LightFields.jl: Fast 3D image reconstruction for VR applications

Héctor Andrade Loarca

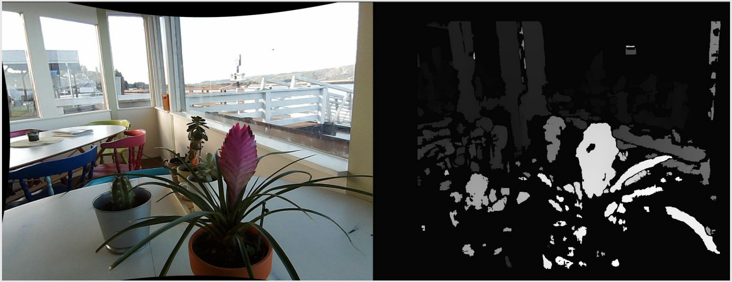
Technical University of Berlin, BMS

8^{th} of August, 2018

JuliaCon 2018

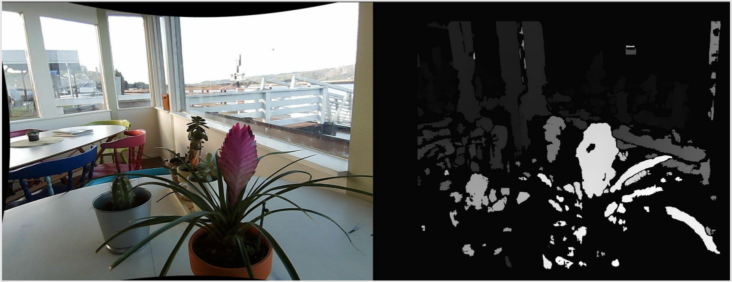
Main goal

- Present a novel technique to reconstruct the **depth map** of a scene from a limited number of views. This can be applied in view synthesis and rendering for free viewpoint VR.



Main goal

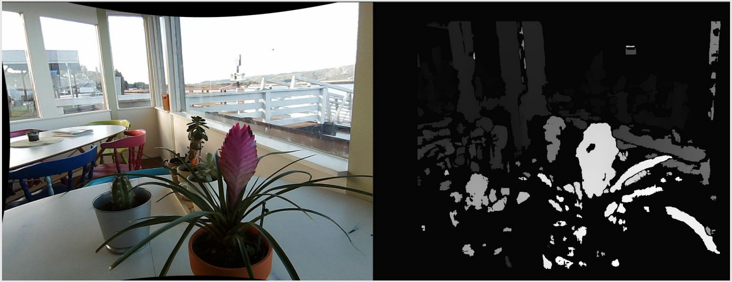
- Present a novel technique to reconstruct the **depth map** of a scene from a limited number of views. This can be applied in view synthesis and rendering for free viewpoint VR.



- Explain the main building blocks of the technique: Light Field and Shearlets.

Main goal

- Present a novel technique to reconstruct the **depth map** of a scene from a limited number of views. This can be applied in view synthesis and rendering for free viewpoint VR.



- Explain the main building blocks of the technique: Light Field and Shearlets.
- Show a free hardware/software implementation using julia, python and Raspberry Pi.

What is a Light Field?

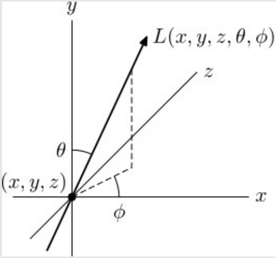
- Light can be interpreted as a field, i.e. assignment of a vector to each point in the space (M. Faraday, 1846).

What is a Light Field?

- Light can be interpreted as a field, i.e. assignment of a vector to each point in the space (M. Faraday, 1846).
- 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**

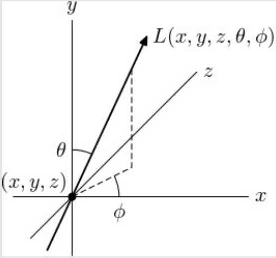
What is a Light Field?

- Light can be interpreted as a field, i.e. assignment of a vector to each point in the space (M. Faraday, 1846).
- 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**



What is a Light Field?

