **position** and **direction** in 3-dimensional space. Thanks to the plenoptic function it is thus possible to define the direction of every ray in the **light field vector function**.

Compared to classical 2D images, **light fields** capture the intensity values **along each ray** and not only the sum of intensities of rays reaching each image point:
- possibility of **changing the angle** of view;
- reconstructing the **scene depth** (depth map);
- and **refocusing** on the desired object in the scene.

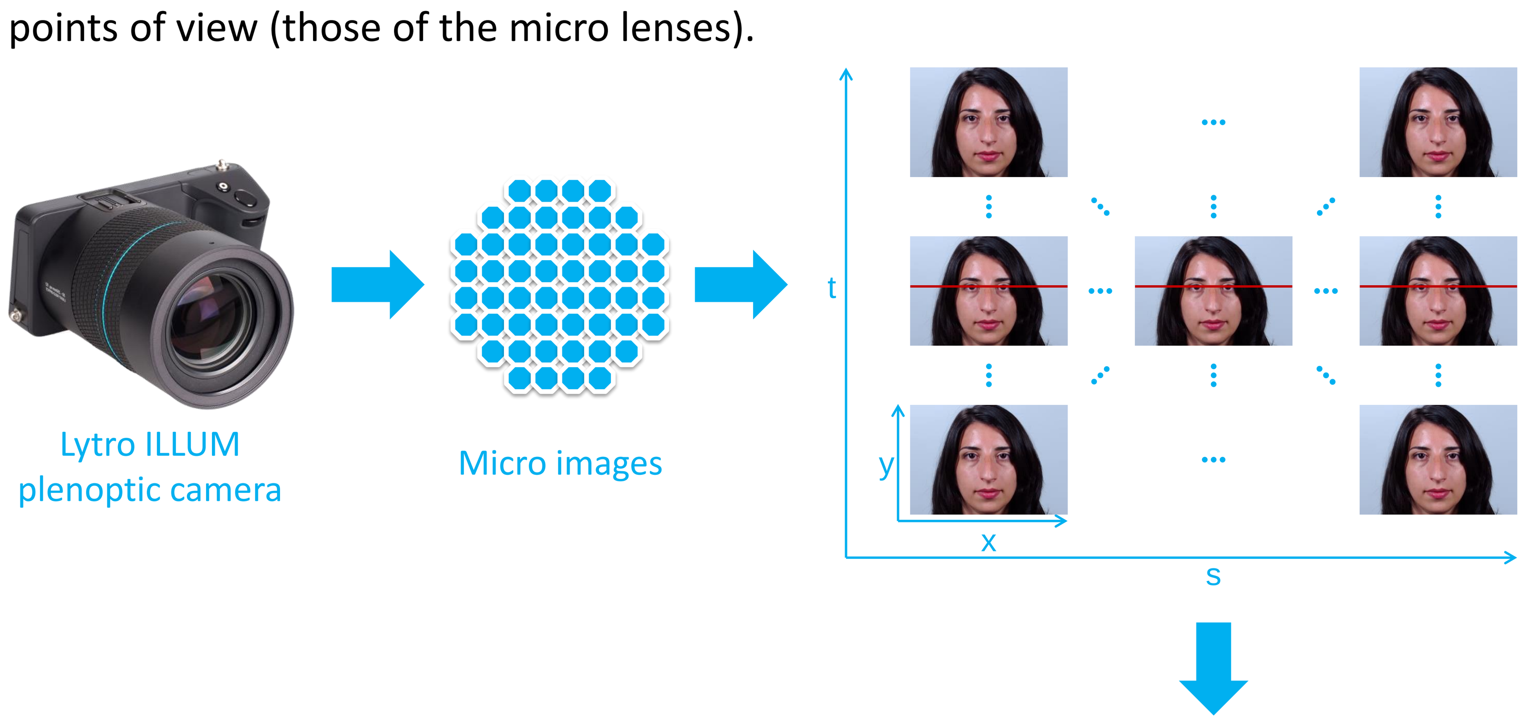
Scan the QR code to read more and watch an animated GIF



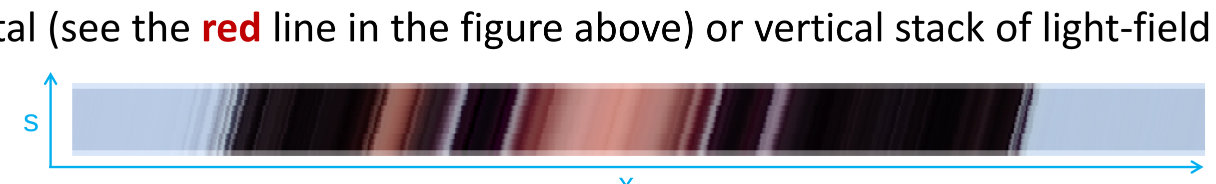
Scan me

Epipolar plane image (EPI)

The plenoptic camera captures a mosaic of micro images that after processing is translated into an array of images depicting the captured scene from different points of view (those of the micro lenses).



