

Master Thesis: Fast Sparse Light Field Reconstruction with Shearlet-Based Inpainting

Héctor Andrade Loarca

TU Berlin, BMS

11th of September, 2017



What is this thesis about?

- ▶ A sparsely sampled light field reconstruction method of a static scene.

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 planes images (EPIs).

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 planes images (EPIs).
- ▶ The EPIs are images with linear structures obtained by tracking different feature points in a sequence of pictures of the scene.

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 planes 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.

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 planes 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.

Involved concepts

- ▶ Early Vision (biology) and Light Field Theory (physics).

Involved concepts

- ▶ Early Vision (biology) and Light Field Theory (physics).
- ▶ Computer Vision for point tracking and line detection (basic concepts of linear algebra).

Involved concepts

- ▶ Early Vision (biology) and Light Field Theory (physics).
- ▶ Computer Vision for point tracking and line detection (basic concepts of linear algebra).
- ▶ Functional Analysis and Computational Harmonic Analysis (sparsifying dictionaries).

Involved concepts

- ▶ Early Vision (biology) and Light Field Theory (physics).
- ▶ Computer Vision for point tracking and line detection (basic concepts of linear algebra).
- ▶ Functional Analysis and Computational Harmonic Analysis (sparsifying dictionaries).
- ▶ Compressed sensing techniques (ℓ^1 optimization algorithms).

Light Field Theory

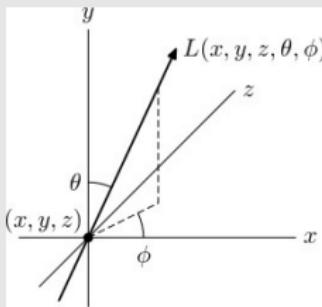
- ▶ In 1846 Michael Faraday proposed for the first time on his lecture "Thoughts on Ray Vibration" that light could be interpreted as a field.

Light Field Theory

- ▶ In 1846 Michael Faraday proposed for the first time on his lecture "Thoughts on Ray Vibration" that light could be interpreted as a field.
- ▶ Propagation of light rays in the 3D space is completely described by a 7D continuous function $L : \mathbb{R}^7 \longrightarrow \mathbb{R}^3$, $L(x, y, z, \theta, \phi, \lambda, \tau)$ called the plenoptic function

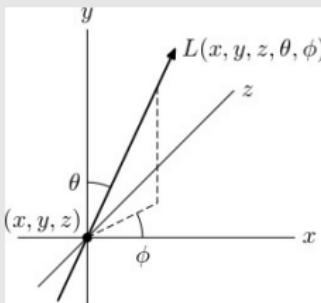
Light Field Theory

- ▶ In 1846 Michael Faraday proposed for the first time on his lecture "Thoughts on Ray Vibration" that light could be interpreted as a field.
- ▶ Propagation of light rays in the 3D space is completely described by a 7D continuous function $L : \mathbb{R}^7 \longrightarrow \mathbb{R}^3$, $L(x, y, z, \theta, \phi, \lambda, \tau)$ called the plenoptic function



Light Field Theory

- ▶ In 1846 Michael Faraday proposed for the first time on his lecture "Thoughts on Ray Vibration" that light could be interpreted as a field.
- ▶ Propagation of light rays in the 3D space is completely described by a 7D continuous function $L : \mathbb{R}^7 \longrightarrow \mathbb{R}^3$, $L(x, y, z, \theta, \phi, \lambda, \tau)$ called the plenoptic function



- ▶ 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.

4D Light Field Representation

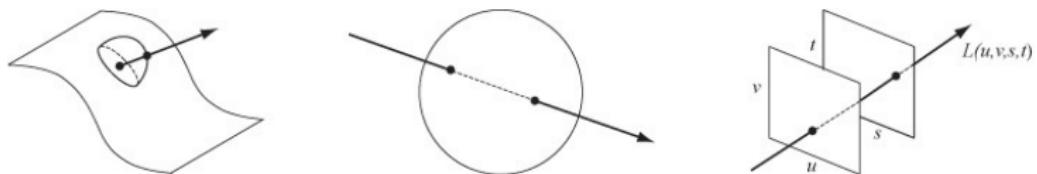


Figure: Three different representation of 4F LF. Left: $L_4(u, v, \phi, \theta)$. Center: $L_4(\phi_1, \theta_1, \phi_2, \theta_2)$. Right: $L_4(u, v, s, t)$.

