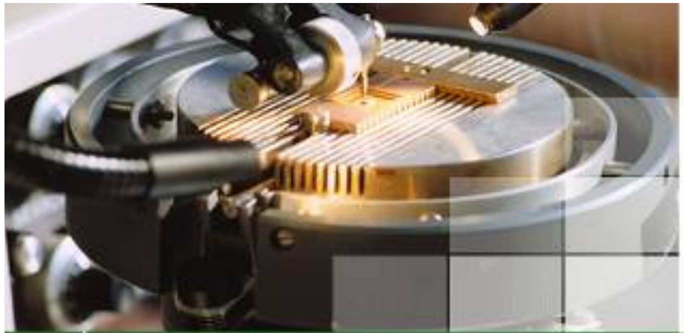


POLITECNICO DI MILANO

Dipartimento di
Elettronica e Informazione

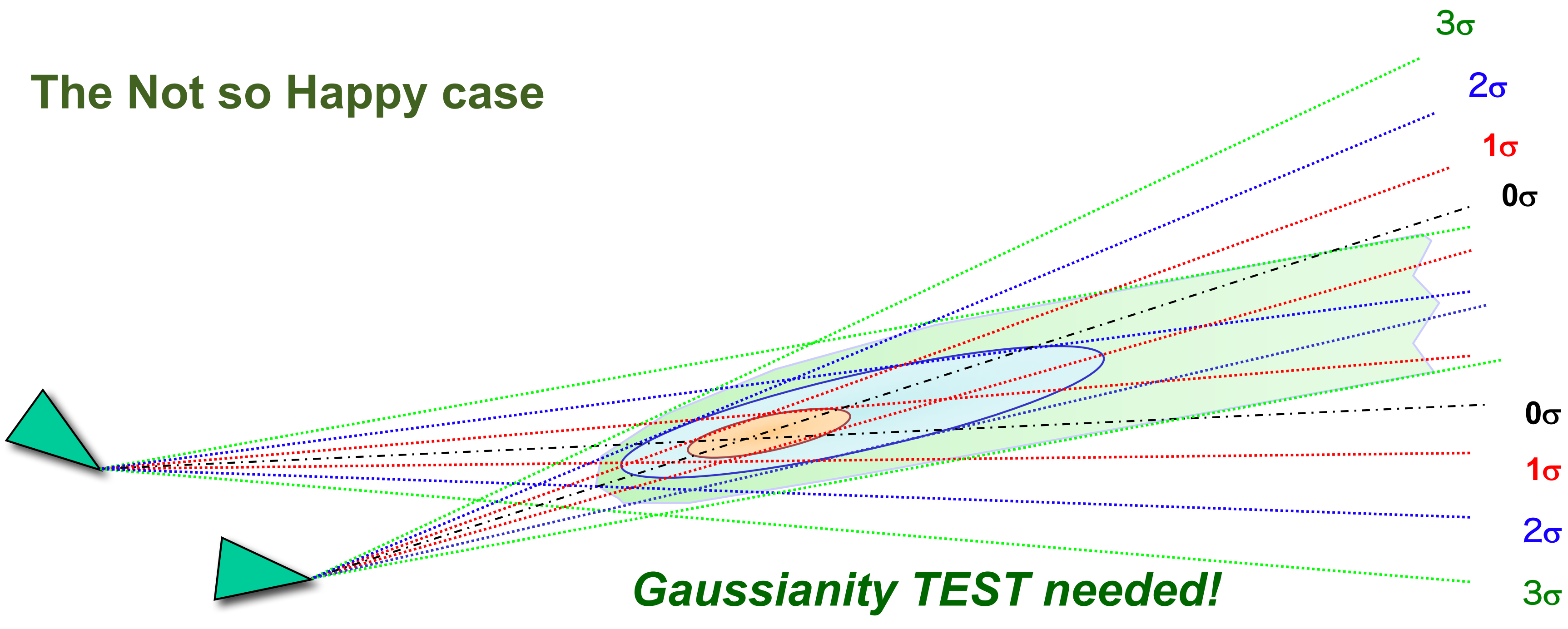
Monocular SLAM: Unified Inverse Depth

3D Structure from Visual Motion

M. Matteucci, Vincenzo Caglioti, Marco Marcon,
*Davide Migliore, **Domenico G. Sorrenti***

sorrenti@disco.unimib.it
Dip. Informatica, Sistemistica e Comunicazione, Università di Milano - Bicocca
Informatica e Robotica per l'Automazione Lab

The problem: Landmark Initialization

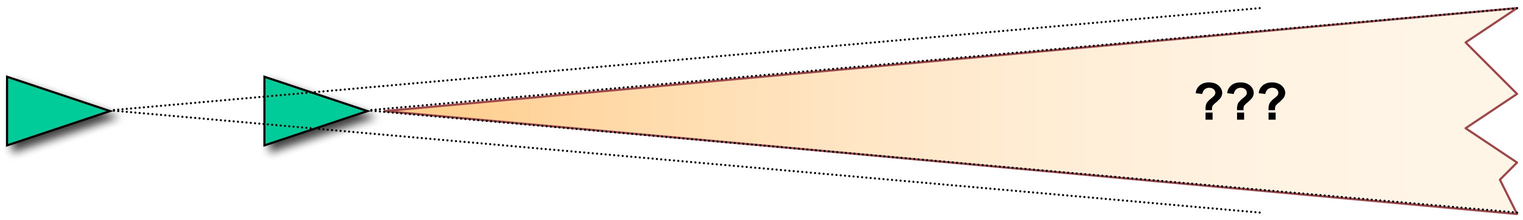


Computation gets risky: A Gaussian does not suit the true PDF:

- **The mean is no longer close to the nominal solution**
- **The covariance is not representative**

The problem: Landmark Initialization

The Unhappy case



There's simply nothing to compute!

Unified Inverse Depth - preliminary considerations

1. **the uncertainty distribution of the Cartesian coordinates of many observed features is not Gaussian (low parallax) => troubles with data association;**

2. **also features with low parallax are important (of course not for the translation, but for the observer orientation);**

3. **if we delay the exploitation of such features we deprive the filter of relevant / critical info;**

4. **nearby features (high-parallax) do not represent a problem, as they are much more Gaussian-distributed than farther ones;**

5. **we look for a parameterization of features that could allow an un-delayed Gaussian initialization of any feature: both far (orientation), and nearby (translation) => this would allow usage of well-established Kalman filtering;**

6. **inverse distances are approximately following a normal distribution**

Unified Inverse Depth - main literature reference

- **Inverse Depth Parametrization for Monocular SLAM**
  Civera, J.; Davison, A.J.; Montiel, J.;
  Robotics, IEEE Transactions on
  Volume 24, Issue 5, Oct. 2008 Page(s):932 - 945

Unified Inverse Depth - “in words” definition

An explicit parameterization of the inverse depth of a feature along a semi-infinite ray from the position from which it was first viewed...

allows a Gaussian distribution to cover uncertainty in depth which spans a depth range from nearby to infinity,

and permits seamless crossing over to finite depth estimates of features which have been apparently infinite for long periods of time.

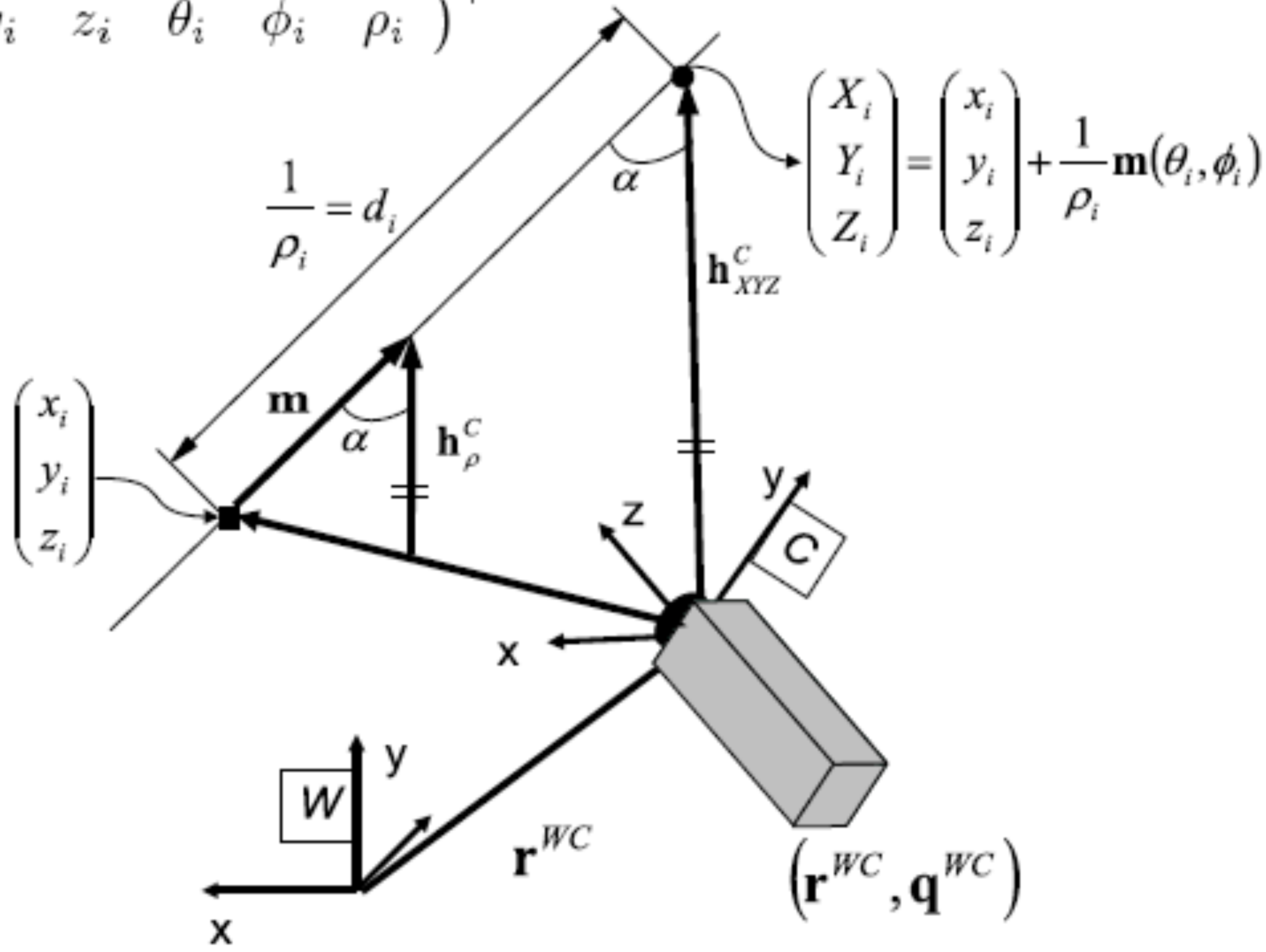
The unified representation means that our algorithm requires no special initialization process for features. They are simply tracked right from the start, immediately contribute to improved camera estimates and have their correlations with all other features in the map correctly modeled.

