

# Notes for thesis defence beamer.

- [Presentation of myself and thanks for the public].

## Slide 1: What is this thesis about?

- A sparsely sampled light field reconstruction method of a static scene.
- This method uses Universal Shearlets to perform an inpainting algorithm in the sparse Epipolar plane images (EPIs).
- The EPIs are images with linear structures obtained by tracking different feature points in a sequence of pictures of the scene.
- The slope of the straight lines in the EPIs can be used to compute the depth of feature points in the scene and therefore the depth map that contains the 3D information of the scene.
- The thesis presents all the steps in the reconstruction pipeline with theory, algorithms and code.
- [Say that I will explain every concept in the next with a limit of time, so one can find a detailed explanation on the thesis].

## Slide 2: Involved concepts.

- Early Vision (biology)[study of how the visual cortex works in order to form the visual information of the world] and Light Field Theory (physics) [how the light as an electromagnetic wave propagates in the 3D space from a scene elements to the human cortex].
- Computer Vision for point tracking and line detection ([which uses] basic concepts of linear algebra).
- Functional Analysis and Computational Harmonic Analysis ([concept of] sparsifying dictionaries [as wavelets and shearlets]).
- Compressed sensing techniques ([as]  $\ell^1$  optimization algorithm).

[When I started to design this thesis I needed to learn light field theory in particular light field reconstruction methods from scratch so I ended up ordering the method on the thesis in the way that they were needed as steps of a pipeline,

## Slide 3: Light Field Theory.

- In 1846 Michael Faraday proposed for the first time on his lecture "Thoughts on Ray Vibrations" that light could be interpreted as a field [this based on his work about electromagnetic fields, but this idea was formalized by Adelson and Bergen until 1991].

- [They proposed that...]The propagation of light rays in the 3D space is completely described by a 7D continuous function  $L(x,y,z,\theta,\phi,\lambda,\tau)$  called the plenoptic function [where  $(x,y,z)$  is the location in the 3D space,  $\theta$  and  $\phi$  are the propagatin angles,  $\lambda$  is the wavelength of the corresponding light wave, and  $\tau$  the time. This function describes the amound of light flowing in every direction through every point in space at any time, the magnitud of  $L$  is know as the radiance, hat is the radiant flux emitted or power (energy per second), reflected, transmitted or received by a given surface per unit solid angle per unit projection area with  $Wsr^{-1}m^{-2}$ ]
- [By fixing  $\lambda$  and  $\theta$ ] The plenoptic function can be simplified to a 4D function  $L_4$ , called 4D Light Field or simply Light Field, which quantifies the intensity of static and monochromatic light rays propagating in half space.
- [This looks like an important reduction of information, but his contraint does not limit us in the accurate 3D description of a scene from where the light rays come from, The Light Field of a 3D scene (subspace) describes the 3D information like the depth of the points in the scene and can be recovered by taking multiple pictures (views) of the scene. The goal of this thesis is to explain a particular fast technique for this purpose).
- [There are different ways to represent a 4D Light Field].

## Slide 4: 4D Light Field Representation.

- [From Left to right three different 4D Light Field represenations. Left: LF rays positions indexed by their Cartesian coordinates on a plane and the directional angles leaving each point. Center: LF rays positions are indexed by pairs of point on the surface of a sphere. Right: LF rays positions are indexed by their Cartesian coordinates on two parallel planes, also called two plane parametrization].
- [Due computational low complexity we used the "Two plane parametrization", how to understand it: Consider a camera with image plane coordinates  $(u,v)$  and the focus  $f$  moving along the  $(s,t)$  plane].
- [Once we have clear what is the light field and Before analysing more deeply the method used to recover it I would like just to present some of the motivation that drive me to study this topic which includes, historical work and interesting applications].

## Slide 5: Motivation.

## Slide 6: Compression of High Resolution Light Field (Wetzsetein et al., 2013).

- [First time I heard about light field was when I found a paper made by MIT Media Lab - present the paper-, I personally find always very interesting all the work made by MIT Media Lab, due its very creative way to tackle fun problems; on this paper the "Camera Culture Group" leaded by Prof. Gordon Wetzstein proposed a method to compress high resolution light fields using masked acquisition -they insert a mask on the lens of a camera to simulate having an array of lenses that take different views of the scene- and learning a sparsifying dictionary called light field atoms, though the paper is very well explained, the approach seemed to me very complicated and it required also very complex and bulky hardware].
- [The most interesting features that I found in this paper was the possibility to perform a refocusing of a scene, when knowing the light field in terms of the depth map].
- [Show the refocusing program by python].
- [After this encounter with the Light Field photography, I searched for a less bulky hardware that could be used to acquire light field and I found two companies that developed commercial light field cameras -also called plenoptic cameras-].