- Light can be interpreted as a field, i.e. assignment of a vector to each point in the space (M. Faraday, 1846).
- 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**



- L 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

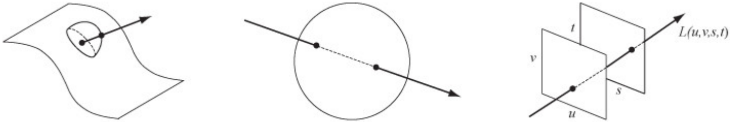


Figure: Three different representation of 4F LF. Left: $L_4(u, v, \phi, \theta)$. Center: $L_4(\phi_1, \theta_1, \phi_2, \theta_2)$. Right: $L_4(u, v, s, t)$.

4D Light Field Representation

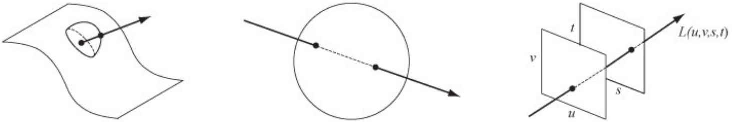


Figure: Three different representation of 4F LF. Left: $L_4(u, v, \phi, \theta)$. Center: $L_4(\phi_1, \theta_1, \phi_2, \theta_2)$. Right: $L_4(u, v, s, t)$.

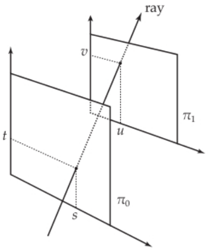


Figure: Used representation: "Two plane parametrization".

From LF to 3D

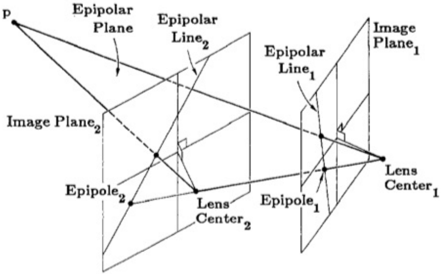
- **Stereo Vision:** The human brain generates the 3D depth perception of its sorroundings by triangulating the points of a scene using the information coming from both eyes.

From LF to 3D

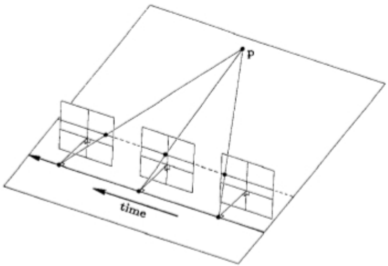
- **Stereo Vision:** The human brain generates the 3D depth perception of its sorroundings by triangulating the points of a scene using the information coming from both eyes.
- **Epipolar Geometry:** Generalization of Stereo Vision with more than two views, assuming the epipolar constraint.

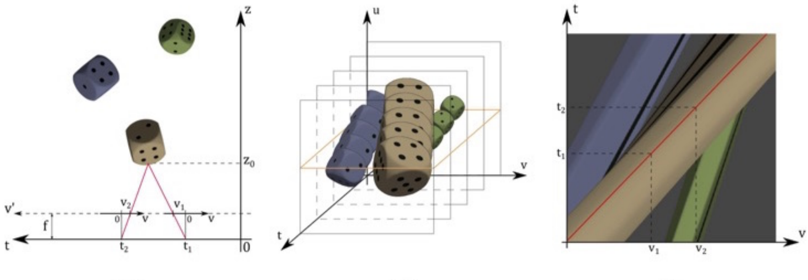
From LF to 3D

- **Stereo Vision:** The human brain generates the 3D depth perception of its sorroundings 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.