The projective nature of a camera means that the image measurement process is nearly linear in this inverse depth coordinate.

The $\mathbf{y}_i$ vector encodes the *ray from the first camera position from which the feature was observed* by x_i, y_i, z_i, the camera optical center, and θ_i, ϕ_i azimuth and elevation (coded in the world frame) defining unit directional vector $\mathbf{m}(\theta_i, \phi_i)$. The point's depth along the ray d_i is encoded by its inverse $\rho_i = 1/d_i$.

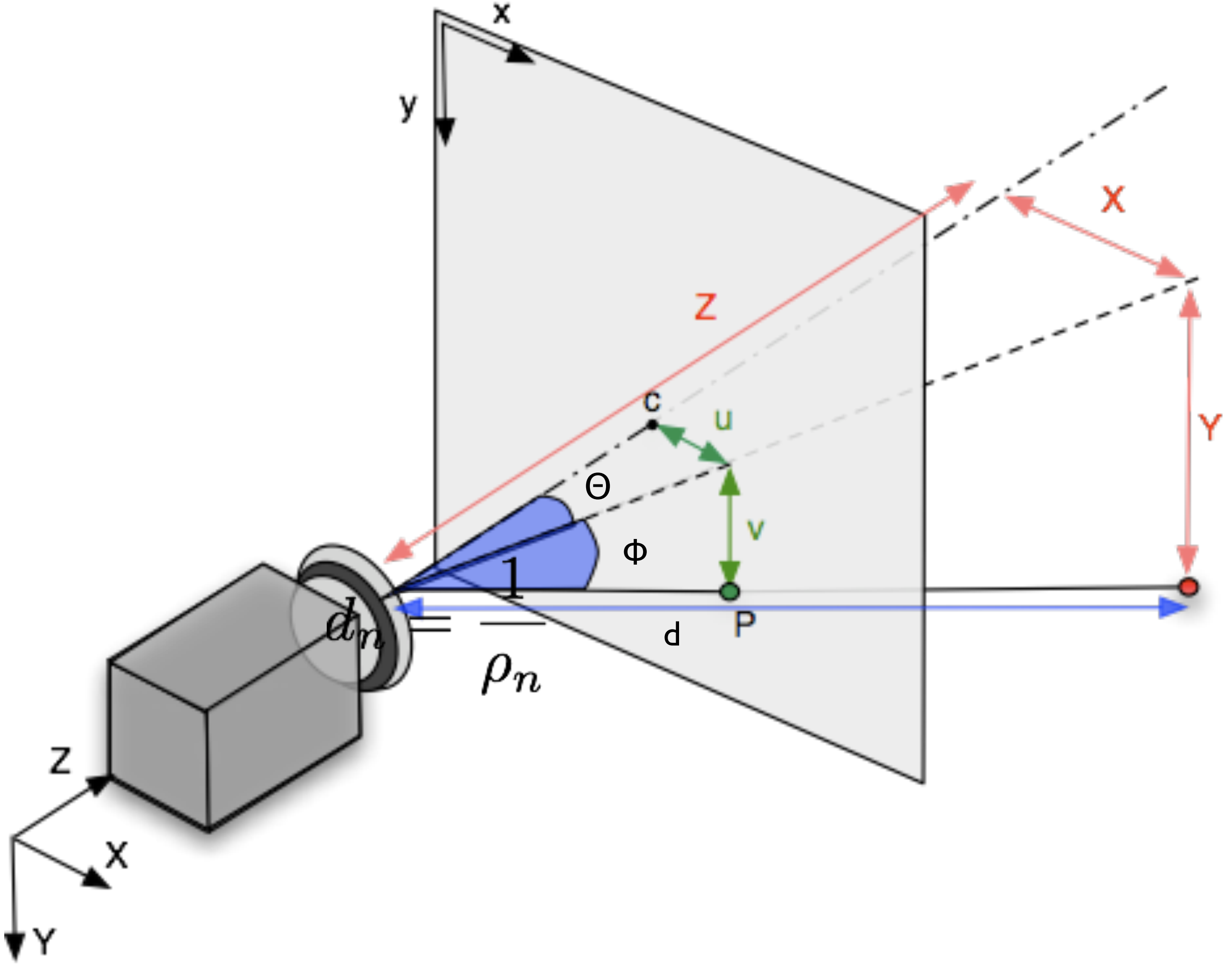
Unified Inverse Depth - the parameterization

$$\mathbf{y}_i = \begin{pmatrix} x_i & y_i & z_i & \theta_i & \phi_i & \rho_i \end{pmatrix}^\top$$



$$\mathbf{m} = (\cos\phi_i \sin\theta_i, -\sin\phi_i, \cos\phi_i \cos\theta_i)^\top$$

Unified Inverse Depth - the parameterization



Unified Inverse Depth - state vector

$$\mathbf{f}_v = \begin{pmatrix} \mathbf{r}^{WC}_{k+1} \\ \mathbf{q}^{WC}_{k+1} \\ \mathbf{v}^{W}_{k+1} \\ \omega^{C}_{k+1} \end{pmatrix} = \begin{pmatrix} \mathbf{r}^{WC}_{k} + \left(\mathbf{v}^{W}_{k} + \mathbf{V}^{W}_{k}\right)\Delta t \\ \mathbf{q}^{WC}_{k} \times \mathbf{q}\left(\left(\omega^{C}_{k} + \Omega^{C}\right)\Delta t\right) \\ \mathbf{v}^{W}_{k} + \mathbf{V}^{W} \\ \omega^{C}_{k} + \Omega^{C} \end{pmatrix}$$

state vector: camera pose sub-part

As in standard EKF SLAM, we use a single joint state vector containing camera pose and feature estimates, with the assumption that the camera moves with respect to a static scene. The whole state vector $\mathbf{x}$ is composed of the camera and all the map features:

$$\mathbf{x} = \left(\mathbf{x}_v^\top, \mathbf{y}_1^\top, \mathbf{y}_2^\top, \ldots \mathbf{y}_n^\top\right)^\top . \quad (6)$$

Unified Inverse Depth - the measurement

For points in XYZ:

$$\mathbf{h}^C = \mathbf{h}^C_{\mathrm{XYZ}} = \mathbf{R}^{CW}\left(\begin{pmatrix} X_i \\ Y_i \\ Z_i \end{pmatrix} - \mathbf{r}^{WC}\right) . \tag{7}$$

For points in inverse depth:

$$\mathbf{h}^C = \mathbf{h}^C_{\rho} = \mathbf{R}^{CW}\left(\rho_i\left(\begin{pmatrix} x_i \\ y_i \\ z_i \end{pmatrix} - \mathbf{r}^{WC}\right) + \mathbf{m}\left(\theta_i, \phi_i\right)\right) , \tag{8}$$

$$\mathbf{h} = \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} u_0 - \frac{f}{d_x}\frac{h_x}{h_z} \\ v_0 - \frac{f}{d_y}\frac{h_y}{h_z} \end{pmatrix}$$