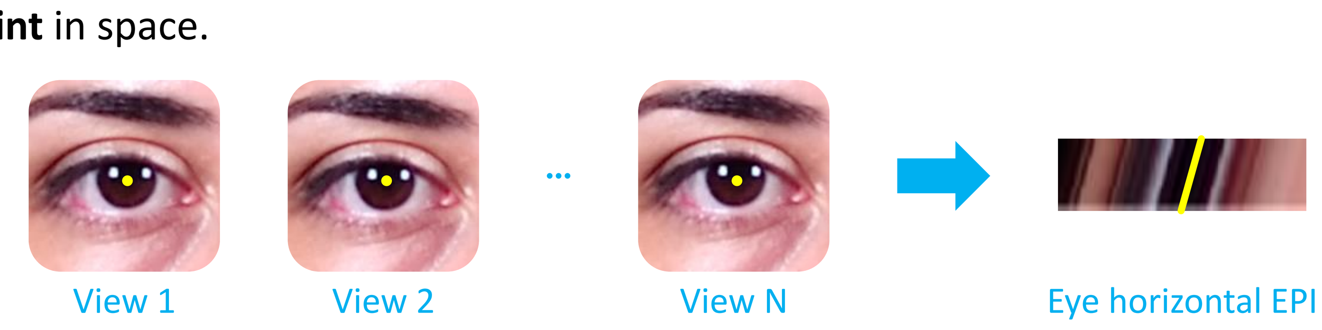
The **EPI** can be represented as a spatio-angular slice of the light field, cut through a horizontal (see the red line in the figure above) or vertical stack of light-field views.



Example of horizontal epipolar plane image (EPI): the pixel vectors corresponding to the horizontal red lines are stacked to form the EPI.

Observation

Following the inherent structure of light fields, the points laying on the **same level line** correspond to the projection, in the different angular (s) views, of the **same 3D point** in space.

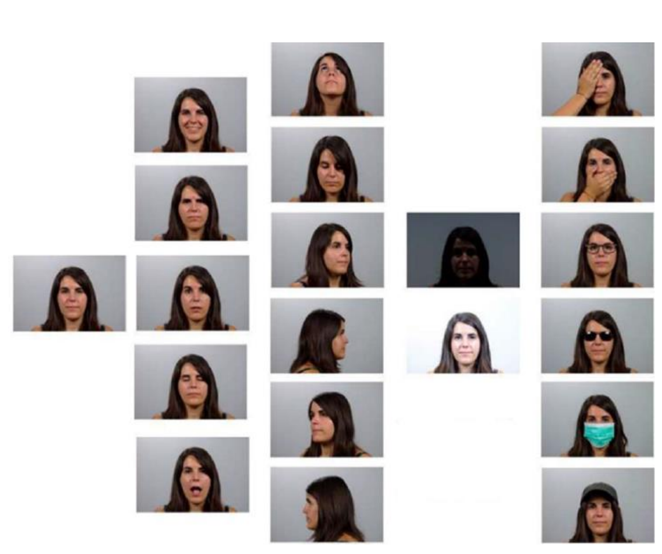


If a given landmark detector performs likewise over all the views of the light-field face image, the **detected landmarks should lay on the same level line in the EPI image**.

Data

Data come from the **IST-EURECOM Light Field Face Database***, a face images database consisting of photos of 100 different persons, captured with **a plenoptic camera** in different face variations and illumination conditions.
> http://lffd.eurecom.fr/

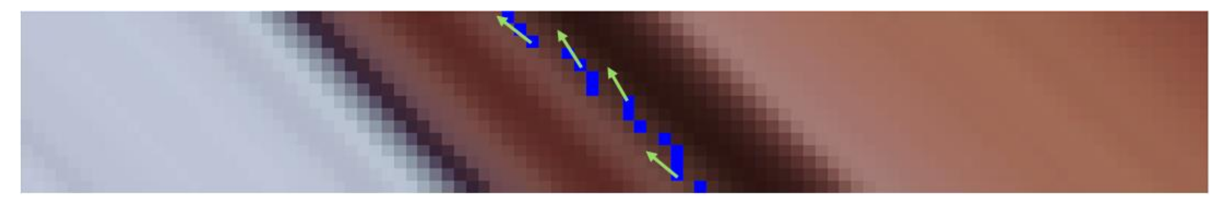
*A. Sepas-Moghaddam, V. Chiesa, P.L. Correia, F. Pereira, J. Dugelay, "The IST-EURECOM Light Field Face Database", International Workshop on Biometrics and Forensics, IWBF 2017, Coventry, UK, April 2017



Method

The idea is thus to detect the level lines on which the detected face landmarks lie and use them to estimate the "true" level line. The coordinates of the points are then corrected accordingly.