4D Light Field Representation

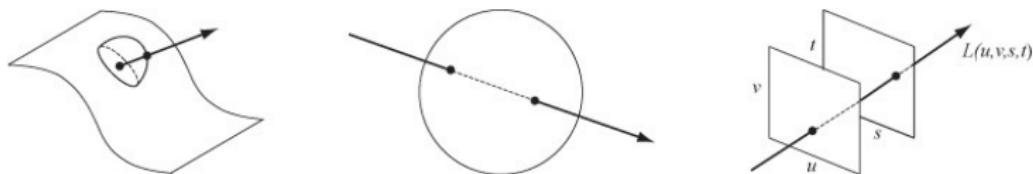


Figure: Three different representation of 4F LF. Left: $L_4(u, v, \phi, \theta)$. Center: $L_4(\phi_1, \theta_1, \phi_2, \theta_2)$. Right: $L_4(u, v, s, t)$.

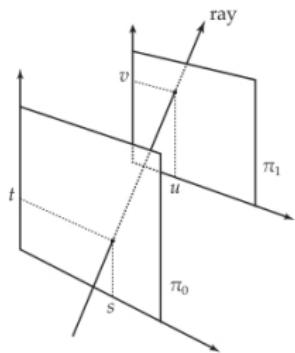


Figure: Used representation: "Two plane parametrization".

Motivation

Compression of High Resolution Light Field (Wetzstein et al., 2013)

Compressive Light Field Photography using Overcomplete Dictionaries and Optimized Projections

Kshitij Marwah¹

Gordon Wetzstein¹

Yosuke Bando^{2,1}

Ramesh Raskar¹

¹MIT Media Lab

²Toshiba Corporation

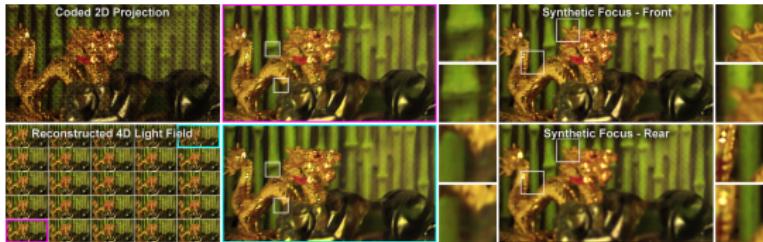


Figure 1: Light field reconstruction from a single coded projection. We explore sparse reconstructions of 4D light fields from optimized 2D projections using light field atoms as the fundamental building blocks of natural light fields. This example shows a coded sensor image captured with our camera prototype (upper left), and the recovered 4D light field (lower left and center). Parallax is successfully recovered (center insets) and allows for post-capture refocus (right). Even complex lighting effects, such as occlusion, specularity, and refraction, can be recovered, being exhibited by the background, dragon, and tiger, respectively.

Abstract

Light field photography has gained a significant research interest in the last two decades; today, commercial light field cameras are widely available. Nevertheless, most existing acquisition approaches either multiplex a low-resolution light field into a single 2D sensor image or require multiple photographs to be taken for acquiring a high-resolution light field. We propose a compressive light field camera architecture that allows for higher-resolution light fields to be recovered than previously possible from a single image. The proposed architecture comprises three key components: light field atoms as a sparse representation of natural light fields, an optical design that allows for capturing optimized 2D light field projections, and robust sparse reconstruction methods to recover a 4D light field from a single coded 2D projection. In addition, we demonstrate a variety of other applications for light field atoms and