Unified Inverse Depth - the measurement

$$\begin{pmatrix} u_u \\ v_u \end{pmatrix} = \mathbf{h}_u \begin{pmatrix} u_d \\ v_d \end{pmatrix} = \begin{pmatrix} u_0 + (u_d - u_0)\left(1 + \kappa_1 r_d^2 + \kappa_2 r_d^4\right) \\ v_0 + (v_d - v_0)\left(1 + \kappa_1 r_d^2 + \kappa_2 r_d^4\right) \end{pmatrix}$$

$$r_d = \sqrt{\left(d_x (u_d - u_0)\right)^2 + \left(d_y (v_d - v_0)\right)^2}$$

$$\begin{pmatrix} u_d \\ v_d \end{pmatrix} = \mathbf{h}_d \begin{pmatrix} u_u \\ v_u \end{pmatrix} = \begin{pmatrix} u_0 + \frac{(u_u - u_0)}{\left(1 + \kappa_1 r_d^2 + \kappa_2 r_d^4\right)} \\ v_0 + \frac{(v_u - v_0)}{\left(1 + \kappa_1 r_d^2 + \kappa_2 r_d^4\right)} \end{pmatrix}$$

$$r_u = r_d \left(1 + \kappa_1 r_d^2 + \kappa_2 r_d^4\right)$$

$$r_u = \sqrt{\left(d_x (u_u - u_0)\right)^2 + \left(d_y (v_u - v_0)\right)^2}$$

Unified Inverse Depth - the measurement

Undistortion Jacobian $\partial \mathbf{h}_u / \partial (u_d, v_d)$ has the following analytical expression:

$$\begin{pmatrix} \begin{array}{c|c} \begin{array}{c} \left(1 + \kappa_1 r_d^2 + \kappa_2 r_d^4\right) + \\ 2\left(\left(u_d - u_0\right) d_x\right)^2 \times \\ \left(\kappa_i + 2\kappa_2 r_d^2\right) \end{array} & \begin{array}{c} 2d_y^2 \left(u_d - u_0\right)\left(v_d - v_0\right) \times \\ \left(\kappa_1 + 2\kappa_2 r_d^2\right) \end{array} \\ \hline \begin{array}{c} 2d_x^2 \left(v_d - v_0\right)\left(u_d - u_0\right) \times \\ \left(\kappa_1 + 2\kappa_2 r_d^2\right) \end{array} & \begin{array}{c} \left(1 + \kappa_1 r_d^2 + \kappa_2 r_d^4\right) + \\ 2\left(\left(v_d - v_0\right) d_y\right)^2 \times \\ \left(\kappa_i + 2\kappa_2 r_d^2\right) \end{array} \end{array} \end{pmatrix} \tag{34}$$

The Jacobian for the distortion is computed by inverting expression (34)

$$\left.\frac{\partial \mathbf{h}_d}{\partial (u_u, v_u)}\right|_{(u_u, v_u)} = \left(\left.\frac{\partial \mathbf{h}_u}{\partial (u_d, v_d)}\right|_{\mathbf{h}_d(u_u, v_u)} \right)^{-1}. \tag{35}$$

Unified Inverse Depth - linearity analysis

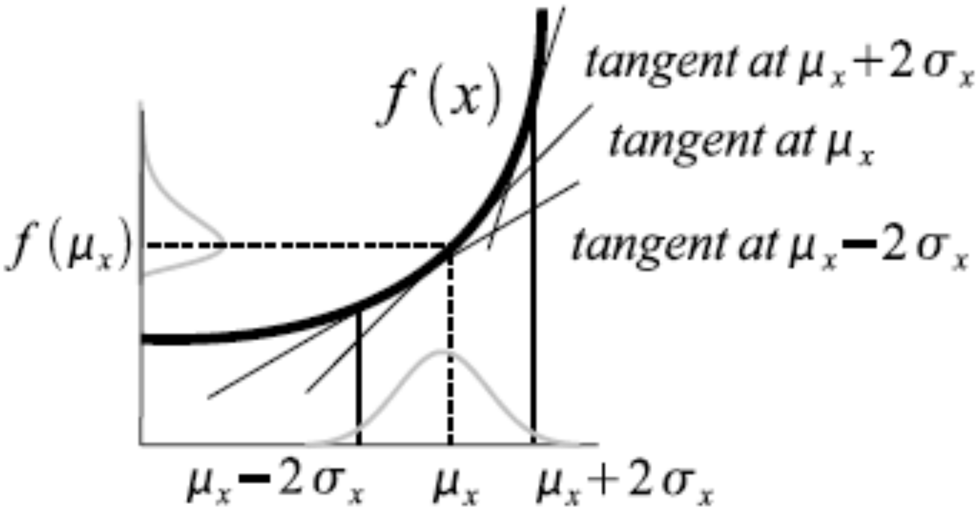


Fig. 2. The first derivative variation in $[\mu_x - 2\sigma_x, \mu_x + 2\sigma_x]$ codes the departure from Gaussianity in the propagation of the uncertain variable through a function.

Unified Inverse Depth - linearity analysis

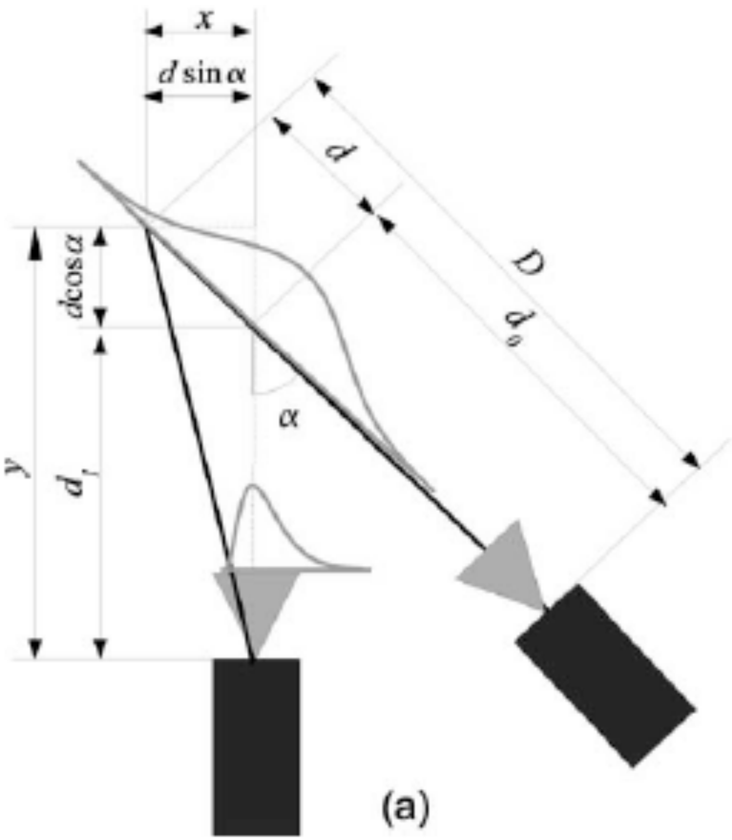
$$\frac{\partial f}{\partial x}(\mu_x + \Delta x) \approx \left.\frac{\partial f}{\partial x}\right|_{\mu_x} + \left.\frac{\partial^2 f}{\partial x^2}\right|_{\mu_x} \Delta x.$$

$$\left| \frac{\frac{\partial f}{\partial x}(\mu_x - 2\sigma_x) - \frac{\partial f}{\partial x}(\mu_x + 2\sigma_x)}{\frac{\partial f}{\partial x}(\mu_x)} \right|$$

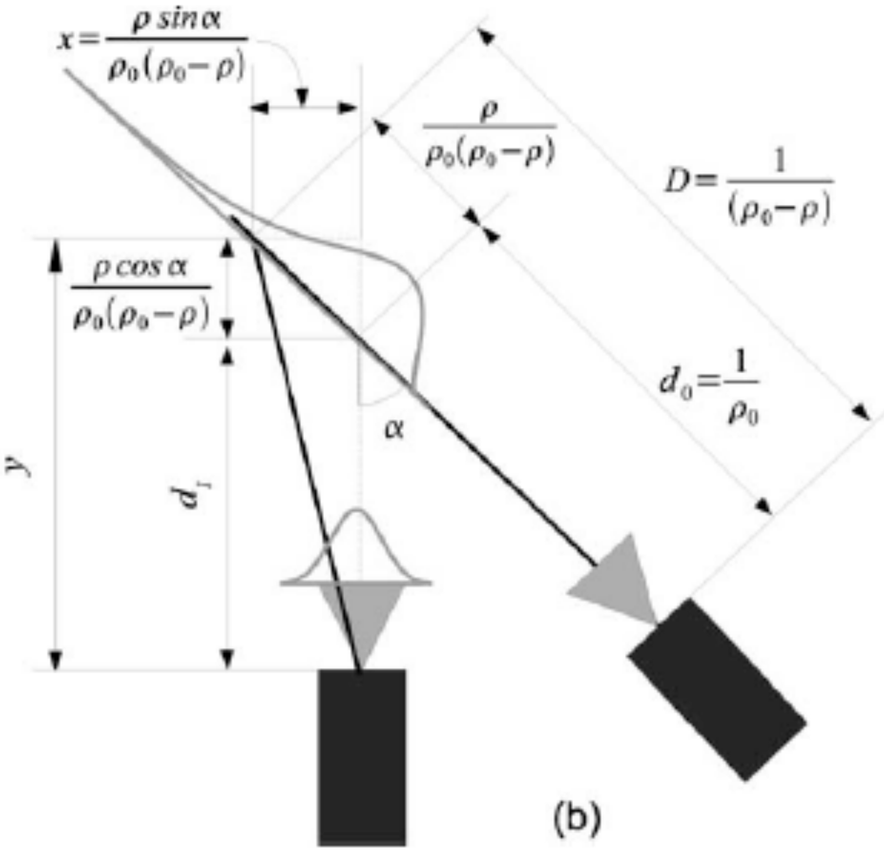
$$L = \left| \frac{\left.\frac{\partial^2 f}{\partial x^2}\right|_{\mu_x} 2\sigma_x}{\left.\frac{\partial f}{\partial x}\right|_{\mu_x}} \right|.$$