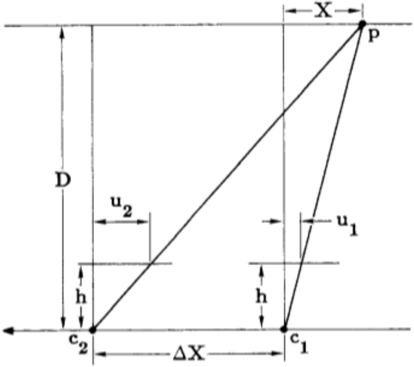


Epipolar Plane Images (EPIs) on Straight Line Trajectories



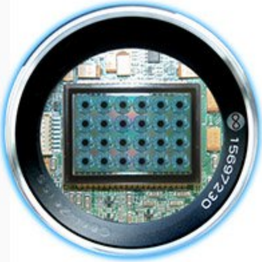


Depth map estimation with EPIs

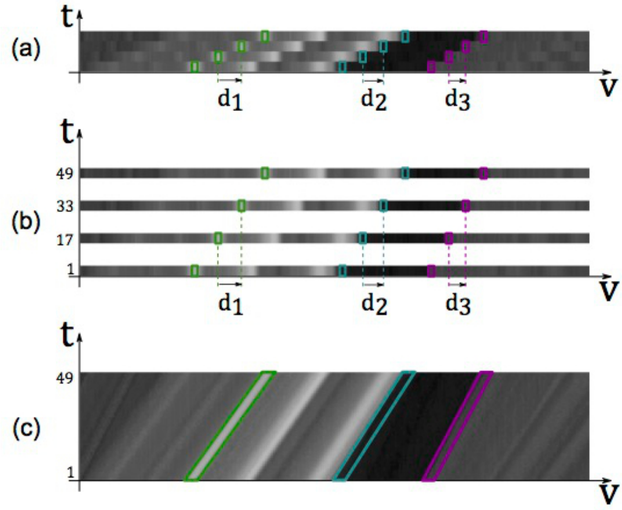


- **Point-depth formula:** $D = h\frac{\Delta X}{\Delta u} = h\frac{\Delta X}{u1-u2}$.
- **Sampling rate (Nyquist criterion):** $\Delta X \leq \frac{D_{min}}{h}\Delta u$.

Commercial LF (Epipolar) camera



Our approach: Sub-Nyquist reconstruction via inpainting



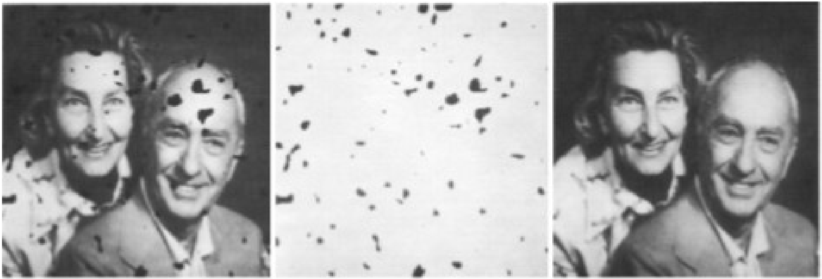
(General) Image inpainting

Mathematical formulation

Recover an image $f \in X$ from known data:

$$g = P_K(f)$$

where P_K is and orthogonal projection onto the known subspace $X_K \triangleleft X$.



How to inpaint?

Frame

A frame for a Hilbert space X is a collection $\Psi = \{\psi_i\}_{i \in \mathcal{I}} \subset X$ satisfying

$$A||f||_2 \leq ||\{\langle f, \psi_i \rangle\}_{i \in \mathcal{I}}||_{\ell^2(\mathcal{I})} \leq B||f||_2 \quad \forall f \in X$$

for some $0 < A \leq B < \infty$.

Sparse Regularization/CS approach (Genzel, Kutyniok, 2014):

" If a signal (image) is sparse within a frame Ψ, it can be recovered from highly underdetermined, non-adaptive linear measurements by ℓ^1-regularization, i.e.

$$\min_{\tilde{f} \in X} ||\{\langle \tilde{f}, \psi_i \rangle\}_{i \in \mathcal{I}}||_{\ell^1(\mathcal{I})} \quad \text{s.t. } P_K(\tilde{f}) = g = P_K(f) \quad "$$

Frames for images and optimal sparsity

- Gabor frames (Gabor, 1946).
- Wavelet frames (Morlet et al., 1984).
- Curvelet frames (Candès et al., 1999).
- Shearlet frames (Kutyniok et al., 2005).

Frames for images and optimal sparsity

- Gabor frames (Gabor, 1946).
- Wavelet frames (Morlet et al., 1984).
- Curvelet frames (Candès et al., 1999).
- Shearlet frames (Kutyniok et al., 2005).

Best N-term approx. error (Donoho, 2001)

Let $\{\psi_\lambda\}_{\lambda\in\Lambda} \subset L^2(\mathbb{R}^2)$ a frame. The optimal best N-Term approximation error for any $f \in \mathcal{E}^2(\mathbb{R}^2)$ is

$$\sigma_N(f, \{\psi_\lambda\}_{\lambda\in\Lambda}) = O(N^{-1})$$

Frames for images and optimal sparsity

- Gabor frames (Gabor, 1946).
- Wavelet frames (Morlet et al., 1984).
- Curvelet frames (Candès et al., 1999).
- Shearlet frames (Kutyniok et al., 2005).

