Inverse Depth Monocular SLAM

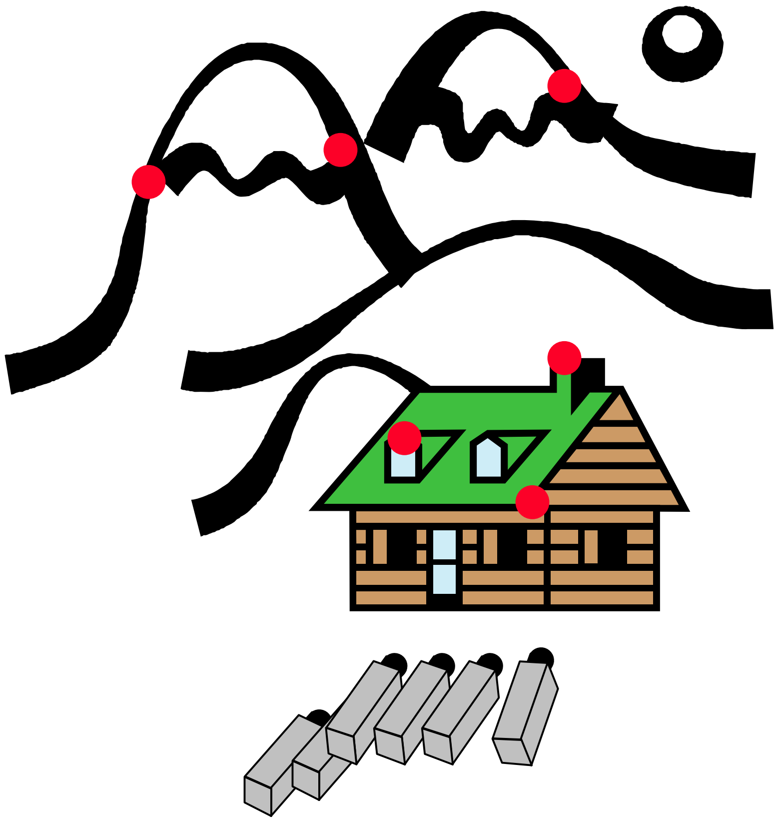
J. Civera, A.J. Davison and J.M.Montiel

Javier Civera

J.M.M. Montiel

A.J. Davison

Problem Statement



- Sequential simultaneous sensor location and map building at frame rate, 30Hz.
- Camera moves freely in 3D, 6dof camera motion
- Outdoors real scenes contains close and distant, even at infinity, features
- Main contribution codifying scene points with its inverse depth:
  1. Deals with low parallax cases
  2. Deals with both distant and close points
  3. Map features are initialised just from one image

ROBOTICS UNIVERSIDAD DE ZARAGOZA

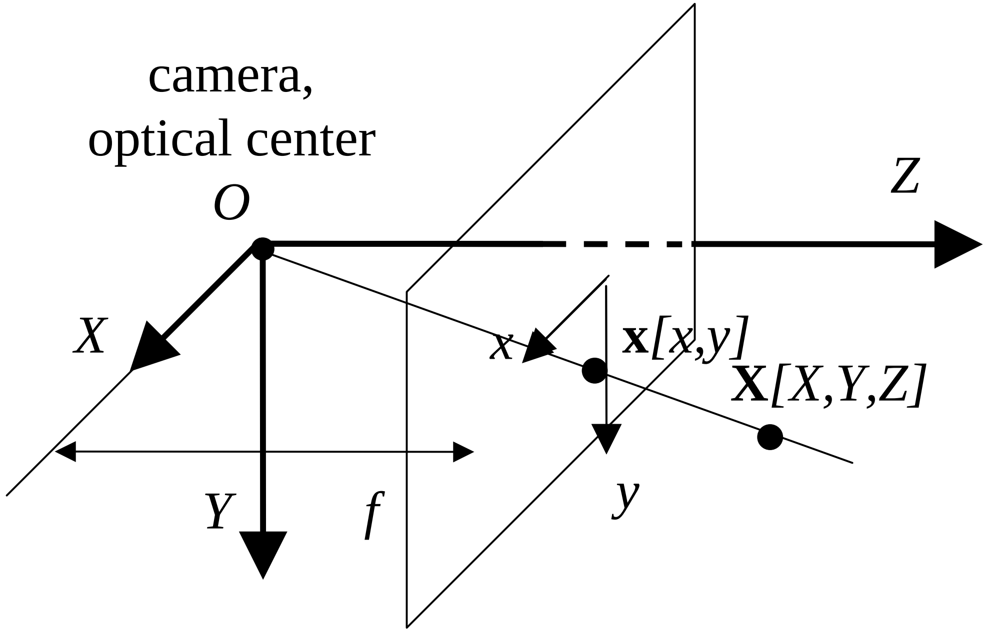
Javier Civera

J.M.M. Montiel

Imperial College London

A.J. Davison

Camera Geometry: Pure Bearing-only Sensor



- Camera detects rays
- A ray is defined by the optical center O and the observed point
- The images is used as the method to determined the detected ray
- Depth is not detected