1 Introduction

Since the invention of the first cameras, photographers have been striving to capture moments on film. Today, camera technology is on the verge of a new era. With the advent of mobile digital photography, consumers can easily capture, edit, and share moments with friends online. Most recently, light field photography was introduced to the consumer market as a technology facilitating novel user experiences, such as digital refocus, and 3D imaging capabilities, thereby capturing moments in greater detail. The technological foundations of currently available light field cameras, however, are more than a century old and have not fundamentally changed in that time. Most currently available devices trade spatial resolution for the ability to capture different views of a light field, oftentimes reducing the final image resolution by orders of magnitude compared to the raw sensor resolution. Unfortunately, this trend directly coun-

Raytrix (Perwass and Wietzke, 2010)



Lytro (Ng, 2012)



Light Field Reconstruction Using Shearlet Transform

Suren Vagharshakyan, Robert Bregovic and Atanas Gotchev, *Member, IEEE*

Abstract—In this article we develop an image based rendering technique based on light field reconstruction from a limited set of perspective views acquired by cameras. Our approach utilizes sparse representation of epipolar-plane images in a directionally sensitive transform domain, obtained by an adapted discrete shearlet transform. The used iterative thresholding algorithm provides high-quality reconstruction results for relatively big disparities between neighboring views. The generated densely sampled light field of a given 3D scene is thus suitable for all applications which requires light field reconstruction. The proposed algorithm is compared favorably against state of the art depth image based rendering techniques.

Index Terms—Image-based rendering, light field reconstruction, shearlets, frames, view synthesis.

1 INTRODUCTION

SYNTHESIS of intermediate views from a given set of captured views of a 3D visual scene is usually referred to as image-based rendering (IBR) [1]. The scene is typically captured by a limited number of cameras which form a rather coarse set of multiview images. However, denser set of images (i.e. intermediate views) is required in immersive visual applications such as free viewpoint television (FVT) and virtual reality (VR) aimed at creating the perception of continuous parallax.

Modern view synthesis methods are based on two,

needs to sample the LF such that the disparity between neighboring views is less than one pixel [9]. Hereafter, we will refer to such sampling as dense sampling and to the correspondingly sampled LF as densely sampled LF. In order to capture a densely sampled LF, the required distance between neighboring camera positions can be estimated based on the minimal scene depth (z_{min}) and the camera resolution. Furthermore, camera resolution should provide enough samples to properly capture highest spatial texture frequency in the scene [10].

An article about computational result is advertising, not scholarship. The actual scholarship is the full software environment, code and data, that produced the result.

Buckheit and Donoho (1995)

Stereo Vision and Epipolar Geometry

- ▶ **Stereo Vision:** The human brain generates the 3D depth perception of its surroundings by triangulating the points of a scene using the information coming from both eyes.

Stereo Vision and Epipolar Geometry

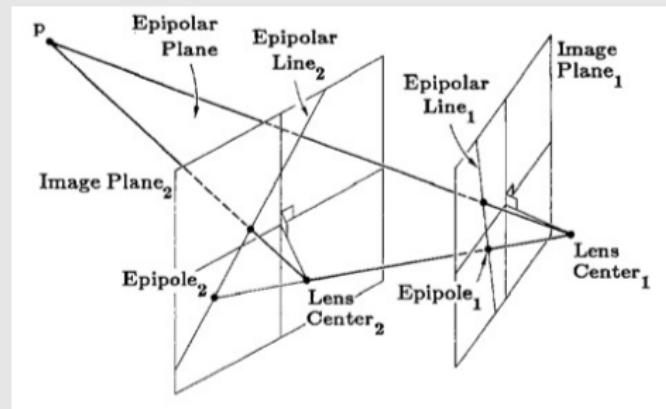
- ▶ **Stereo Vision:** The human brain generates the 3D depth perception of its surroundings by triangulating the points of a scene using the information coming from both eyes.
- ▶ **Epipolar Geometry:** Generalization of Stereo Vision with more than two views, assuming the epipolar constraint.