## **Slide 6: Raytrix (Perwass and Wietzke, 2010).**

- [The first commercialized plenoptic cameras were developed by the german computer scientist Christian Perwass and Lennart Wietzke, they created the company Raytrix (based on Kiel) that develops since then plenoptic cameras with array of lenses mostly focused on industrial applications, they have very high resolution and their cheapest camera costs 3500 euros].

## **Slide 7: Lytro (Ng, 2012).**

- [While in Germany people were producing industrial Light Field cameras, a Stanford PhD Student in Computer Science and photography aficionado, Ren Ng was developing the first light field camera focused on general public. He created a company Lytro based in Silicon Valley and developed the Lytro cameras in 2012 that also use lenses arrays and that cost around 500 euros, which is still too expensive for my budget].
- [At that moment I was developing a Julia version of the Shearlet Transform library developed by Professor Kutyniok Group, Shearlab so I was interested in the question, could we use Shearlets to recover a sparse sampled light field, in other words, can we compress the Light Field using Shearlets].

## **Slide 8: Shearlet Representation of LF (Vagharshakya et al.,**

**2015).**

- [After this question I actually found a paper that did exactly that].
- [Why to do something that is already done?].
- [Mathematical curiosity: best way to learn how the things are done is actually do it your self].
- [The commercial cameras are very expensive.]
- [All the papers that I found on this topic, including the ones related to the Lytro and Raytrix cameras, are very obscure when presenting the technical parts; the mathematics behind the reconstruction methods are clear, but they do not give their data set, their code or the side steps to actually go from a raw sequence of pictures to a depth map representing the light field. They also use very slow-high level privative software for some computer vision tasks that can be done in a free and easier way. The reason for this black boxes is commercial interest].

## **Slide 9: Freedom of knowledge.**

- [Present Donoho and Buckheit quote and explain that this is the main philosophy followed in the thesis].
- [Now that we know the motivation and the philosophy behind this thesis we can proceed on the theory behind the method we are presenting].

## **Slide 10: Stereo Vision and multiview Epipolar Geometry.**

- The human brain generates the 3D depth perception by triangulating the points of a scene using the information coming from both eyes [that can be interpreted as cameras] [humans can actually percieve 3D with one eye, using the so called "pinhole effect", but this is more complex to explain and does not concern this thesis].
- Epipolar Geometry: Generalization of Stereo Vision with more than two views interpreted as [images taken by a camera that moves withing a trajectory], assuming the epipolar constraint.
- Epipolar Constraint: Supply the separate analysis of both camera motion and object position by an unified treatment of parameters and concentrate solely in object position while knowing the camera followed trajectory.
- [The Epipolar Constraint reduces the complexity of the feature point matching in different views from two dimensions to one dimension (Epipolar Line).]
- [The application of Epipolar Geometry to describe the threedimensional information of an static scene was introduced in 1987, he used for this task a dense sequence of images of a static scene -dense in the sens that it forms a solid block of data in which tempoal

continuity from image to image is equal to spatial continuity].

### Slide 11: EPIs on straight line trajectories.

- [Epipolar lines are the same in the multiple views when the camera trajectory describes a straight line]
- [Tracking the different feature points in the scene corresponding to each of one of this lines we can obtain an image with linear structures that describe the points trajectories, this are called epipolar plane images].

### Slide 12: Functional analysis of EPIs.

- 3D Light Field: Fixing one coordinate in the focal  $\pi_0$ -plane of the two-parallel plane approach of 4D Light Field, one reduces the  $\pi_0$  to a line, and the resulting field  $L_3: \mathbb{R}^3 \rightarrow \mathbb{R}^3$  is called 3D Light Field with radiance  $\mathbf{r} = L_3(u, v, s)$  [this means fixing the camera movement to a straight line with  $t$  constant, the remaining dimension can be understood as the time]
- Epipolar Plane Image: By fixing on the 3D Light Field  $L_3$  the  $v$ - coordinate on the image  $uv$ -plane one obtain the field  $E_v: \mathbb{R}^2 \rightarrow \mathbb{R}^3$  known as the Epipolar Plane Image with radiance  $\mathbf{r} = E_v(u, s)$ .
- [Show the BMS picture, with the scenes at different times and the EPI-polar plane images].

### Slide 13: Depth map estimation with EPIs.

- [Show the picture of two vies of the lateral motion, and the formula how to measure the depth of points knowing the distance the moved in the epipolar line, with the focal distance.]
- [Show depth map formula]
- [Show the sparse sampling rate (Nyquist criterion): the sampling frequency is at least the double of the scene's highest texture frequency, that is maximum of 1 pixel disparity between nearby views to avoid aliasing and other artifacts]

### Slide 14: EPIs physical acquisition setup.

- [I did not have the hardware to acquire the pictures]
- Data set: Sequence of 101 rectified pictures of a scene generated by Professor markus Gorss' group in the Disney.