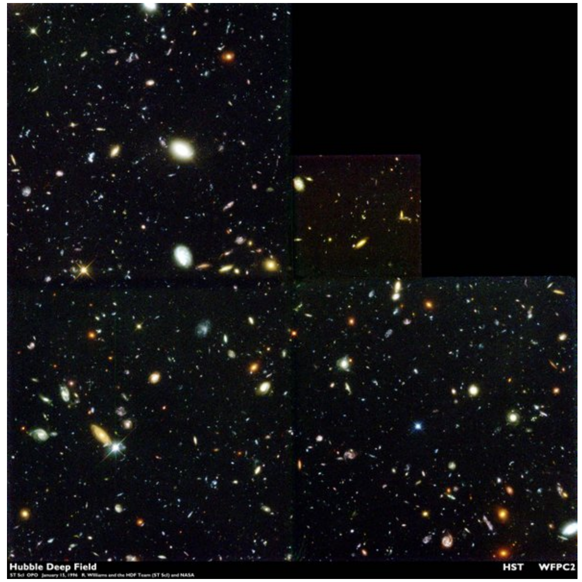


image of gallaxies

at 10^9 years light far from

the Earth

ROBOTICS
UNIVERSIDAD DE ZARAGOZA

Javier Civera

J.M.M. Montiel

Imperial College
London

A.J. Davison

Points at Infinity

- **projective cameras *do* observe points at infinity**
- **parallel lines meet at infinity, a projective camera does observe this intersection point as vanishing point**
- **we intend to code and exploit this points at infinity in the monocular SLAM problem**

ROBOTICS UNIVERSIDAD DE ZARAGOZA

Javier Civera

J.M.M. Montiel

Imperial College London

A.J. Davison

Parallax

- **No parallax geometries**
  - **Camera rotation**
  - **Camera observing a scene plane**
- **Low parallax cases**
  - **Distant features compared with camera translation**
  - **Initial feature observation**

Javier Civera

J.M.M. Montiel

Imperial College London

A.J. Davison

State of the Art

- **SLAM, initially proposed by Smith and Cheesman, 1986, widespread usage in robotics for multisensor fusion [Casetellanos 1999], [Ferder 1999] [Thrun 2004]**
  - **Sequential approach**
  - **Ability to close loops, identifying features previously observed as reobserved. Complexity is linked to the scene not to the number of observations processed.**
- **SLAM used for computer vision, [Castellanos 2000], [Davison 1998] combined with odometry**
- **Monocular SLAM vision [Davison 2003]**
  - **Camera "following the laws of mechanics" motion model**
  - **Vision as the only sensor, no odometry.**
  - **Synergic usage of vision geometry and vision photometric map**
  - **Low parallax points avoided:**
    - » **Points represented as XYZ, only works with points close to the camera**
    - » **Delayed initialization**
- **Monocular SLAM Fast SLAM, inverse depth delayed initialization [Eade 2006]**

SFM,computer vision methods

ROBOTICS UNIVERSIDAD DE ZARAGOZA

Javier Civera

J.M.M. Montiel

Imperial College London

A.J. Davison

State of the Art

SLAM methods

- **Photogrammetric bundle adjustment, 60's**
  - **Normally only close points**
- **Computer vision geometry Hartley & Zisserman [Hartley 2003]**
  - **Robust statistics**
  - **Matching between several shots enforcing a coherence with a projective camera model**
  - **Applied to individual shots, to sequences with varying camera parameters**
  - **Applied for robot navigation [Nister 2003, Mouragnon 2006]**
  - **Not sequential**
  - **Wide-baseline performance**
  - **Routine usage of points at infinity**
- **Model selection problem [Torr ICCV98]**
  - **No parallax, homography model**
  - **Parallax epipolar geometry model**
  - **Increasing the frame rate, the interframe motion closes to a homography**

ROBOTICS UNIVERSIDAD DE ZARAGOZA

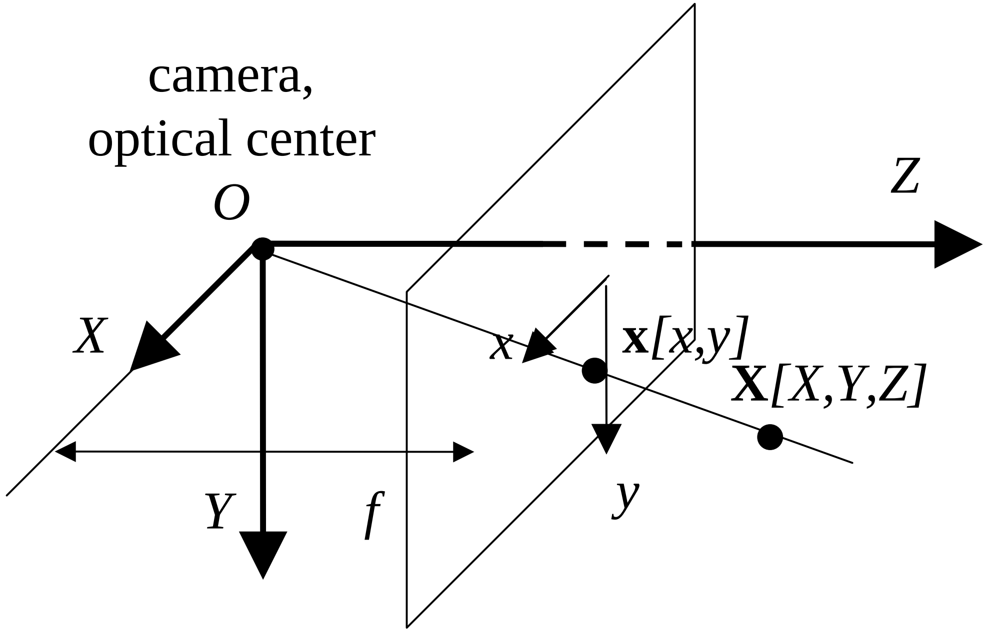
Javier Civera

J.M.M. Montiel

Imperial College London

A.J. Davison

Camera Geometry: Pure Bearing-only Sensor



- Camera detects rays
- A ray is defined by the optical center O and the observed point
- The images is used as the method to determined the detected ray
- Depth is not detected