Unified Inverse Depth - linearity analysis

XYZ | **Inverse Depth**



$$\frac{4\sigma_d}{d_1}|\cos\alpha|$$



$$\frac{4\sigma_\rho}{\rho_0}\left|1-\frac{d_0}{d_1}\cos\alpha\right|$$

Unified Inverse Depth - feature initialization

The initial location for a newly observed feature inserted into the state vector is

$$\hat{\mathbf{y}}\left(\hat{\mathbf{r}}^{WC}, \hat{\mathbf{q}}^{WC}, \mathbf{h}, \rho_0\right) = \left(\hat{x}_i \quad \hat{y}_i \quad \hat{z}_i \quad \hat{\theta}_i \quad \hat{\phi}_i \quad \hat{\rho}_i\right)^\top \quad (21)$$

$$\left(\hat{x}_i \quad \hat{y}_i \quad \hat{z}_i\right)^\top = \hat{\mathbf{r}}^{WC}$$

$$\mathbf{h}^W = \mathbf{R}_{WC}\left(\mathbf{q}^{\hat{W}C}\right)\left(\upsilon \quad \nu \quad 1\right)^\top$$

$$\begin{pmatrix} \theta_i \\ \phi_i \end{pmatrix} = \begin{pmatrix} \arctan\left(\mathbf{h}_x^W, \mathbf{h}_z^W\right) \\ \arctan\left(-\mathbf{h}_y^W, \sqrt{\mathbf{h}_x^{W\,2} + \mathbf{h}_z^{W\,2}}\right) \end{pmatrix}$$

$$\hat{\rho}_i = \rho_0$$

Unified Inverse Depth - feature initialization

The covariance of $\hat{x}_i, \hat{y}_i, \hat{z}_i, \hat{\theta}_i$, and $\hat{\phi}_i$ is derived from the image measurement error covariance $\mathbf{R}_i$ and the state covariance estimate $\hat{\mathbf{P}}_{k|k}$.

The initial value for ρ_0 and its standard deviation are set empirically such that the 95% confidence region spans a range of depths from close to the camera up to infinity. In our experiments, we set $\hat{\rho}_0 = 0.1, \sigma_\rho = 0.5$, which gives an inverse depth confidence region $[1.1, -0.9]$. Notice that infinity is included in this range. Experimental validation has shown that the precise values of these parameters are relatively unimportant to the accurate operation of the filter as long as infinity is clearly included in the confidence interval.

Unified Inverse Depth - improvements

1. switching from Unified Inverse Depth to XYZ;
2. sharing camera pose for all feature newly observed at the same time (T. Pietzsch BMVC 2008; Imre et al. ICRA 2009)

Unified Inverse Depth

inverse_out.davison.nontro.avi (58.9Mb) the movie shows the method initialising features during a challenging outdoor motion with some features at very large depths. Some of these features retain very high depth uncertainties throughout the sequence.

inverseDepth_indoor.avi (11.7 MB) shows simultaneous localization and mapping, from a hand-held camera observing an indoor scene. All the processing is automatic, the image sequence being the only sensorial information used as input. It is shown as a top view of the computed camera trajectory and 3-D scene map. Image sequence is acquired with a hand-held camera 320 £ 240 at 30 frames/second.

inverseDepth_outdoor.avi (12.4 MB) shows real-time simultaneous localization and mapping, from a hand-held camera observing an outdoor scene, including rather distant features. All the processing is automatic, the image sequence being the only sensorial information used as input. It is shown as a top view of the computed camera trajectory and 3-D scene map. Image sequence is acquired with a hand-held camera 320 £ 240 at 30 frames/second. The processing is done with a standard laptop.

inverseDepth_loopClosing.avi (10.2MB) shows simultaneous localization and mapping, from a hand-held camera observing a loop-closing indoor scene. All the processing is automatic, the image sequence being the only sensorial information used as input. It is shown as a top view of the computed camera trajectory and 3-D scene map. Image sequence is acquired with a hand-held camera 320 £ 240 at 30 frames/second.

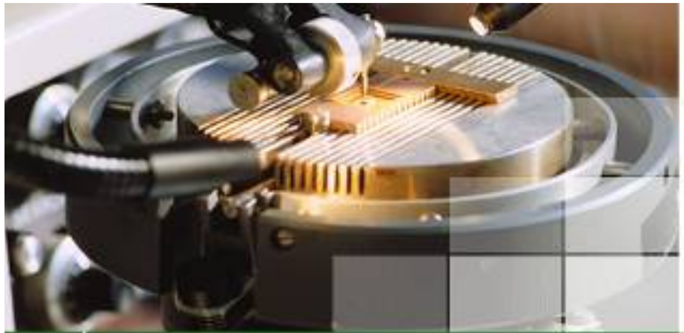
Unified Inverse Depth

inverseDepth_indoor.avi

inverseDepth_outdoor.avi

inverseDepth_loopClosing.avi

inverse_out.davison.nontro.avi



POLITECNICO DI MILANO

Dipartimento di
Elettronica e Informazione

Monocular SLAM: Inverse Scaling

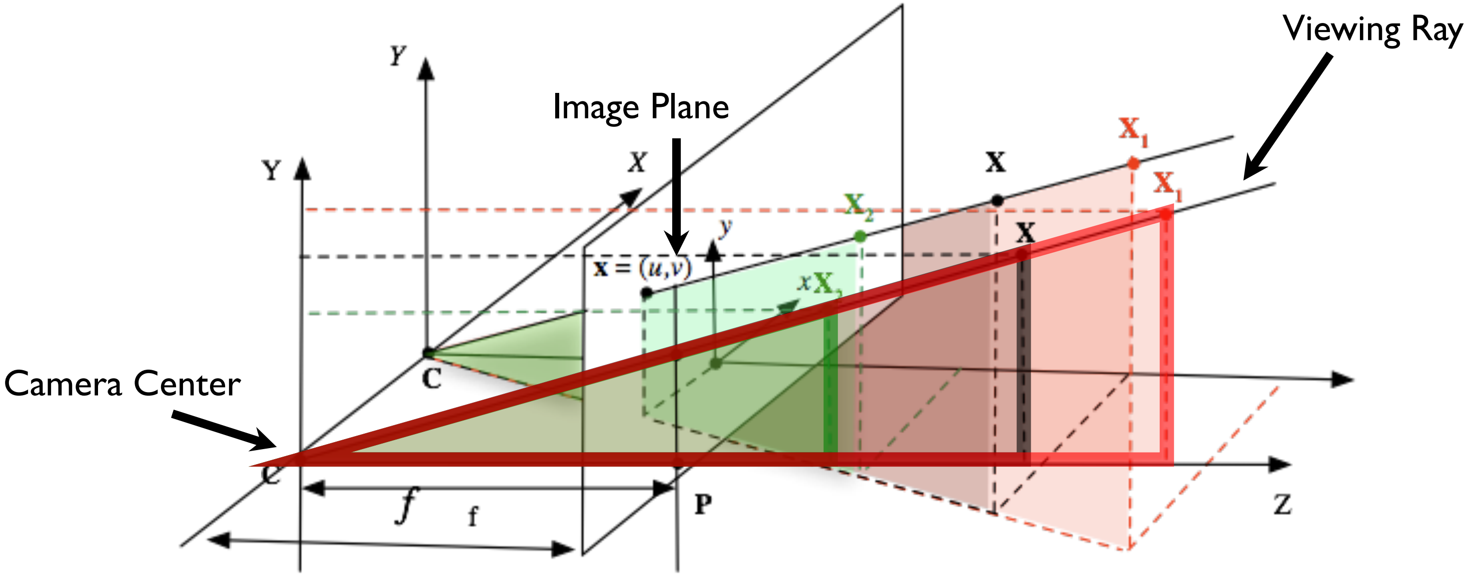
3D Structure from Visual Motion

M. Matteucci, Vincenzo Caglioti, Marco Marcon,
Davide Migliore, ***Domenico G. Sorrenti***

sorrenti@disco.unimib.it
Dip. Informatica, Sistemistica e Comunicazione, Università di Milano - Bicocca
Informatica e Robotica per l'Automazione Lab

Inverse Scaling Parametrization

Idea:



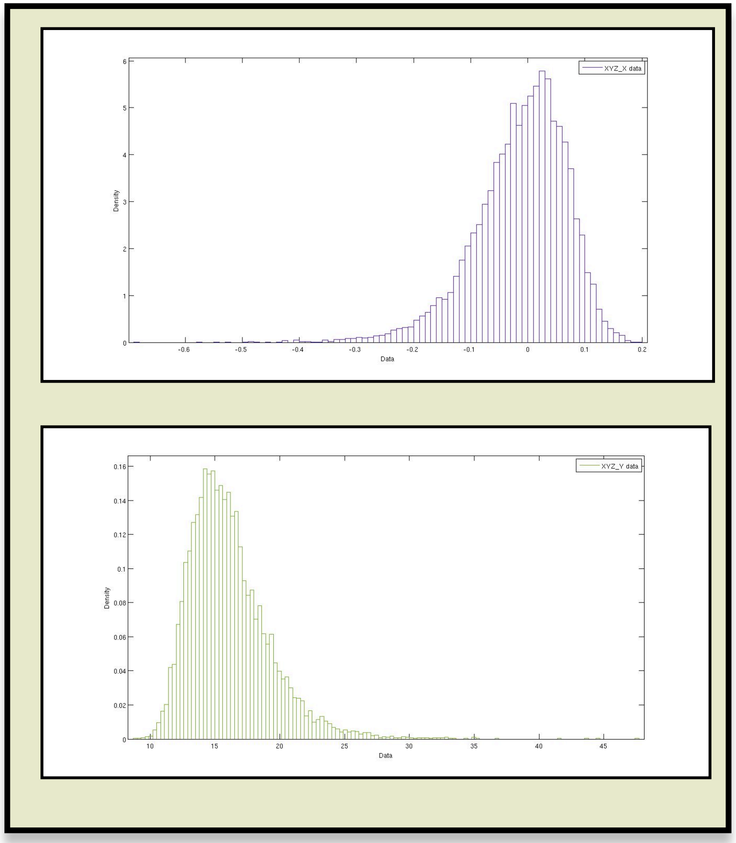
$$\mathbf{X} = \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \qquad \mathbf{X}_i = \begin{pmatrix} \alpha_i X \\ \alpha_i Y \\ \alpha_i Z \\ 1 \end{pmatrix} \longrightarrow \mathbf{X}_i \equiv \frac{1}{\alpha_1}\mathbf{X}_i = \begin{pmatrix} X \\ Y \\ Z \\ 1/\alpha_i \end{pmatrix}$$