## Slide 15: Used scene.

- [Present image of first and last scene and show how the things are moved].

## Slide 16: Followed pipeline.

- [Present image of pipeline, mention that if I have time I can explain the Shi-Tomasi corner detector, the Point tracking algorithm and the Hough Line Transform]

## Slide 16.5.1: Point tracking results.

- [Present pictures of (336 strong corners) points detected and their traces].

## Slide 16.5.2: Particular EPI example.

- [Present pictures of EPI of points for particular EPI, with dense EPI and corresponding sparse EPI with sampling rate of 1/4]

## Slide 17: Reconstruction method with inpainting.

- [Present the image with the sampled lines on the EPIs]
- [Explain that reducing the number of samples it is reduced the complexity on the EPIs acquisition algorithm but adds a new problem, since the sparse EPI are not suitable to apply line detection algorithm to measure the slope of the lines and therefore is not useful to estimate the depth map].
- [We need to somehow patch the unknown parts that can be interpreted as lost parts of a corrupted image, this task is known as inpainting as is present in different applications on data processing, since frequently the technologies of data acquisition fail resulting on missing traces].

## Slide 17.5.1: Important Tool: Frame Theory.

- Definition (Frame): Let  $I$  be a set of countable indices. A sequence  $\{\phi_i\}_{i \in I}$  in a Hilbert space  $\mathcal{H}$  is called a **frame** of  $\mathcal{H}$ , if there exist constants  $0 < A \leq B < \infty$  such that

$$A\|x\|^2 \leq \sum_{i \in I} |\langle x, \phi_i \rangle|^2 \leq B\|x\|^2, \forall x \in \mathcal{H}$$

$A$  and  $B$  are called lower and upper frame bound. Moreover, if  $A$  and  $B$  can be

chosen to be equal, we call it  $(A-)$  tight frame. If  $A=B=1$  is possible,  $\{\phi_i\}_{i \in I}$  forms a Parseval frame [a particular case of parseval frame is the orthonormal basis, so frames generalize its concept with more flexibility]

## Slide 17.5.2: Untitled

- [Present the forms of the analysis and synthesis operator, the frame operator as well as the Reconstruction and decomposition formulas].

## Slide 17.5.3: Abstract Inpainting Framework.

- [There are different approaches to tackle the inpainting task, but we will use the one based on applied harmonic analysis combined with ideas of compressed sensing, this approach usually assumes that we know the positions and shape of the lost areas (referred as masks)].
- Let  $\mathcal{H}$  a separable Hilbert space and  $x^0 \in \mathcal{H} = \mathcal{H}_K \oplus \mathcal{H}_M$ ,  $M = P_K \mathcal{H} \oplus P_M \mathcal{H}$ . Then, given a corrupt signal  $P_K x^0$ , we want to recover the missing part  $P_M x^0$ .
- In image inpainting  $\mathcal{H} = L^2(\mathbb{R}^2)$ , the missing space  $\mathcal{H}_M = L^2(\mathcal{M})$  for some measurable set  $\mathcal{M} \subset \mathbb{R}^2$ . Given  $x_K \in \mathcal{H}_K$  we want to find  $x^0 \in \mathcal{H}$  such that  $x_K = P_K x^0$  (underdetermined problem). [seen as a mask covering the corrupted parts].
- [So is basically a separation problem where  $x$  and  $x_K$  are morphologically different].
- We will assume that  $x^0$  can be efficiently represented by some Parseval frame  $\{\phi_i\}_{i \in I}$  for  $\mathcal{H}$ , this is translated as asking for the solution of the  $\ell^0$ -minimization problem [give the problem and explain the  $\ell^0$  norm].
- [Explain that this minimization problem is not convex, so it is a big issue in the solution method].

## Slide 17.5.4: Analysis Approach.

- [Show the new approach that can be interpreted as the  $\ell^1$  relaxation of the original problem, explain that Algorithm 1 minimizes the  $\ell^1$  norm among all possible reconstruction candidates, since  $T\{\Phi\}$  gives you the frame coefficients of the candidates, and that the error can be estimated using Compressed sensing tools as  $\delta$ -cluster sparsity, Cluster coherence, which broadly says that if  $\Phi$  sparsifies  $x^0$  and the missing space does not contain important information of the signal then the algorithm outputs a good, the measurement is sufficiently decoherent, it also can be seen as a separation problem sinusoids and spike sampled at  $n$  points.]