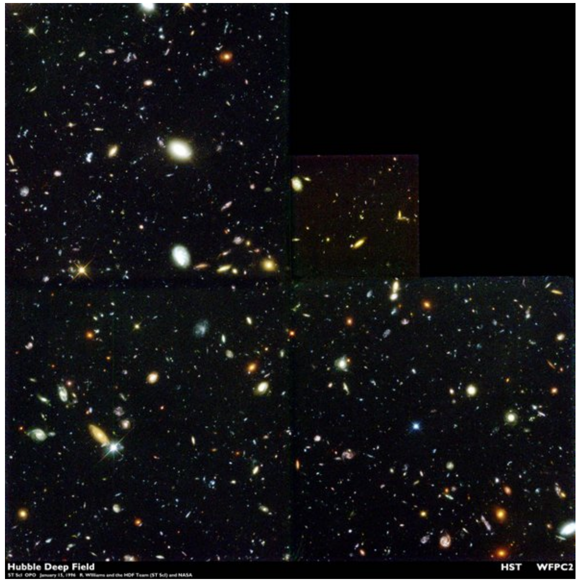


image of gallaxies

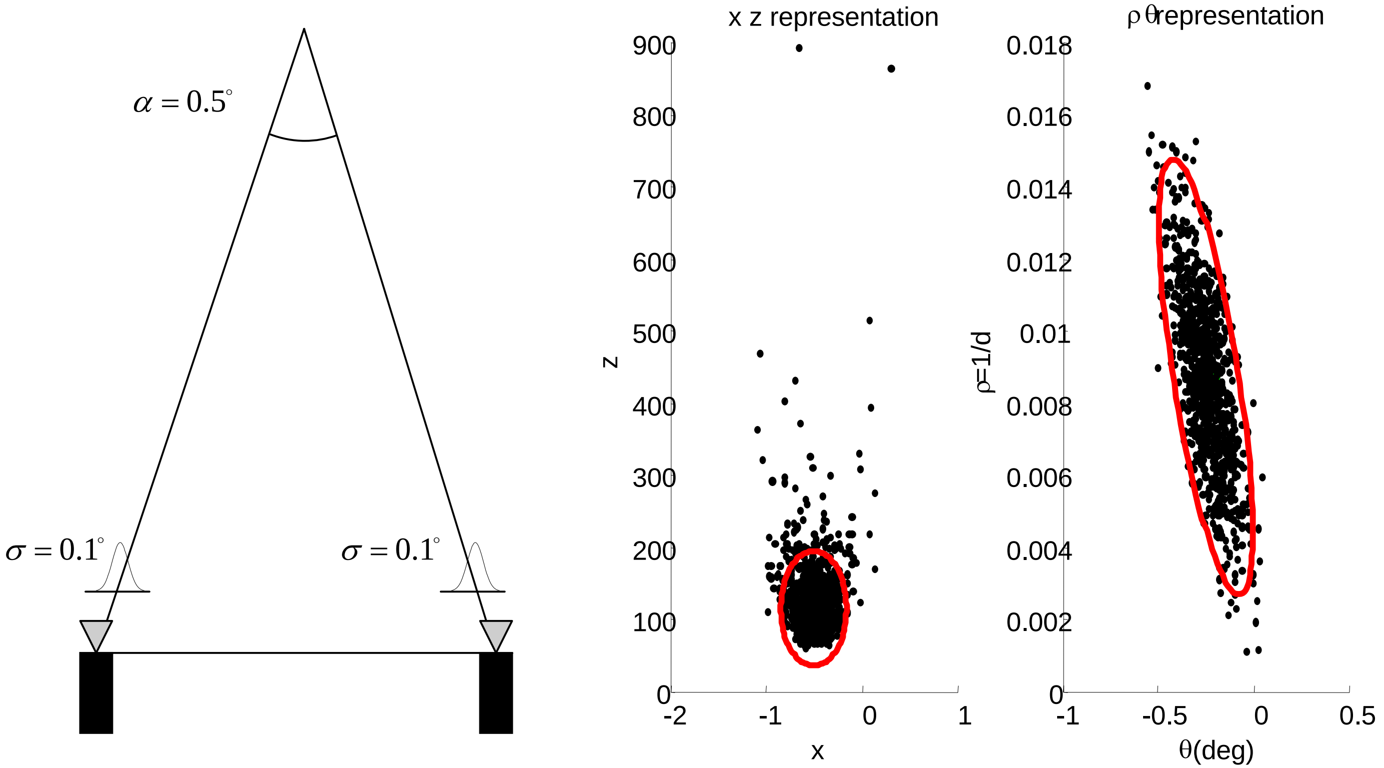
at 10^9 years light far from

the Earth

ROBOTICS
UNIVERSIDAD DE ZARAGOZA
Javier Civera
J.M.M. Montiel

Imperial College London
A.J. Davison

Gaussianity of the Inverse Depth Coding



Simulation: computing depth of a point from 2 views at known camera locations

- Non Gaussian in XZ
- Gaussian in 1/d, theta

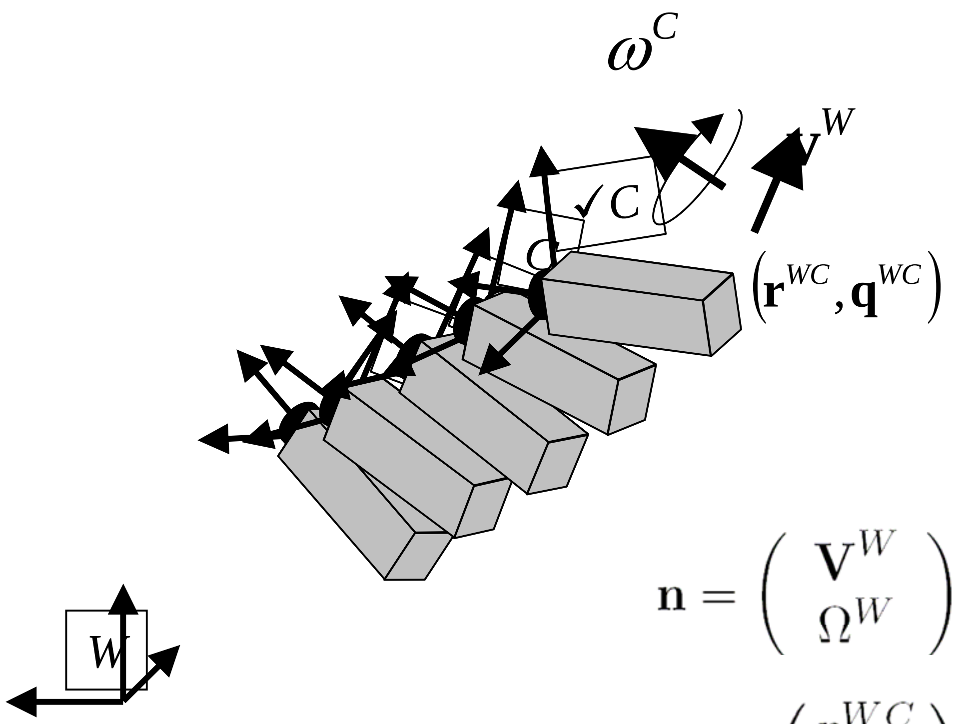
Javier Civera

J.M.M. Montiel

A.J. Davison

Camera Motion Priors

•



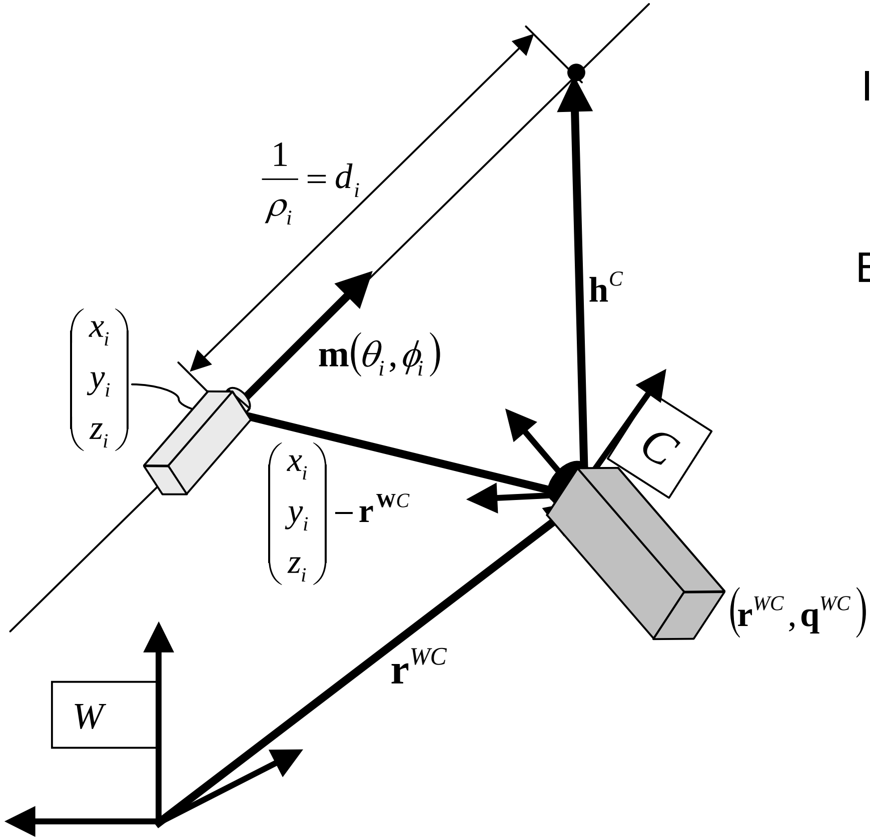
$$\mathbf{x}_v = \begin{pmatrix} \mathbf{r}^{WC} \\ \mathbf{q}^{WC} \\ \mathbf{V}^W \\ \omega^C \end{pmatrix}$$

Constant Velocity Motion Model
smooth camera motion
impulse acceleration noise

$$\mathbf{n} = \begin{pmatrix} \mathbf{V}^W \\ \Omega^W \end{pmatrix} = \begin{pmatrix} \mathbf{a}^W \Delta t \\ \alpha^W \Delta t \end{pmatrix}$$

$$\mathbf{f_v} = \begin{pmatrix} \mathbf{r}^{WC}_{k+1} \\ \mathbf{q}^{WC}_{k+1} \\ \mathbf{v}^W_{k+1} \\ \omega^C_{k+1} \end{pmatrix} = \begin{pmatrix} \mathbf{r}^{WC}_k + v^W_k \Delta t + a^W_k \Delta t^2 \\ \mathbf{q}^{WC}_k \times \mathbf{q}(\omega^C_k \Delta t + \alpha^C_k \Delta t^2) \\ \mathbf{v}^W_k + a^W_k \Delta t \\ \omega^C_k + \alpha^C_k \Delta t \end{pmatrix}$$

Scene Point Coding in Inverse Depth. Measurement Equation



Inverse dept point coding

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

Bearing only camera measurement

$$u = u_0 - \frac{f}{d_x}\frac{h_x}{h_z} \qquad v = v_0 - \frac{f}{d_y}\frac{h_y}{h_z}$$