Stereo Vision and Epipolar Geometry

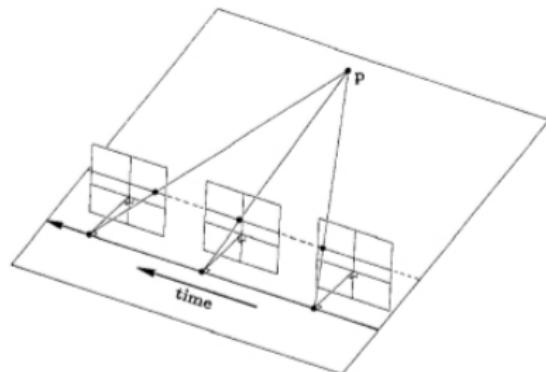
- ▶ **Stereo Vision:** The human brain generates the 3D depth perception of its surroundings by triangulating the points of a scene using the information coming from both eyes.
- ▶ **Epipolar Geometry:** Generalization of Stereo Vision with more than two views, assuming the epipolar constraint.
- ▶ **Epipolar Constraint:** Analysis of object position while assuming the knowledge of the camera motion.

Stereo Vision and Epipolar Geometry

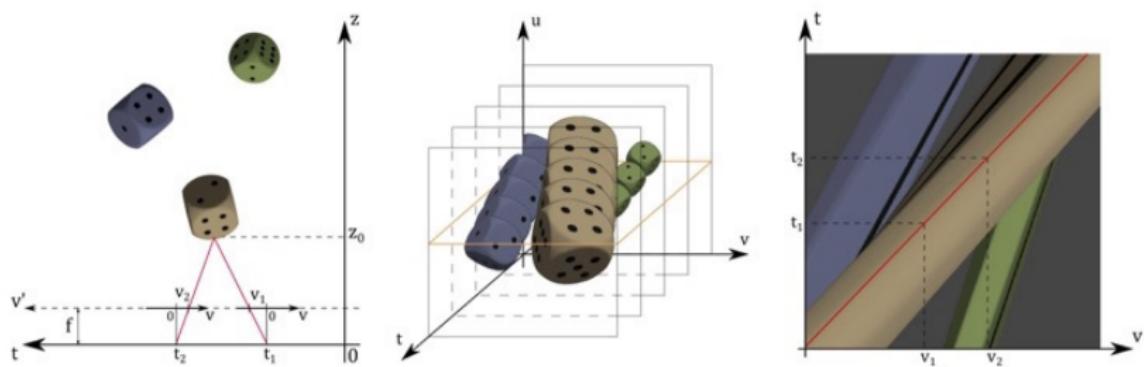
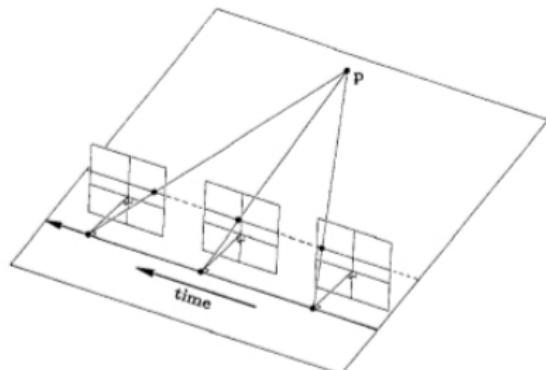
- ▶ **Stereo Vision:** The human brain generates the 3D depth perception of its surroundings by triangulating the points of a scene using the information coming from both eyes.
- ▶ **Epipolar Geometry:** Generalization of Stereo Vision with more than two views, assuming the epipolar constraint.
- ▶ **Epipolar Constraint:** Analysis of object position while assuming the knowledge of the camera motion.



Epipolar Plane Images (EPIs) on Straight Line Trajectories



Epipolar Plane Images (EPIs) on Straight Line Trajectories



Functional Analysis of EPIs

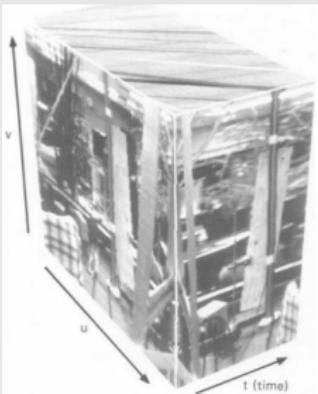
- ▶ **3D Light Field:** Fixing one coordinate in the focal st -plane π_0 of the two-parallel plane approach of 4D Light Field, one reduces the π_0 to a line, and the resulting field $L_3 : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ is called 3D Light Field with radiance $\mathbf{r} = L_3(u, v, s)$.