Best N-term approx. error (Donoho, 2001)

Let $\{\psi_\lambda\}_{\lambda \in \Lambda} \subset L^2(\mathbb{R}^2)$ a frame. The optimal best N-Term approximation error for any $f \in \mathcal{E}^2(\mathbb{R}^2)$ is

$$\sigma_N(f, \{\psi_\lambda\}_{\lambda \in \Lambda}) = O(N^{-1})$$

Error of 2D-wavelets

$$\sigma_N(f, \{\psi_\lambda\}_\Lambda) \sim N^{-1/2}$$

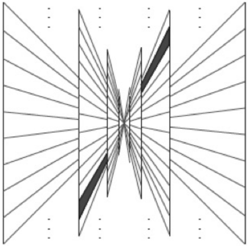
Shearlet Transform (Kutyniok, Guo, Labate, 2005)

Classical Shearlet Transform

$$\langle f, \psi_{j,k,m} \rangle = \int_{\mathbb{R}^2} f(x)\overline{\psi_{j,k,m}(x)}dx$$

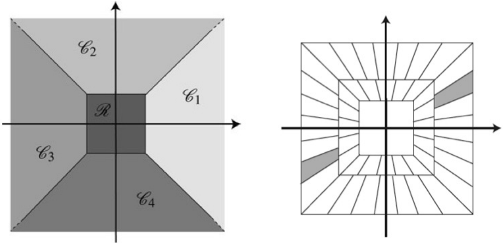
where

$$\mathcal{SH}(\psi) = \{\psi_{j,k,m}(x) = 2^{3j/4}\psi(S_k A_j x - m) : (j,k) \in \mathbb{Z}^2, m \in \mathbb{Z}^2\}$$



Modification: Cone-adapted Shearlet transform

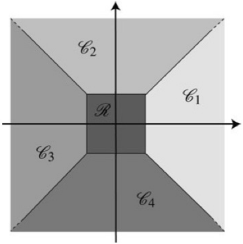
$$\mathcal{SH}(\phi, \psi, \tilde{\psi}, c) := \mathcal{P}_{\mathcal{R}}\Phi(\phi, c1) \cup \mathcal{P}_{\mathcal{C}_1}\Psi(\psi, c) \cup \mathcal{P}_{\mathcal{C}_2}\tilde{\Psi}(\tilde{\psi}, c)$$



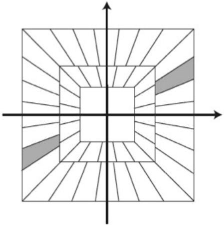
(a) (b)

Modification: Cone-adapted Shearlet transform

$$\mathcal{SH}(\phi,\psi,\tilde{\psi},c) := \mathcal{P}_{\mathcal{R}}\Phi(\phi,c1) \cup \mathcal{P}_{\mathcal{C}_1}\Psi(\psi,c) \cup \mathcal{P}_{\mathcal{C}_2}\tilde{\Psi}(\tilde{\psi},c)$$



(a)



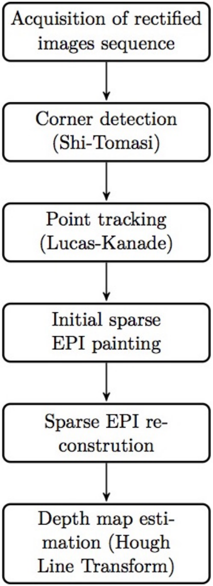
(b)

Cone shearlets sparsity (Band limited case: Lim, Labate; 2006), (Compactly supported case: Kutyniok, Lim, 2011)

Best N-term approximation error

$$\sigma_N(f,\{\psi_{j,k,m}\}_{j,k,m}) \sim N^{-1}(\log(N))^{3/2}$$

Followed Pipeline



Physical Acquisition Setup

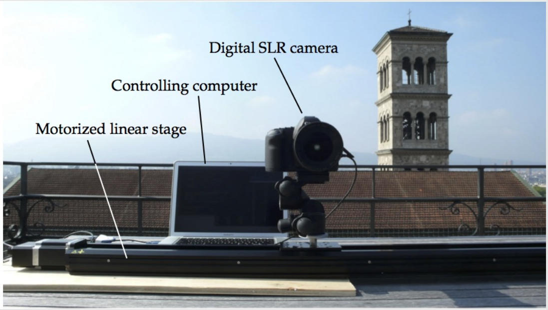
- **Data set:** Sequence of 101 rectified pictures of a scene generated by Professor Markus Gross' group in the Disney Research Center at Zürich.