Measurement equation

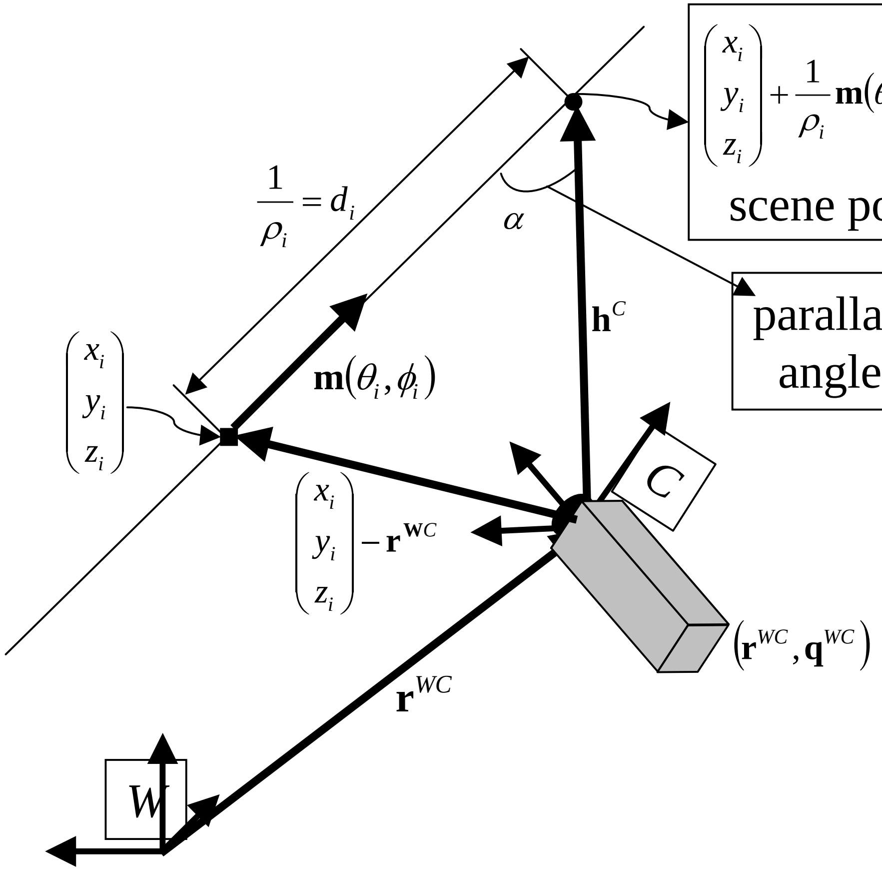
$$\mathbf{h}^C = \begin{pmatrix} h_x \\ h_y \\ h_z \end{pmatrix} = \mathbf{R}^{CW}\left(\rho_i\left(\begin{pmatrix} x_i \\ y_i \\ z_i \end{pmatrix} - \mathbf{r}^{WC}\right) + \mathbf{m}(\theta_i, \phi_i)\right)$$

Javier Civera

J.M.M. Montiel

A.J. Davison

Scene Point Coding in Inverse Depth. Measurement Equation



$$u = u_0 - \frac{f}{d_x}\frac{h_x}{h_z} \qquad v = v_0 - \frac{f}{d_y}\frac{h_y}{h_z}$$

$$\mathbf{h}^C = \begin{pmatrix}h_x\\h_y\\h_z\end{pmatrix} = \mathbf{R}^{CW}\left(\begin{pmatrix}x_i\\y_i\\z_i\end{pmatrix} + \frac{1}{\rho_i}\mathbf{m}(\theta_i,\phi_i) - \mathbf{r}^{WC}\right) =$$

$$\mathbf{R}^{CW}\left(\rho_i\left(\begin{pmatrix}x_i\\y_i\\z_i\end{pmatrix} - \mathbf{r}^{WC}\right) + \mathbf{m}(\theta_i,\phi_i)\right)$$

$\begin{pmatrix}x_i\\y_i\\z_i\end{pmatrix}$, $\mathbf{m}(\theta_i,\phi_i)$, those of the first time the feature is observed

$\rho_i\left(\begin{pmatrix}x_i\\y_i\\z_i\end{pmatrix} - \mathbf{r}^{WC}\right)$, parallax

distant point, $\rho_i \to 0, \Rightarrow$ parallax goes to zero

close camera locations, low baseline, $\begin{pmatrix}x_i\\y_i\\z_i\end{pmatrix} - \mathbf{r}^{WC}$, goes to zero