Functional Analysis of EPIs

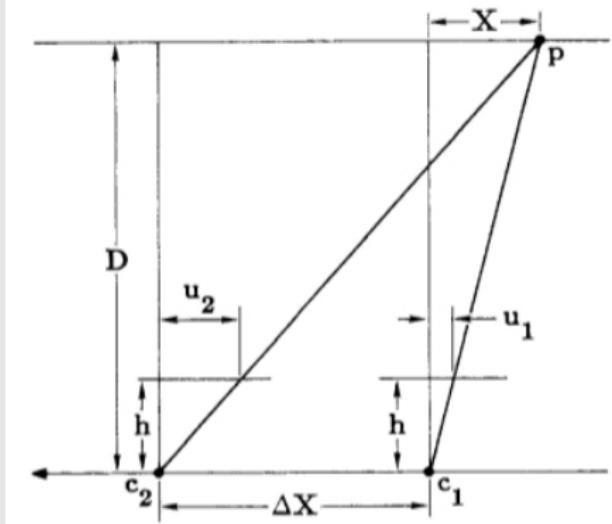
- ▶ **3D Light Field:** Fixing one coordinate in the focal st -plane π_0 of the two-parallel plane approach of 4D Light Field, one reduces the π_0 to a line, and the resulting field $L_3 : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ is called 3D Light Field with radiance $\mathbf{r} = L_3(u, v, s)$.
- ▶ **Epipolar 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 $\mathbf{r} = E_v(u, s)$.

Functional Analysis of EPIs

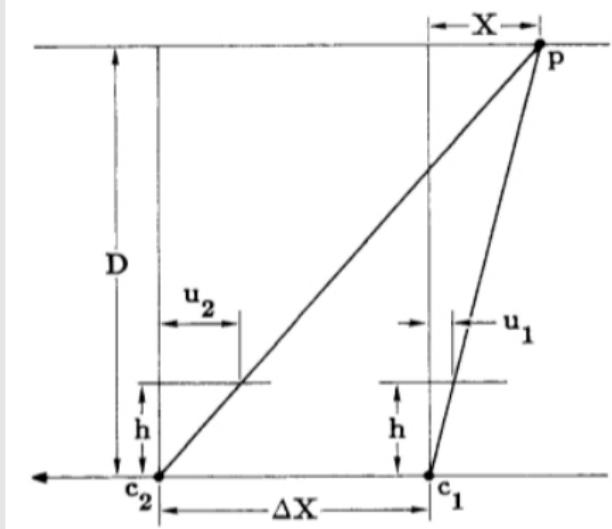
- ▶ **3D Light Field:** Fixing one coordinate in the focal st -plane π_0 of the two-parallel plane approach of 4D Light Field, one reduces the π_0 to a line, and the resulting field $L_3 : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ is called 3D Light Field with radiance $\mathbf{r} = L_3(u, v, s)$.
- ▶ **Epipolar 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 $\mathbf{r} = E_v(u, s)$.



Depth map Estimation with EPIs

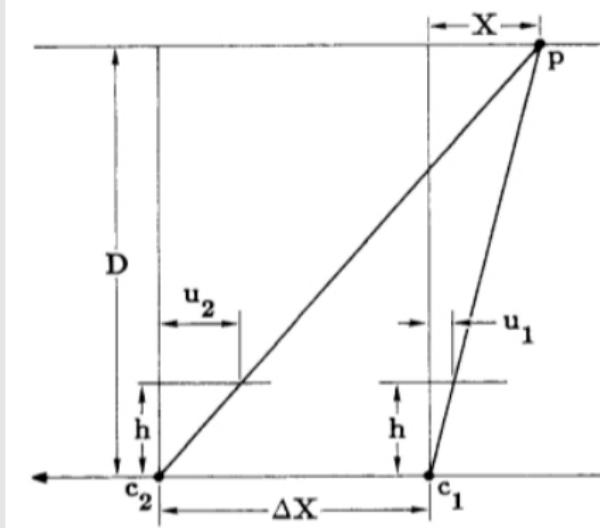


Depth map Estimation with EPIs



- ▶ **Point-depth formula:** $D = h \frac{\Delta X}{\Delta u} = h \frac{\Delta X}{u_1 - u_2}$.