- [ $\delta$ -cluster sparsity:  $\Gamma \subset I$ ,  $x$  is called  $\delta$ -clustered sparse in  $\Phi$  with respect to  $\Gamma$ , if  $\|1_{\Gamma^c} T(\Phi)x\| \leq \delta$ , and  $\Gamma$  is a  $\delta$ -cluster for  $x$  in  $\Phi$ , i.e. the analysis coefficients are highly concentrated on  $\Gamma$ . Cluster coherence  $\mu_c(\Gamma, P_M \Phi) = \max_{j \in \Gamma} \sum_{i \in \Gamma} |\langle P_M \Phi_i, P_M \Phi_j \rangle|$ , where  $P_M \Phi := (P_M \Phi_i)_{i \in I}$ , (abstract cmeasure for the gap size, maximal amount of missing information; We use the fact that is a parseval frame when defining a good norm to measure the error,  $\ell^1$  analysis norm, since parseval frame has injective analysis operator and therefore the  $\ell^1$  analysis norm define a Hilbert space, and is embedded in  $\mathcal{H}$  so small recovery errors in the analysis norm imply small recovery error in the original space. ]
- [We want know to choose a suitable representation systems which provides an optimally sparse representations of principally two morphologically distinct components, points and curves; present result by donoho, and wavelets case, explain why it is not sufficient just doing isotropic scaling in natural images].

### Slide 18.1: Shearlets.

- [Explain one solution, the shearing operator, the parabolic scaling operator].
- [Show the Discrete shearlet transform].

### Slide 18.2: Classical Shearlets.

- [Show the Classical Shearlets, how to pick the generating function in order to generate a Parseval frame]
- [Show the frequency tiling of the classical shearlets, not uniform at all, is very biased towards the  $\xi_2$ -axis, that will lead to some issues when trying to analyze singularities aligned with the  $\xi_1$ -axis]

### Slide 18.3: Cone Adapted Shearlets.

- [Show the solution, explain the functions, and how to choose them in order to get a Parseval frame (compactly supported)].

### Slide 19.1: Universal Shearlets and $\mathcal{H}^s$ -Shearlets.

- Parabolic scaling is well suited to approximate functions with singularities over parabolic curves; we would like to inpaint the EPIs that have singularities over lines.
- For more flexibility on the "level of anisotropy" of the functions that one would like to approximate, one can use a different scaling parameter in each scale (scaling sequence),



$\{\alpha_j\}_{j \in \mathbb{I}} \subseteq (-\infty, 2)$ , with associated scaling matrices

$$A_j, \alpha_j = \begin{pmatrix} 2^j & 0 & 0 & 2^{\alpha_j/2} \end{pmatrix}$$

- With  $A_j, \alpha_j$  we can define the \textbf{Universal Shearlet System} (Kutyniok, Genzel, 2014), a generalization of the cone-adapted shearlet system for different level of anisotropy [and also permits a unified treatment of wavelets, shearlets, curvelets and ridgelets].

## Slide 19.2: Untitled

- [Present the Schwartz Functions Space].
- [The Meyer and Corona Scaling Functions, explain that it can be interpreted as the wavelet-like function considered in the Classical Shearlets construction.]

## Slide 19.3: Untitled

- [As in the classical shearlets one define a bump-like function].
- [Show the Meyer function, the Corona Shaped domains, and the cones, explain that for the same reason than in the classical shearlets one uses the cone-adapted approach].

## Slide 19.4: Untitled

- [Show the new scaling and shearing Matrices for each scale, cone, and scaling parameter]
- [Show the Adapted cone functions, using the bump functions in each cone]

## Slide 19.5 and 19.6...: Untitled

- [Show the ingredients, coarse scaling functions, Interior functions, boundary shearlets]
- [Explain that to make the Fourier transform continuous (and then cover the fourier domain) the exponent  $2^{\{(2-\alpha_j)/2\}}$  needs to be integer-valued, therefore one define a set of permitted scaling parameters].
- [Show the ingredients of the Universal Shearlet Transform].
- [Show the form of the universal Shearlet transform].
- [Show that it also constitutes a Parseval frame].
- [Mention that for the implementation on julia one just need to select the scaling levels segun the bound of  $|k|$  to perform it].
- [Mention that when selecting  $\alpha_j=1$  then you have Cone-adapted shearlets, by

selecting  $\alpha_j=2$  you obtain the wavelets].

- [Show the 0-shearlets that approximate lines, maximum anisotropic].

## **Slide 20: Shearlet-based inpainting with iterative hard thresholding.**

- [Show the algorithm, explain the synthesis and analysis operator, the thresholding operator, mostly how it works, and that choosing  $\alpha_n=1$  is quite slow, but with  $\alpha$  too high it has instability]
- [Show the results on image inpainting results with disparity of 7pix and 16pix band].

## **Slide 21: Results on line detection and depth map estimation.**

- [Show them the results on line detection using Hough line detector and a depth map]

## **Slide 23: Conclusions and outlook.**

- [On the next pages].

## **Slide 24: Thanks.**