ROBOTICS UNIVERSIDAD DE ZARAGOZA

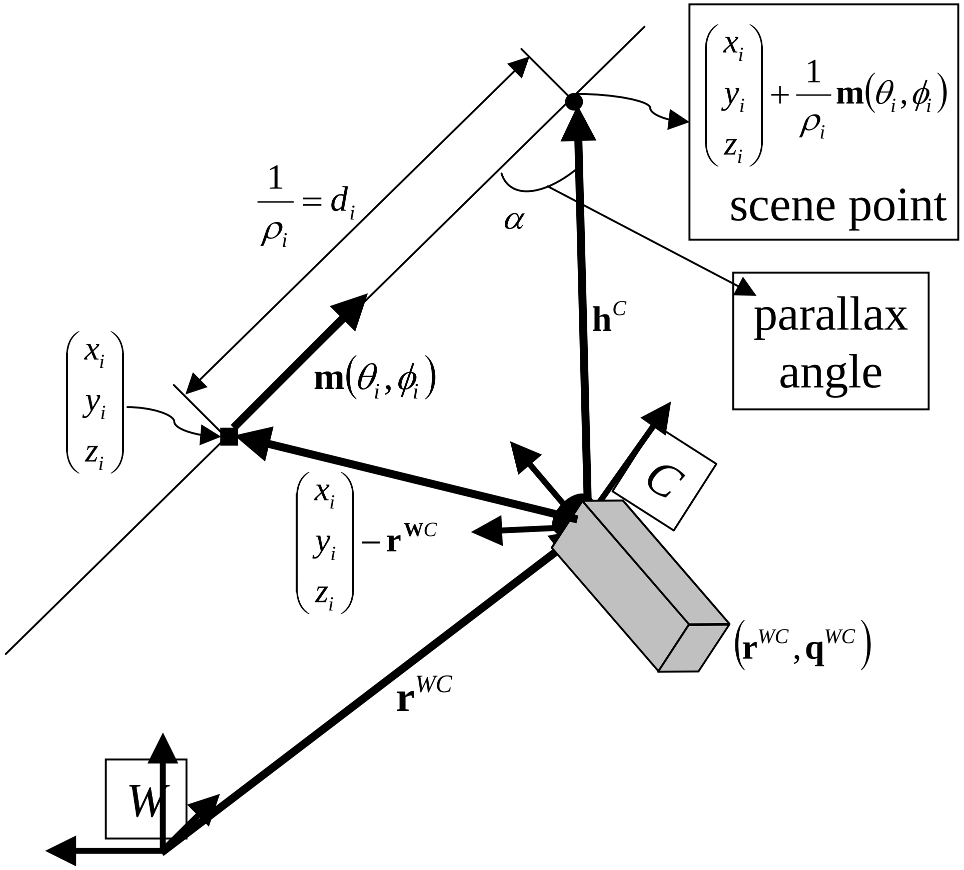
Javier Civera

J.M.M. Montiel

Imperial College London

A.J. Davison

Scene Point Coding in Inverse Depth. Measurement Equation



$$u = u_0 - \frac{f}{d_x}\frac{h_x}{h_z} \qquad v = v_0 - \frac{f}{d_y}\frac{h_y}{h_z}$$

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

low parallax, only point at infinity is observed:

$$\mathbf{h}^C \approx \mathbf{R}^{CW}(\mathbf{m}(\theta_{i,}\phi_i))$$

Javier Civera

J.M.M. Montiel

A.J. Davison

EKF sequential processing

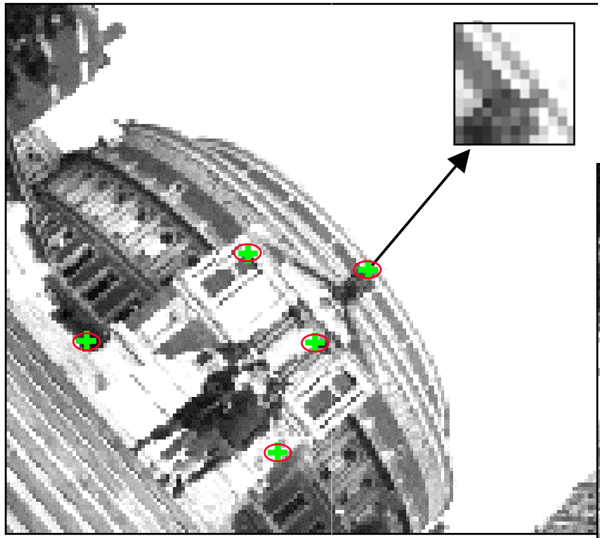
$$\hat{\mathbf{x}}_{k+1|k} = \begin{pmatrix} \mathbf{f}_v\left(\hat{\mathbf{x}}_{v,k|k}, 0\right) \\ \hat{\mathbf{y}}_{1,k|k} \\ \vdots \\ \hat{\mathbf{y}}_{n,k|k} \end{pmatrix}$$

$$\hat{\mathbf{P}}_{k+1|k} = \mathbf{F}\hat{\mathbf{P}}_{k+1|k}\mathbf{F}^\top + \mathbf{G}\begin{pmatrix} (\Delta t)^2\,\mathbf{P}_\alpha & 0 \\ 0 & 0 \end{pmatrix}\mathbf{G}^\top$$

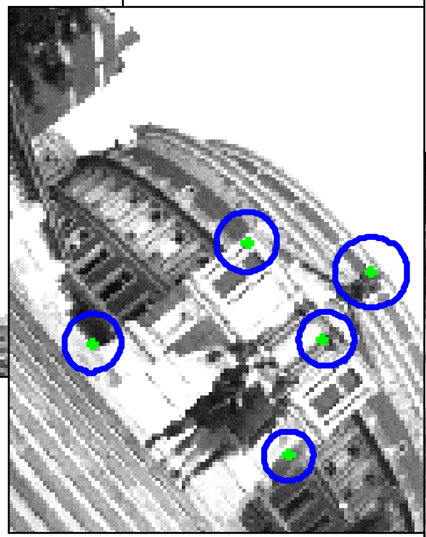
$$\hat{\mathbf{h}}_{i,k+1|k} = \mathbf{h}\left(\hat{\mathbf{y}}_{i,k+1|k}, \hat{\mathbf{x}}_{k+1|k}\right)$$

$$\mathbf{S}_{k+1|k} = \mathbf{H}\mathbf{P}_{k+1|k}\mathbf{H}^\top + \mathrm{diag}\left(\mathbf{R}_1, \ldots, \mathbf{R}_n\right)$$

stored patch



Estimation at step

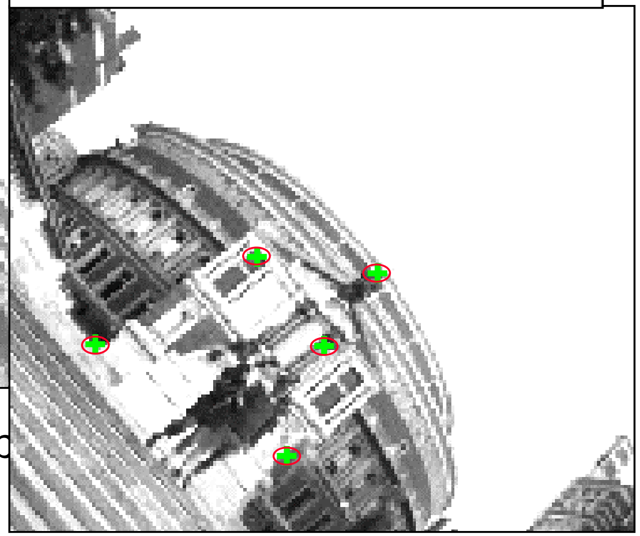


Prediction at step

$$\hat{\mathbf{x}}_{k+1|k+1} = \hat{\mathbf{x}}_{k+1|k} + \mathbf{K}\left(\mathbf{z}_m - \hat{\mathbf{h}}_m\right)$$

$$\mathbf{K} = \hat{\mathbf{P}}_{k+1|k}\mathbf{H}_m^\top\left(\mathbf{H}_m\hat{\mathbf{P}}_{k+1|k}\mathbf{H}_m^\top\right)^{-1}$$

$$\hat{\mathbf{P}}_{k+1|k+1} = \left(I - \mathbf{K}\mathbf{H}_m\right)\hat{\mathbf{P}}_{k+1|k}$$