Depth map Estimation with EPIs



- ▶ **Point-depth formula:** $D = h \frac{\Delta X}{\Delta u} = h \frac{\Delta X}{u_1 - u_2}$.
- ▶ **Sparse sampling rate (Nyquist criterion):** $\Delta X \leq \frac{D_{min}}{h} \Delta u$.

Physical Acquisition Setup

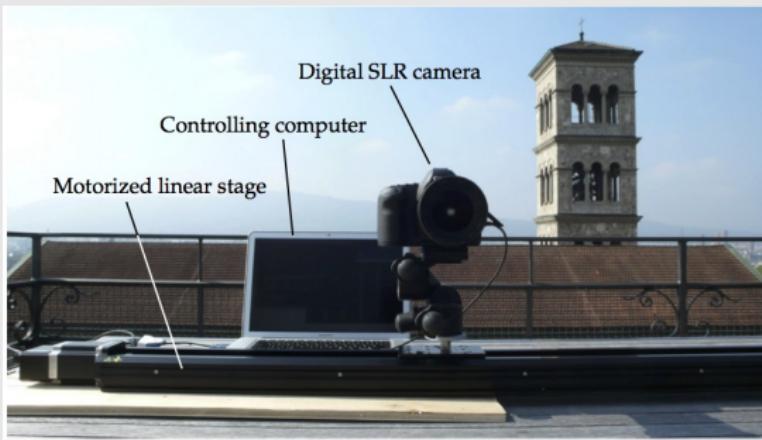
- ▶ **Data set:** Sequence of 101 rectified pictures of a scene generated by Professor Markus Gross' group in the Disney Research Center at Zurich.

Physical Acquisition Setup

- ▶ **Data set:** Sequence of 101 rectified pictures of a scene generated by Professor Markus Gross' group in the Disney Research Center at Zurich.
- ▶ **Technical details of physical setup:** Canon EOS 4D Mark II DSLR camera, Canon EF 50 mm f/1.4 USM lens and a Zaber T-LST1500D motorized linear stage to drive the camera to the shooting positions with 10 mm of distance between each other.

Physical Acquisition Setup

- ▶ **Data set:** Sequence of 101 rectified pictures of a scene generated by Professor Markus Gross' group in the Disney Research Center at Zurich.
- ▶ **Technical details of physical setup:** Canon EOS 4D Mark II DSLR camera, Canon EF 50 mm f/1.4 USM lens and a Zaber T-LST1500D motorized linear stage to drive the camera to the shooting positions with 10 mm of distance between each other.