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 Zürich.
- **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 Zürich.
- **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

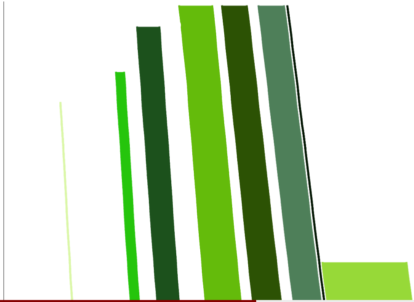
Used Data Set: Church

Point Tracking Results

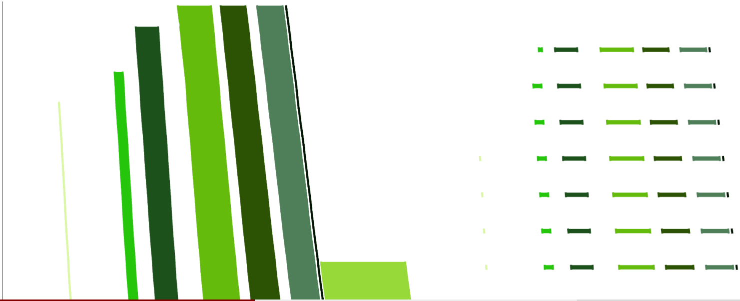
Point Tracking Results

Example of EPI

Example of EPI

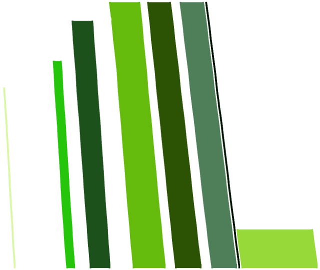


Example of EPI

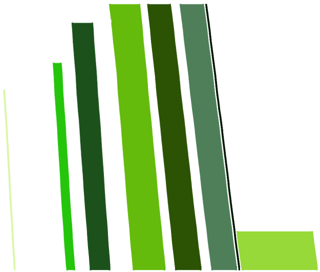


Results on EPIs inpainting

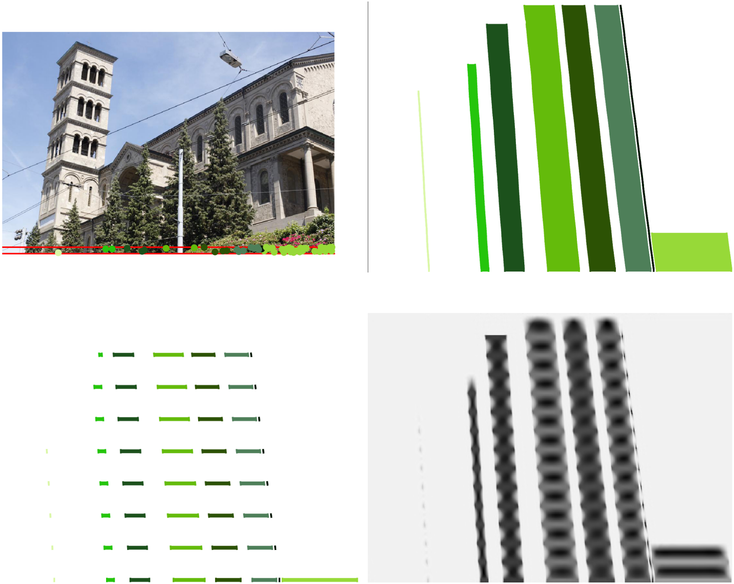
Results on EPIs inpainting