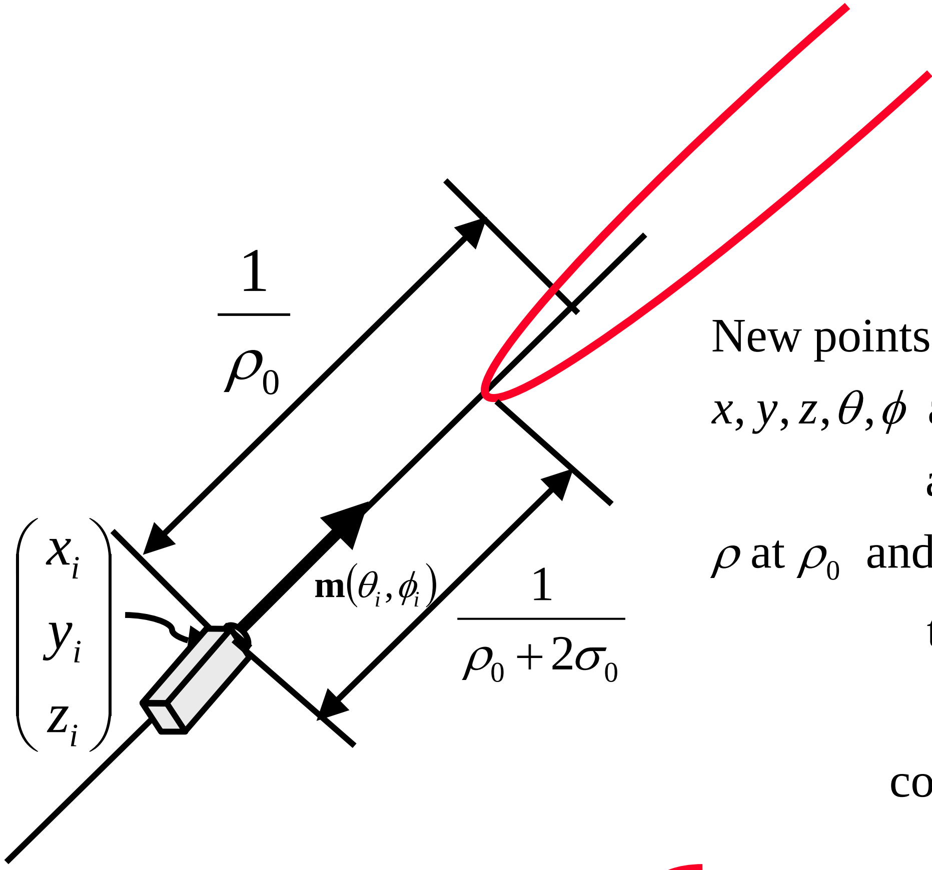


Update at step k

stored patch is searched in all acceptance region for the feature prediction

Javier Civera
J.M.M. Montiel
Imperial College London
A.J. Davison

Inverse Depth Feature Initialization Priors



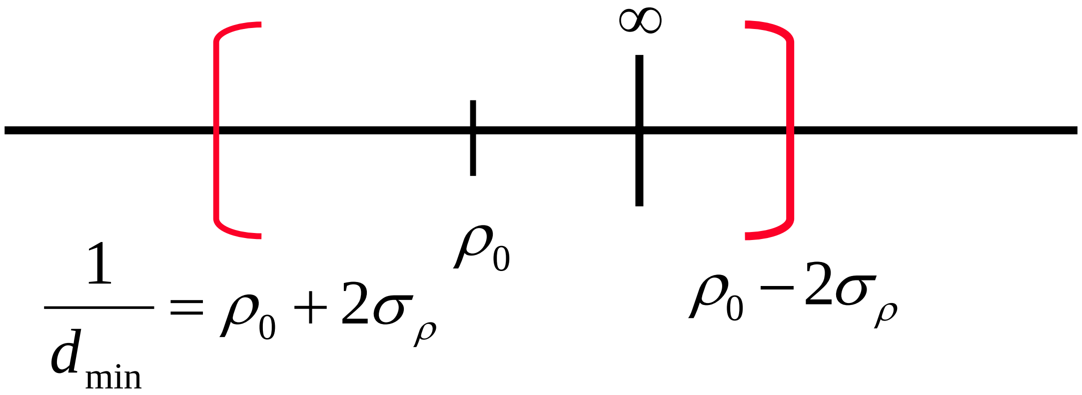
New points are observed from just an observation

x, y, z, θ, ϕ and the corresponding covariance are initialized from the first feature observation