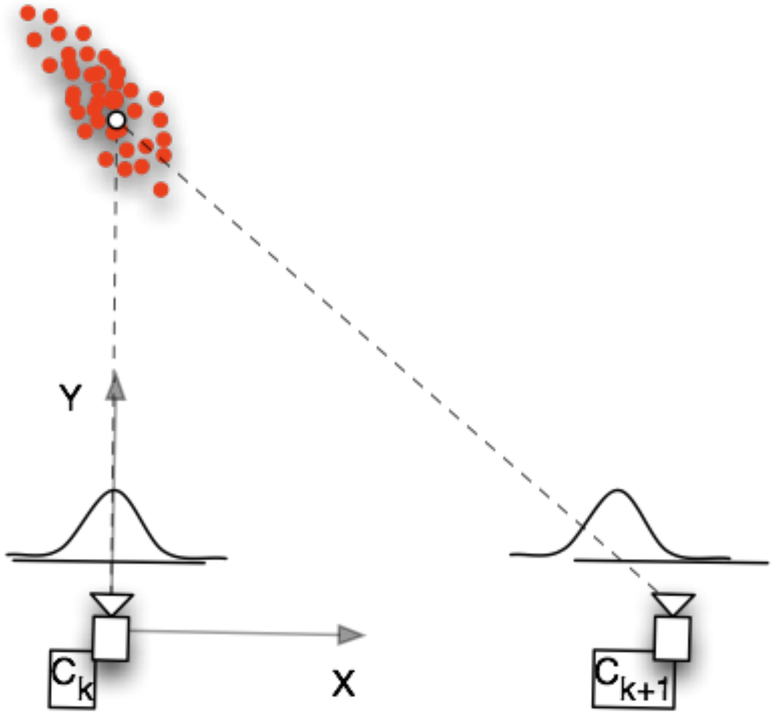
$$\mathbf{X} = \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \equiv \begin{pmatrix} u \\ v \\ f \\ 1/\alpha \end{pmatrix} \equiv \begin{pmatrix} u \\ v \\ f \\ \omega \end{pmatrix},$$