Results on EPIs inpainting

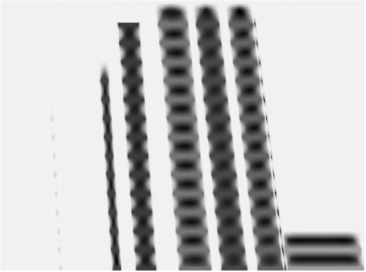


Results on EPIs inpainting

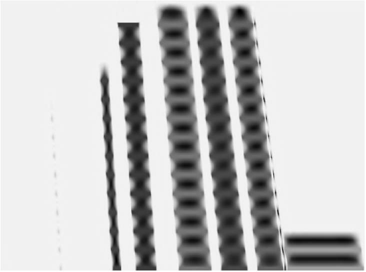


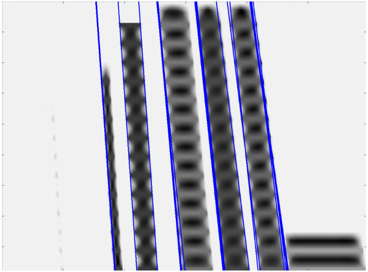
Results on line detection and depth map estimation

Results on line detection and depth map estimation

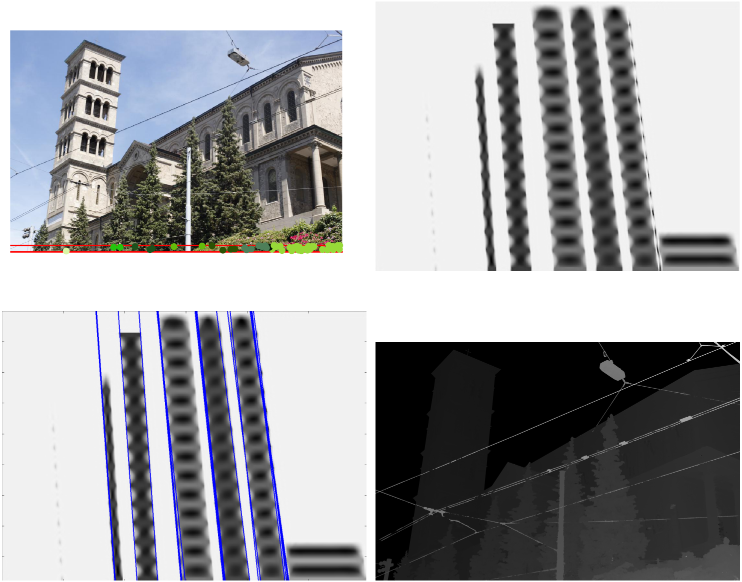


Results on line detection and depth map estimation

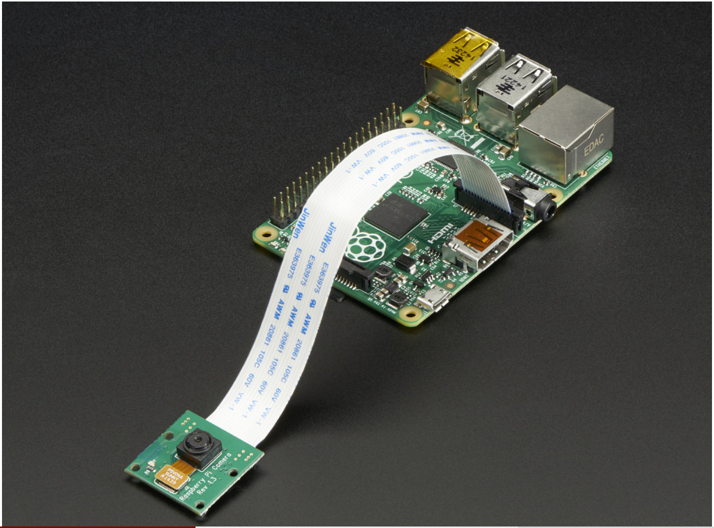




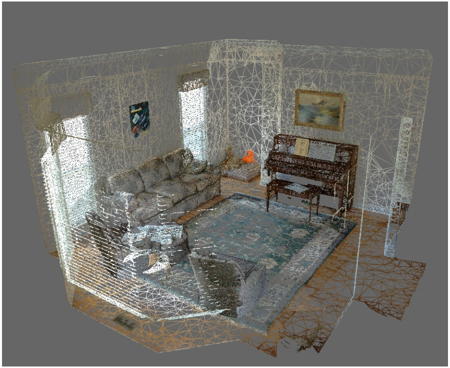
Results on line detection and depth map estimation



Raspberry π + Camera module v2



Future work



Thanks!

Questions?