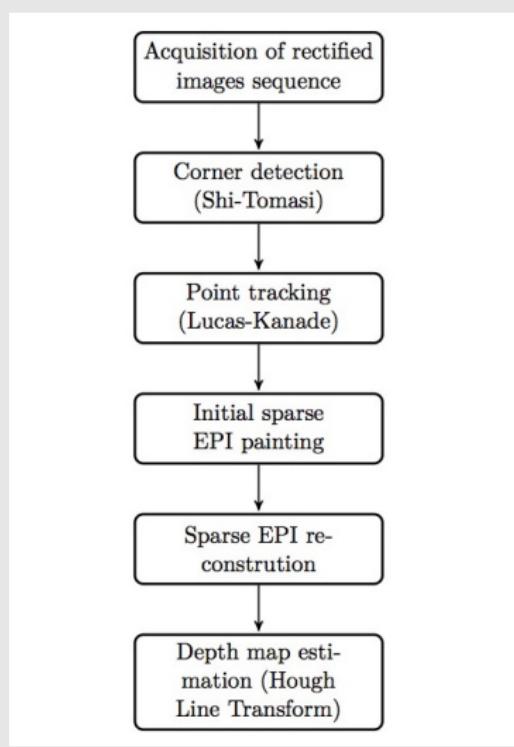
Used Data Set: Church



Used Data Set: Church



Followed Pipeline



Point Tracking Results



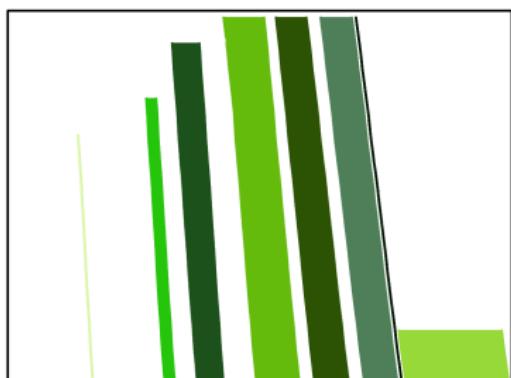
Point Tracking Results



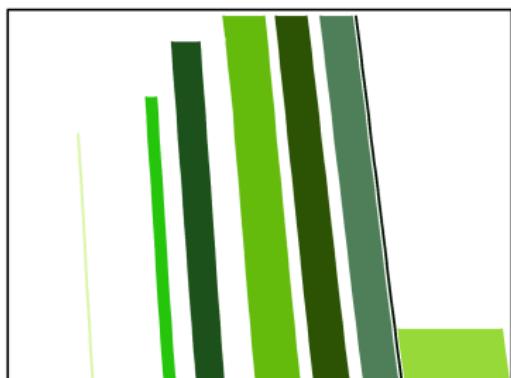
Particular EPI Example



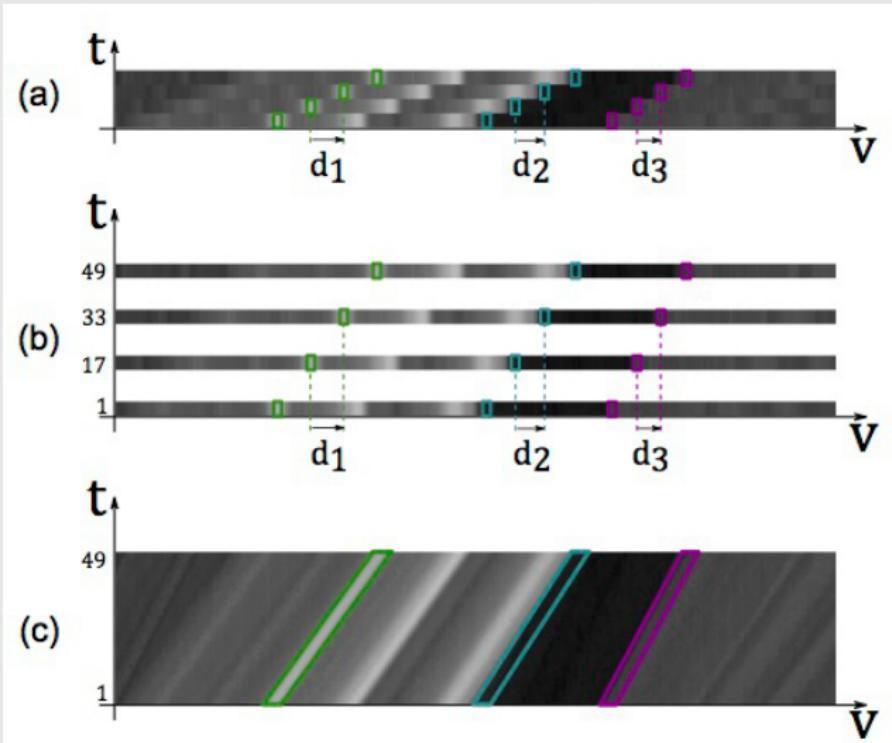
Particular EPI Example



Particular EPI Example



Reconstruction Method with Inpainting



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 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 frame $(\phi_i)_{i \in I}$ span \mathcal{H} .