Level lines are computed using the **Structure Tensor**.



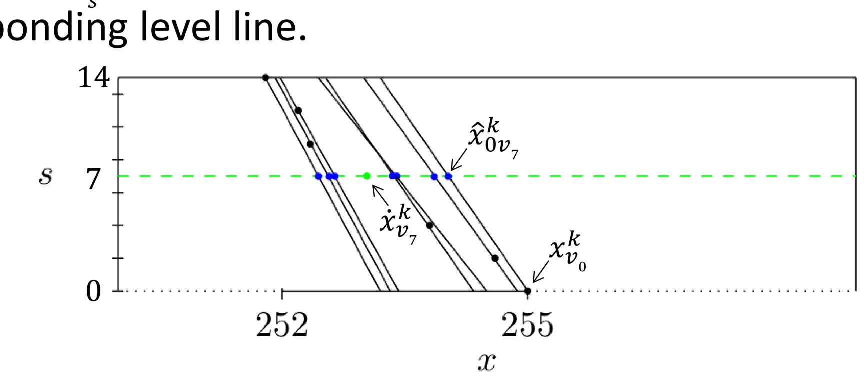
Detected landmarks (blue points) and corresponding director vectors for the level lines computed from the structure tensor (green arrows).

DEF: The structure tensor, also referred to as the second-moment matrix, is a matrix derived from the gradient of a function. It summarizes the predominant directions of the gradient in a specified neighborhood of a point, and the degree to which those directions are coherent.

Given a face landmark k, let's consider the **black** points in the figure below its position estimation over the views (v_s). For each black point, its level line is computed using structure tensor (**black** lines). The black points are then projected following the black lines to the central view (**blue** points, $\hat{x}^k_{sv_c}$). The coordinates of the point on the central view (v_c) are then corrected using weighted sum (**green** point):

$$\dot{x}^k_{v_c} = \frac{\sum_{s=0}^{P} w^k_{v_s} \cdot \hat{x}^k_{sv_c}}{\sum_{s=0}^{P} w^k_{v_s}}$$

where the weight $w^k_{v_s}$ is the number of black points at a distance less of 0.1 pixels form the corresponding level line.



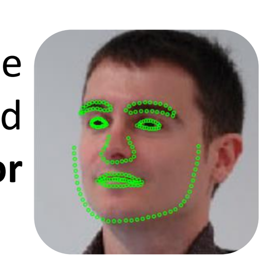
Experiments

A set of light-field face images (50 subjects × 2 sessions × 4 pose variations = **400 images**) has been selected from the IST-EURECOM Light Field Face Database.

For each light field, 15 horizontal and 15 vertical views have been extracted with regular angular sampling in the perspective range [−0.5, 0.5] thanks to the LYTRO POWER TOOLS BETA. For a total of **12 000 images**.

Faces on the central views have been manually annotated with 32 landmarks to constitute a set of **ground-truth** points.

Face landmarks are detected using DLIB, and then corrected with the proposed method. The original detected coordinates and the corrected ones are then evaluated using **normalized root mean square error** (NRMSE). DLIB > http://dlib.net/



Results

The NRMSE between the ground-truth coordinates (x, y) and the estimated coordinates $(\tilde{x}, \tilde{y})$, is defined as:

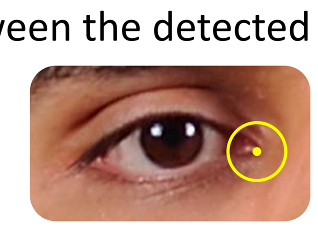
$$\delta^k_v = \frac{d\{(x^k_v, y^k_v), (\tilde{x}^k_v, \tilde{y}^k_v)\}}{IOD}$$

Where $d()$ indicates the Euclidean distance, k indicates the landmark index (e.g. eye corner, nose tip), v is the image angular coordinate (view), and IOD is the inter ocular distance.

The overall landmark detector performance in terms of percentage of detected landmarks, is computed by the following formula:

$$P = 100 \frac{\sum_{k=1}^{K} \sum_{i=1}^{I} [i: {\delta_i}^k < Th]}{K \times I}$$

Where $[i: {\delta_i}^k < Th]$ is the indicator of value 1 if the distance is smaller that *Th*, 0 otherwise. *I* is the number of test images and *K* the number of landmarks per face image. And δ is the NRMSE between the detected landmark and the ground truth.



Example of error threshold for Th = 0.1 (1/10 of *IOD*)

Scan the QR code to watch an animated GIF correction example >

Scan me

	Neutral Frontal Face		Action Mouth Open		Pose Up Looking		Pose Half-profile Left	
	Original	Corrected	Original	Corrected	Original	Corrected	Original	Corrected
P(%)	97.81	**98.11**	95.70	**96.37**	91.80	**92.66**	77.68	**79.13**