Inverse Scaling Parametrization

Uncertainty Modeling - Gaussian approximation

MonteCarlo simulation

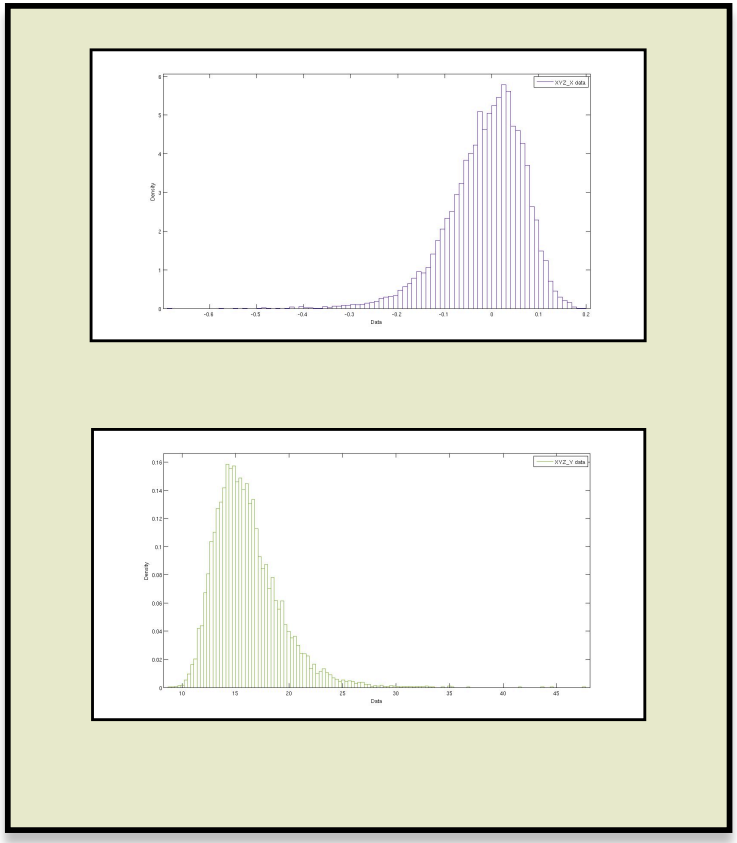




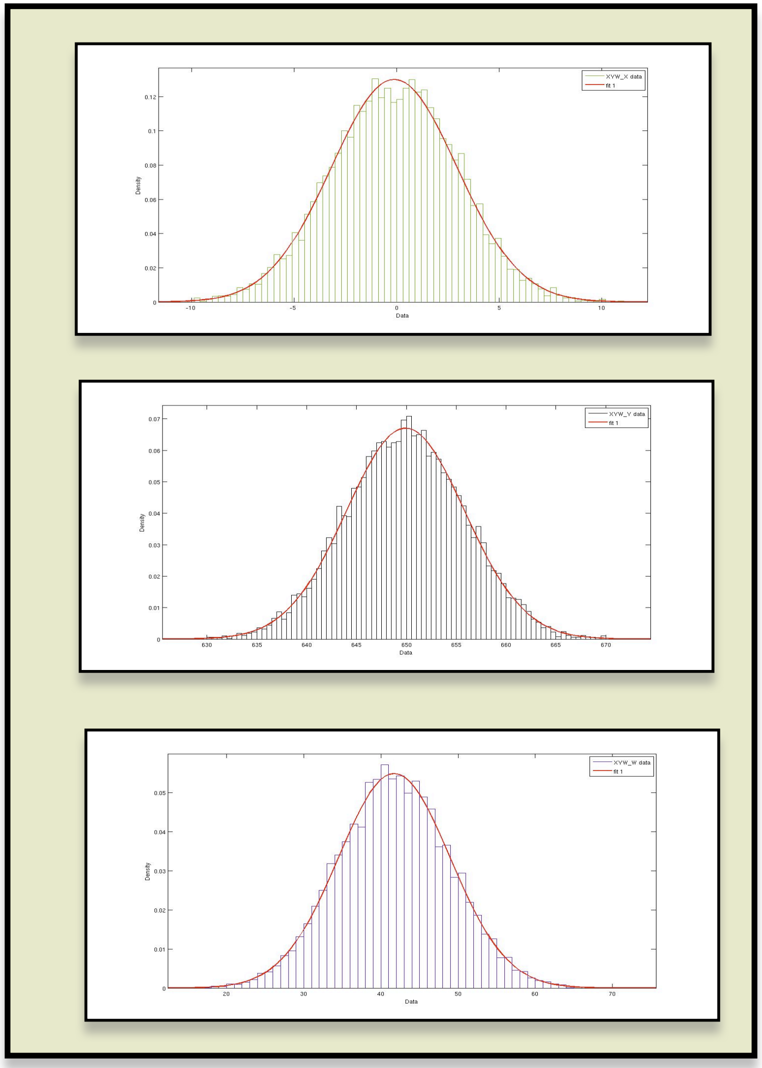
Inverse Scaling Parametrization

Uncertainty Modeling - Gaussian approximation

MonteCarlo simulation

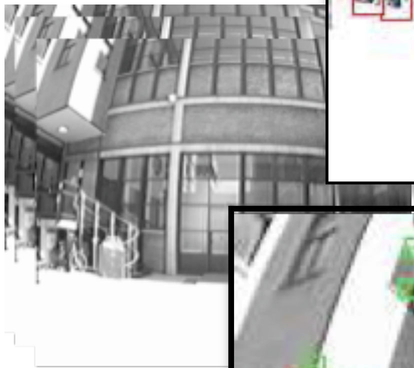


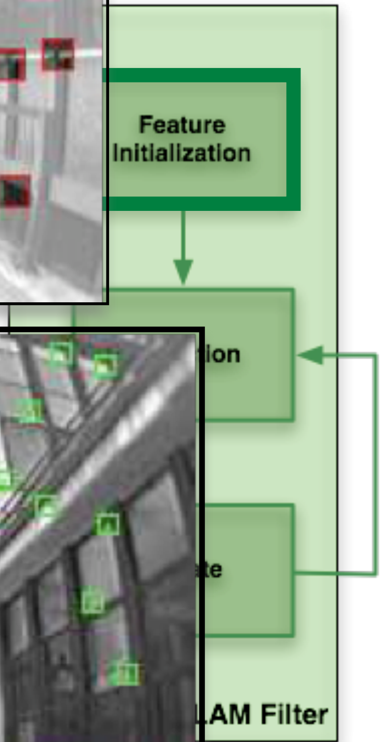
Inverse Scaling Representation



MonoSLAM wit

Extended Kalman



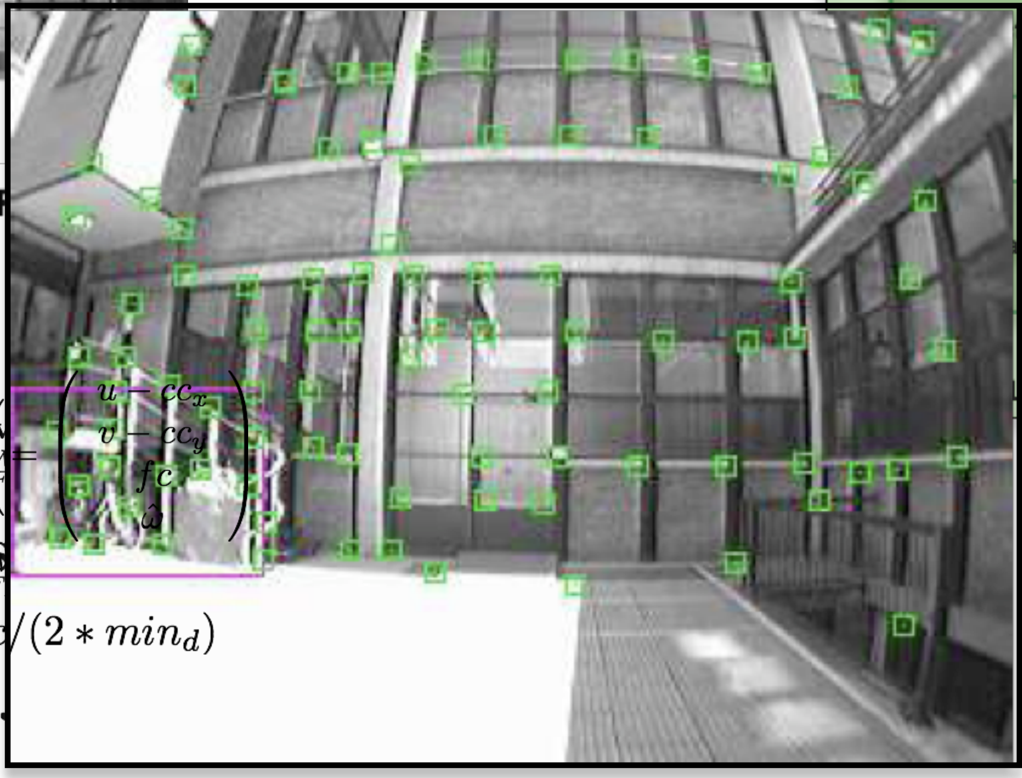


Video F

$$\mathbf{x}_k = [\ \mathbf{x}^W_{C_k}$$

$$\hat{\omega} = fc/(2 * min_d)$$

$$\mathbf{J}_1 = [$$

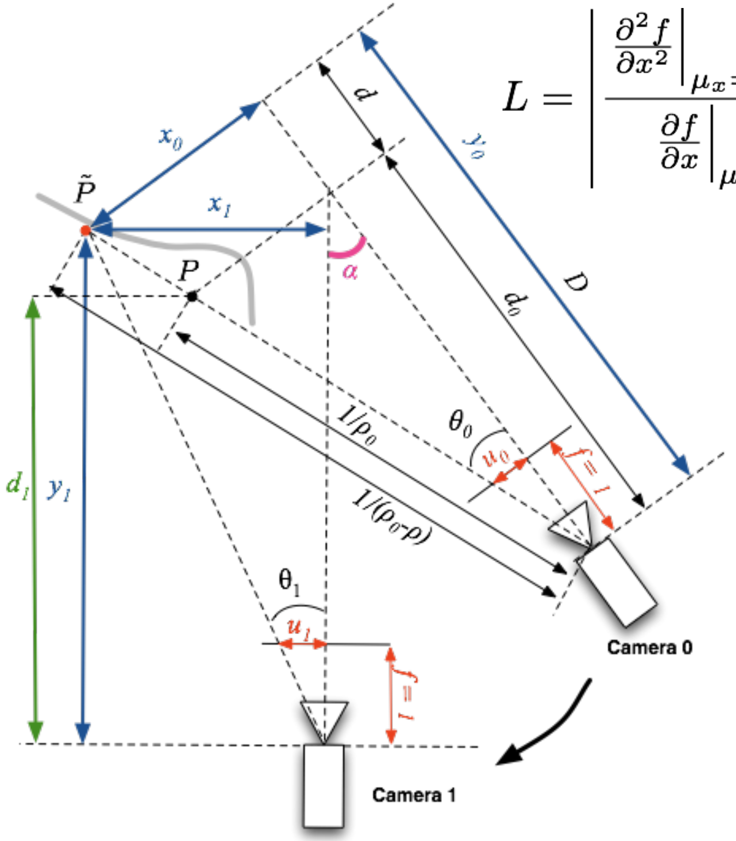


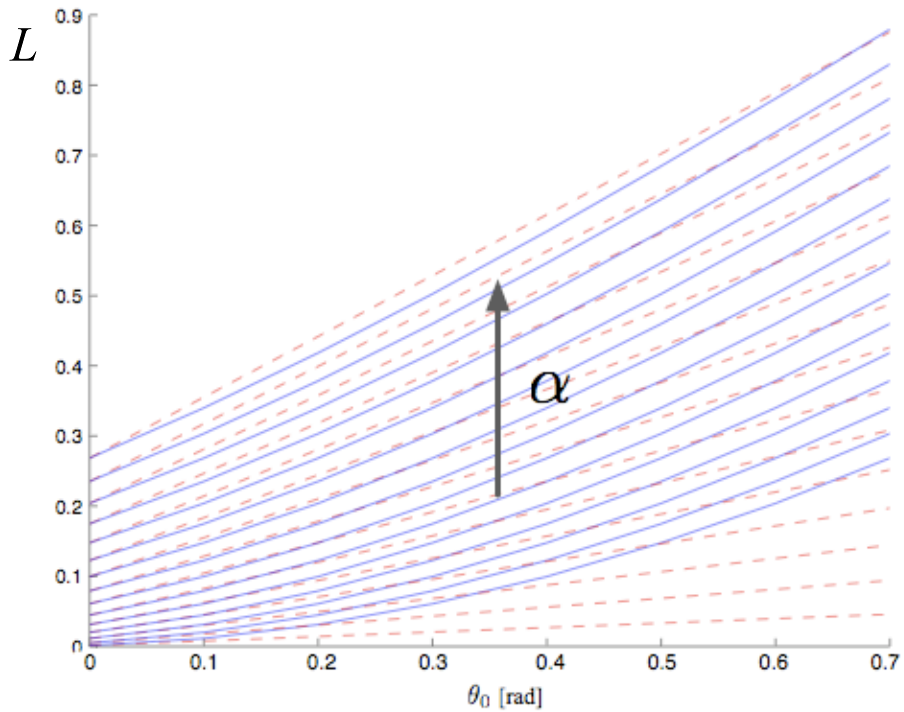
Inverse Scaling Parametrization

Improvements w.r.t. Unified Inverse Depth:

- Measurement model non-linearity

$$L = \left| \frac{\left.\frac{\partial^2 f}{\partial x^2}\right|_{\mu_x=0} 2\sigma_x}{\left.\frac{\partial f}{\partial x}\right|_{\mu_x=0}} \right|,$$

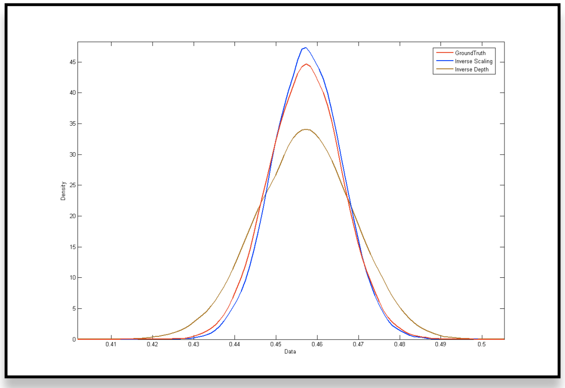


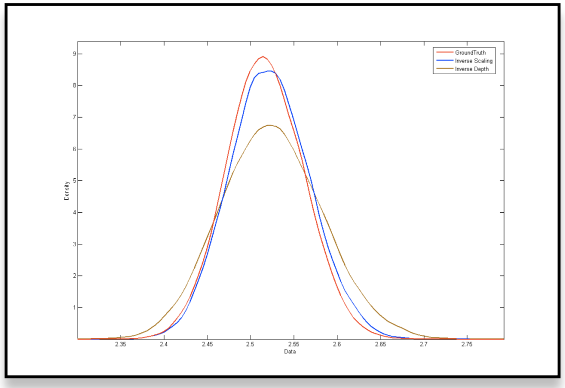


Inverse Scaling Parametrization

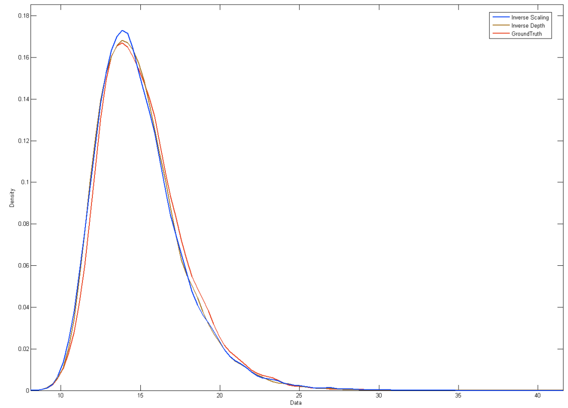
Improvements w.r.t. Unified Inverse Depth:

- Uncertainty Modeling - Comparison with Inverse Depth





Feature - 2.5m

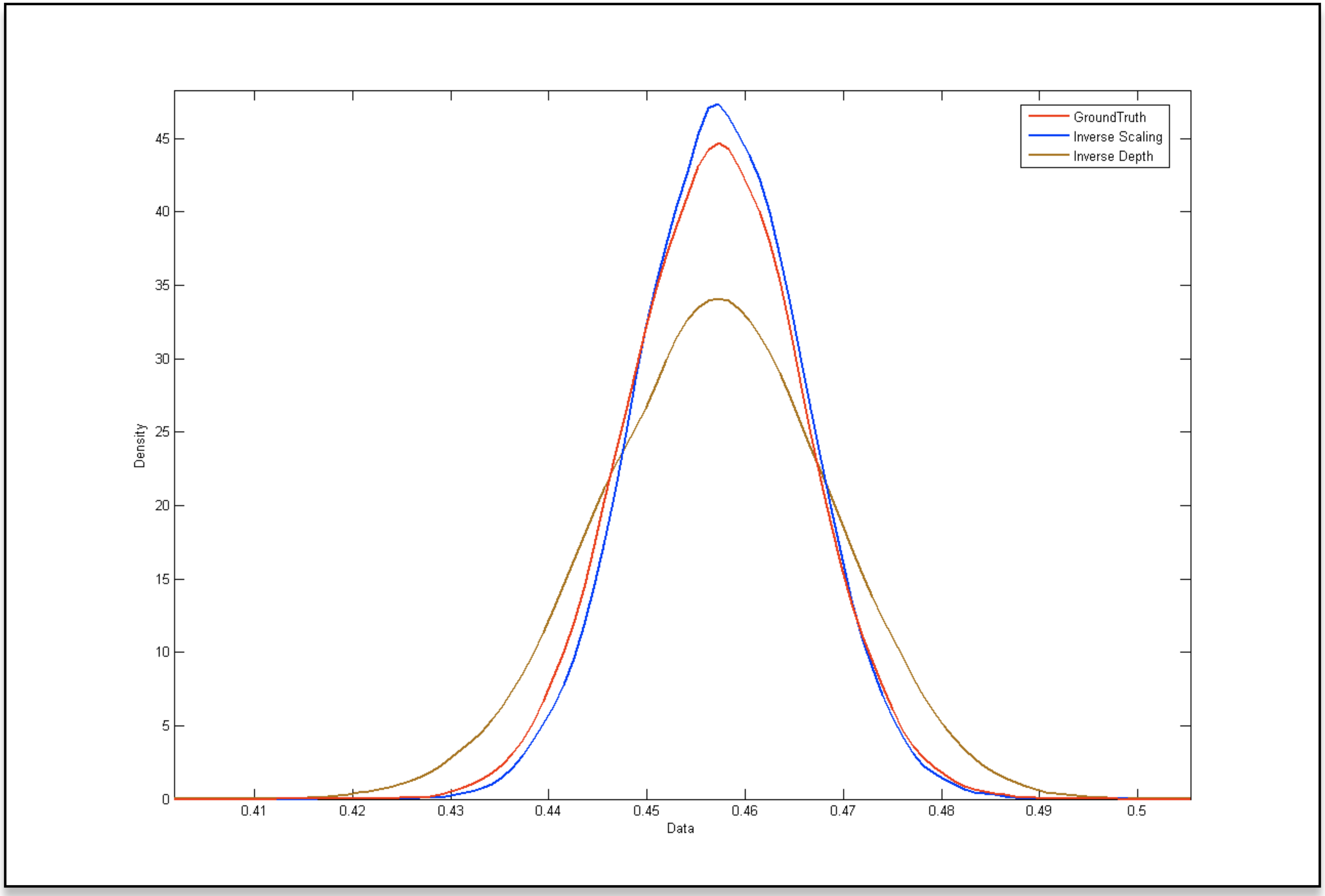


Feature - 15m

Inverse Scaling Parametrization

Improvements w.r.t. Unified Inverse Depth:

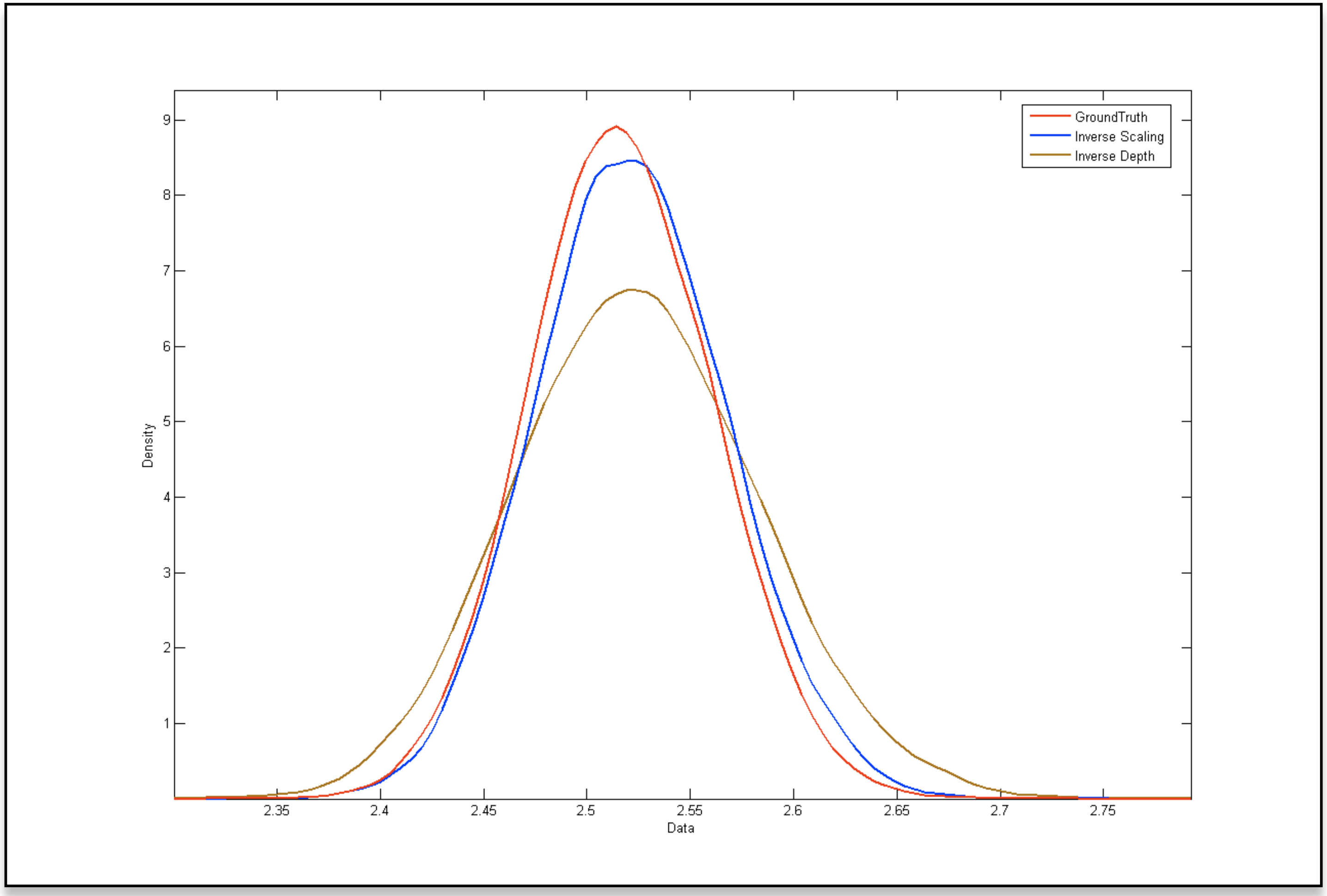
- Uncertainty Modeling - Comparison with Inverse Depth - X



Inverse Scaling Parametrization

Improvements w.r.t. Unified Inverse Depth:

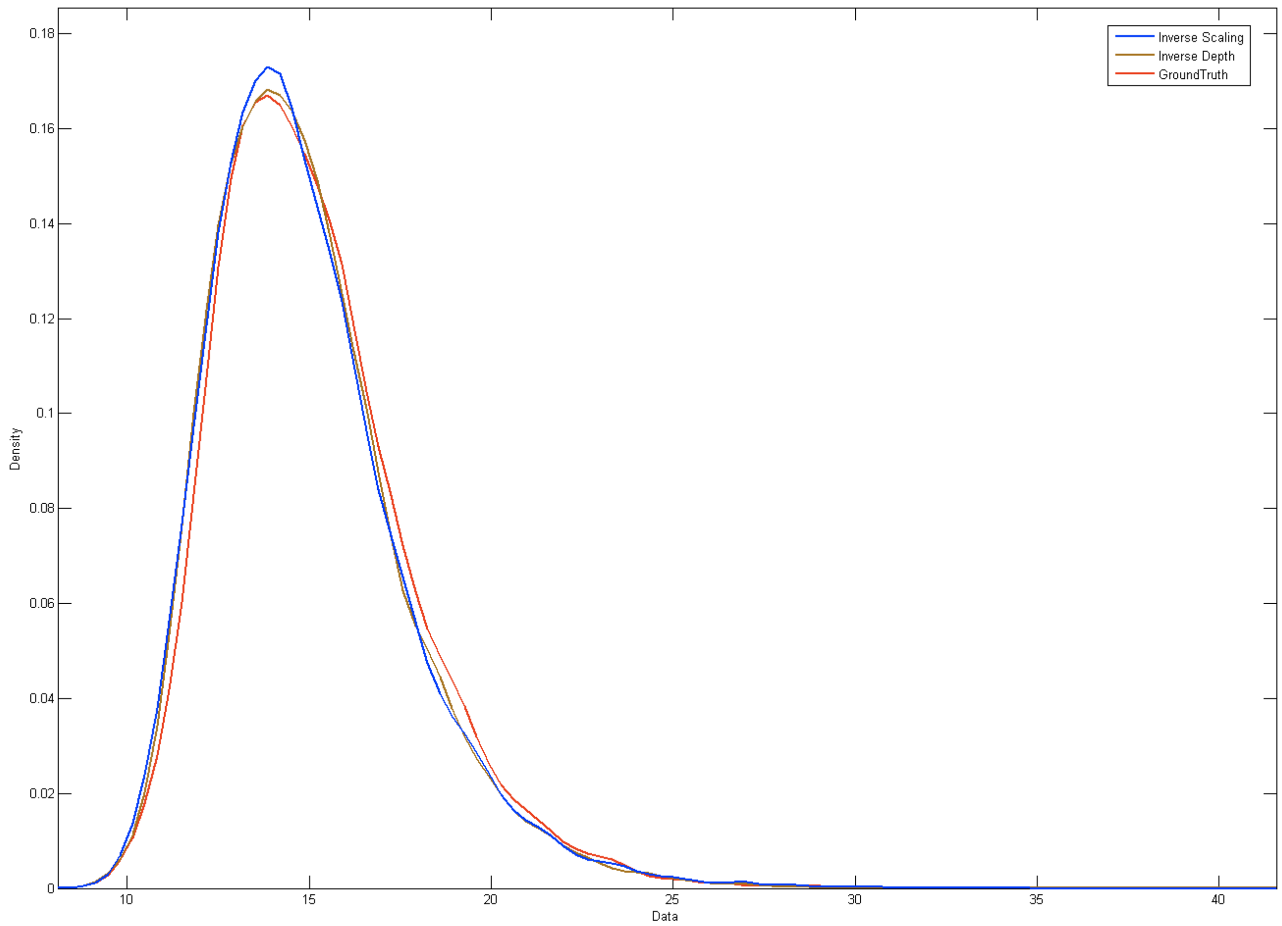
- Uncertainty Modeling - Comparison with Inverse Depth - Y



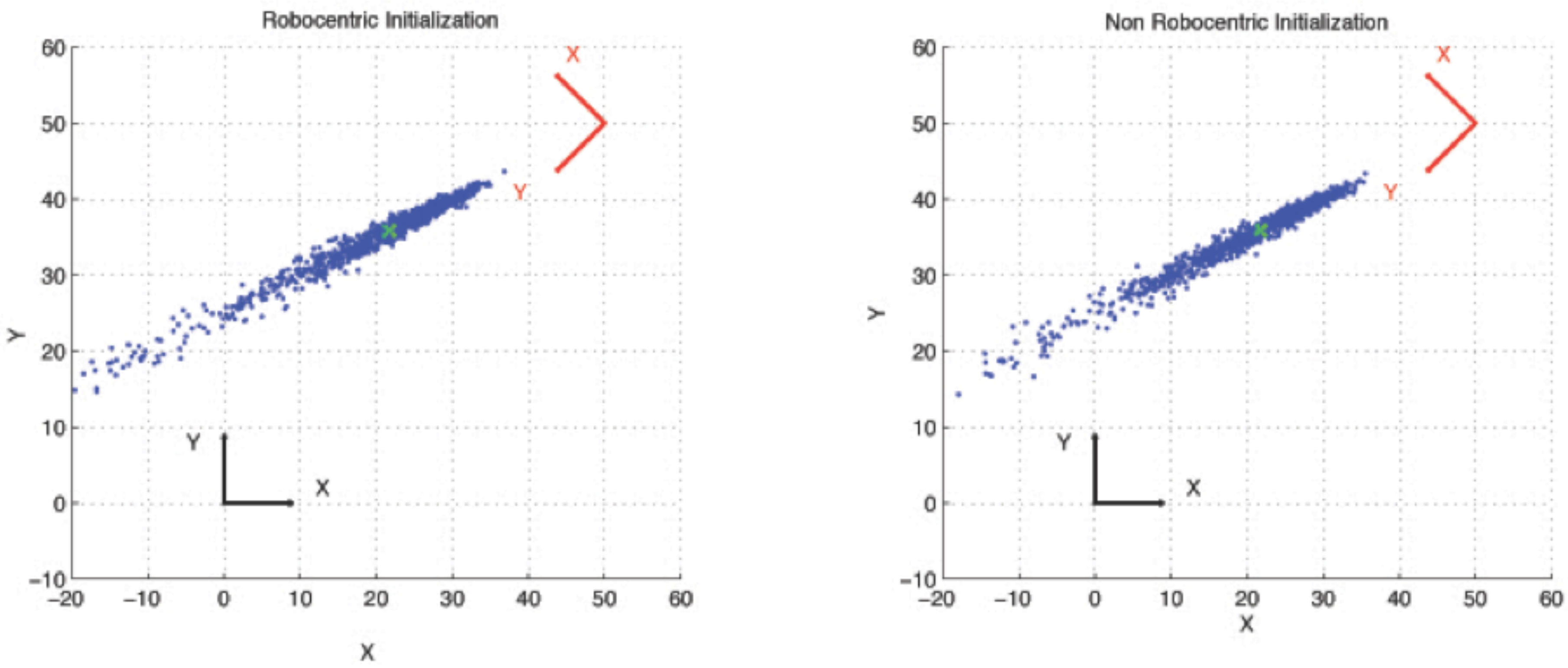
Inverse Scaling Parametrization

Improvements w.r.t. Unified Inverse Depth:

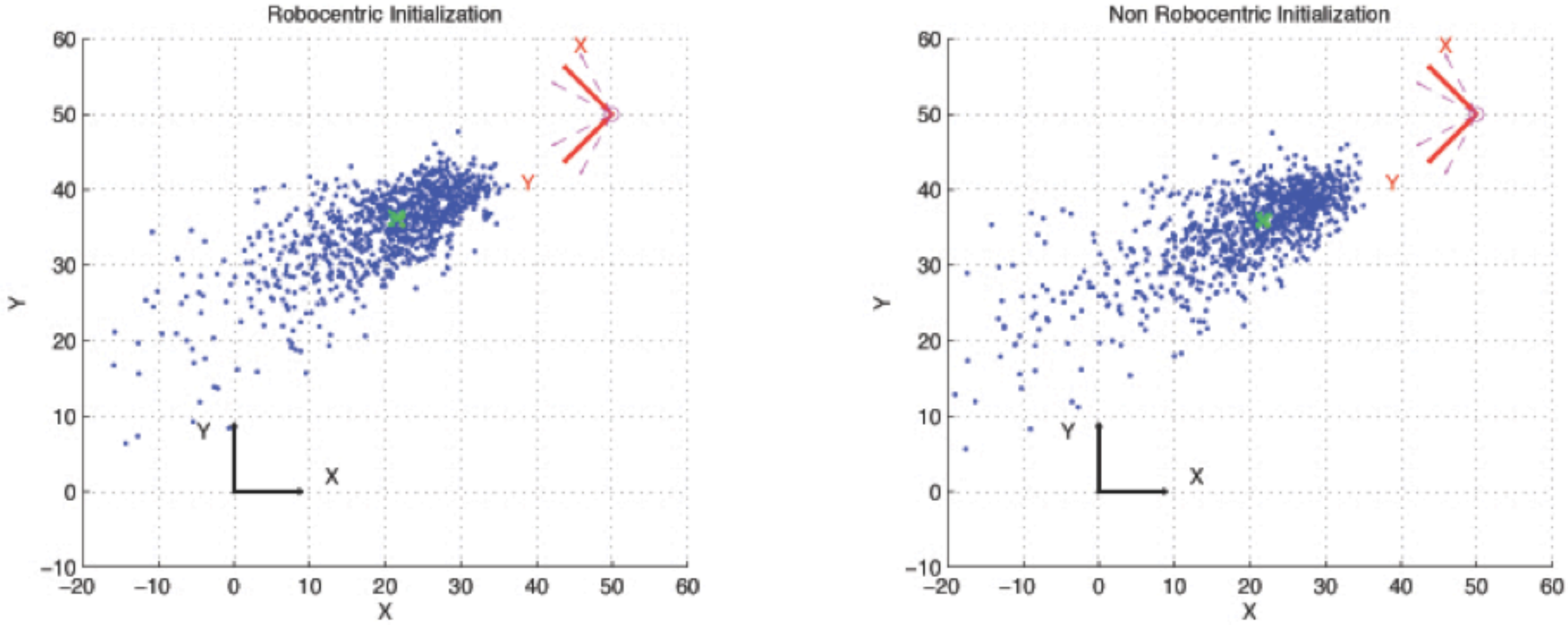
- Uncertainty Modeling - Comparison with Inverse Depth - Y



Changing Reference Frame



Changing Reference Frame



With uncertainty into rototranslation

Experimental Results

Simulated Dataset

movimento dall'interno

movimento dall'esterno

Experimental Results

Real dataset

desk and wall scene

desks, floor, wall, etc scene

Conclusions

Presented a new parametrization for Monocular SLAM

- Proper Uncertainty Modeling for both low and high parallax
- More linear than Unified Inverse Depth
- Only 4 parameters required
- Undelayed initialization

Ongoing works:

- MonoSLAM with Bearing Only Tracking
- Integration on Large Maps
- Self calibration

Thanks! Any question?