Analysis, synthesis and frame operator

$T : \mathcal{H} \rightarrow \ell_2(I)$ given by $f \mapsto (\langle f, \phi_i \rangle)_{i \in I}$ is called **the analysis operator**, and $T^* : \ell_2(I) \rightarrow \mathcal{H}$, given by $(c_i)_{i \in I} \mapsto \sum_{i \in I} c_i \phi_i$ is called **the synthesis operator**.

$S = T^*T : \mathcal{H} \rightarrow \mathcal{H}$, given by $f \mapsto \sum_{i \in I} \langle f, \phi \rangle \phi$ is called **the frame operator** and is an invertible operator.

Analysis, synthesis and frame operator

$T : \mathcal{H} \rightarrow \ell_2(I)$ given by $f \mapsto (\langle f, \phi_i \rangle)_{i \in I}$ is called **the analysis operator**, and $T^* : \ell_2(I) \rightarrow \mathcal{H}$, given by $(c_i)_{i \in I} \mapsto \sum_{i \in I} c_i \phi_i$ is called **the synthesis operator**.

$S = T^*T : \mathcal{H} \rightarrow \mathcal{H}$, given by $f \mapsto \sum_{i \in I} \langle f, \phi_i \rangle \phi_i$ is called **the frame operator** and is an invertible operator.

Reconstruction and Decomposition Formula

If $(\phi_i)_{i \in I} \subseteq \mathcal{H}$ be a frame for \mathcal{H} , and S its frame operator, then

$$f = \sum_{i \in I} \langle f, \phi_i \rangle S^{-1} \phi_i, \quad \forall f \in \mathcal{H} \text{ (Reconstruction)}$$

$$f = \sum_{i \in I} \langle f, S^{-1} \phi_i \rangle \phi_i, \quad \forall f \in \mathcal{H} \text{ (Decomposition)}$$

Abstract Inpainting Framework

- ▶ Let \mathcal{H} a separable Hilbert space and $x^0 \in \mathcal{H} = \mathcal{H}_K \oplus \mathcal{H}_M = P_K \mathcal{H} \oplus P_M \mathcal{H}$. Then, given a corrupt signal $P_K x^0$, we want to recover the missing part $P_M x^0$.

Abstract Inpainting Framework

- ▶ Let \mathcal{H} a separable Hilbert space and $x^0 \in \mathcal{H} = \mathcal{H}_K \oplus \mathcal{H}_M = P_K \mathcal{H} \oplus P_M \mathcal{H}$. 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 $H_M = L^2(\mathcal{M})$ for some measurable set $\mathcal{M} \subset \mathbb{R}^2$.

Abstract Inpainting Framework

- ▶ Let \mathcal{H} a separable Hilbert space and $x^0 \in \mathcal{H} = \mathcal{H}_K \oplus \mathcal{H}_M = P_K \mathcal{H} \oplus P_M \mathcal{H}$. 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 $H_M = L^2(\mathcal{M})$ for some measurable set $\mathcal{M} \subset \mathbb{R}^2$.
- ▶ We will assume that x^0 can be efficiently represented by some Parseval frame $\Phi = (\phi_i)_{i \in I}$ for \mathcal{H} , this is translated as asking for the solution of the ℓ^0 -minimization problem

$$\min_{c \in \ell^2(I)} \|c\|_{\ell^0(I)} \quad \text{subject to} \quad x^0 = T_\Phi^* c = \sum_{i \in I} c_i \phi_i$$

Analysis approach

Algorithm 1: Inpainting via ℓ^1 -minimization

Input : Corrupted signal $P_Kx^0 \in \mathcal{H}_K$, Parseval frame $\Phi = (\phi_i)_{i \in I}$ for \mathcal{H}
Compute:

$$x^* = \underset{x \in \mathcal{H}}{\operatorname{argmin}} \|T_\Phi x\|_{\ell^1(I)} \quad \text{subject to} \quad P_Kx^0 = P_Kx$$

Output : recovered signal $x^* \in \mathcal{H}$

Cone Adapted Shearlets

Universal Shearlets and α -Shearlets

Shearlet-based inpainting with iterative hard thresholding

Line detection and depth map estimation

Conclusions and Outlook

Thanks!

Questions?