ρ at ρ_0 and its covarinace σ_ρ is initialized so that the interval

$$\left[\rho - 2\sigma_\rho, \rho + 2\sigma_\rho\right]$$

covers a region from $d_{\min}$ and including infinite



ROBOTICS UNIVERSIDAD DE ZARAGOZA

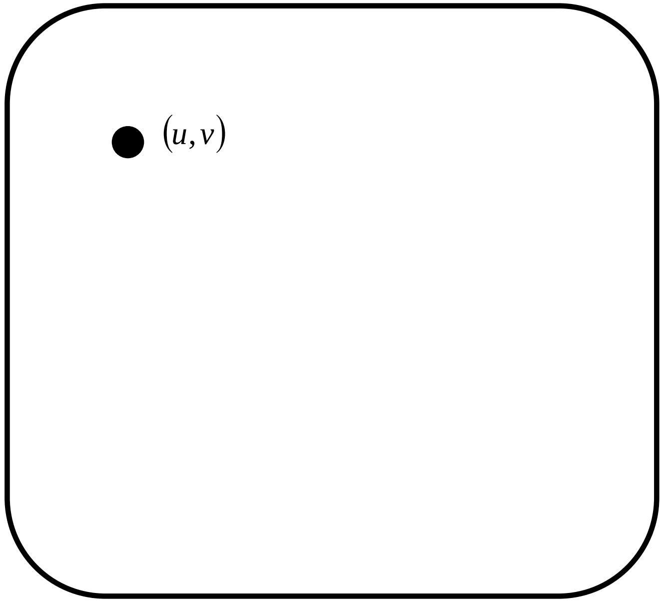
Javier Civera

J.M.M. Montiel

Imperial College London

A.J. Davison

Inverse Depth Feature Initialization



$$\begin{pmatrix} \hat{x}_i & \hat{y}_i & \hat{z}_i & \hat{\theta}_i & \hat{\phi}_i & \hat{\rho}_i \end{pmatrix}^\top$$

$$\begin{pmatrix} \hat{x}_i \\ \hat{y}_i \\ \hat{z}_i \end{pmatrix} = \hat{\mathbf{r}}^{WC}$$

$$\rho_i = \rho_0$$

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

$$\mathbf{h}^W = \mathbf{R}_{WC}\left(\mathbf{q}_{k|k}^{WC}\right)\begin{pmatrix} \upsilon \\ \nu \\ 1 \end{pmatrix}$$

$$\hat{\mathbf{x}}_v = \begin{pmatrix} \hat{\mathbf{r}}_{k|k}^{WC} \\ \hat{\mathbf{q}}_{k|k}^{WC} \\ \hat{\mathbf{V}}_{k|k}^{W} \\ \hat{\omega}_{k|k}^{C} \end{pmatrix} \begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_n \end{bmatrix} \hat{\mathbf{P}}_{k|k}$$

$$\hat{\mathbf{P}}_{k|k}^{\text{new}} = \mathbf{J}\begin{pmatrix} \hat{\mathbf{P}}_{k|k} & 0 & 0 \\ 0 & \mathbf{R}_j & 0 \\ 0 & 0 & \sigma_\rho^2 \end{pmatrix}\mathbf{J}^\top$$

Current camera and map location estimate

$$\mathbf{J} = \left(\begin{array}{c|c} I & 0 \\ \hline \frac{\partial \mathbf{y}}{\partial \mathbf{r}^{WC}}, \frac{\partial \mathbf{y}}{\partial \mathbf{q}^{WC}}, 0, \ldots, 0, & \frac{\partial \mathbf{y}}{\partial \mathbf{h}}, \frac{\partial \mathbf{y}}{\partial \rho} \end{array}\right)$$

Javier Civera

J.M.M. Montiel

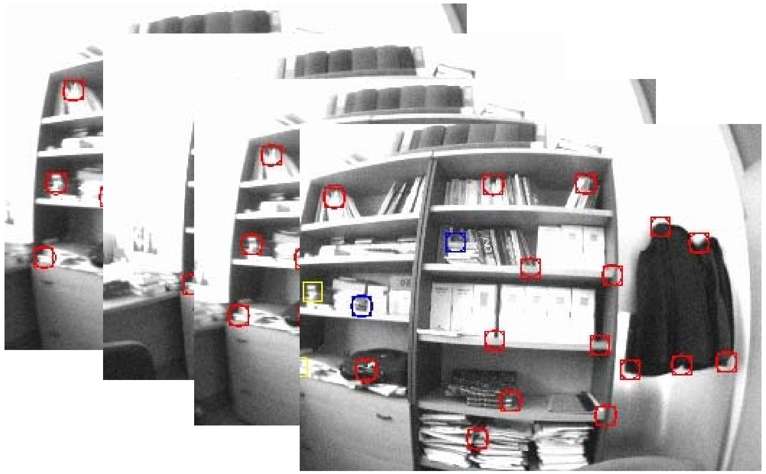
Imperial College London

A.J. Davison

Dimensional Monocular SLAM

Calibrated image measurements $\mathbf{z}_i, \sigma_z$

Frame Rate Δt



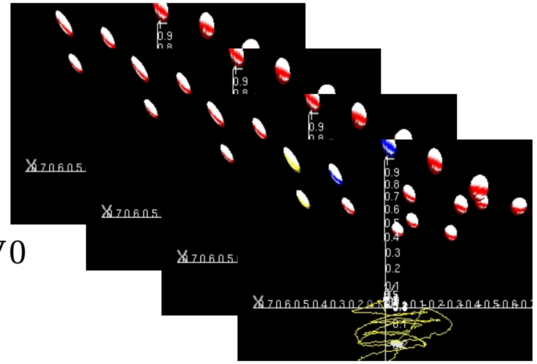
Monocular SLAM Processing

Camera Motion Priors σ_a, σ_α ; $\sigma_{\omega 0}, \sigma_{v0}$

Intial depth Prior ρ_0, σ_ρ

camera location **scene map**

$$\left(\mathbf{r}^{WC}, \mathbf{q}^{WC}, \mathbf{v}^{W}, \omega^{C}, \mathbf{y}_1 \cdots \mathbf{y}_n\right)^T$$

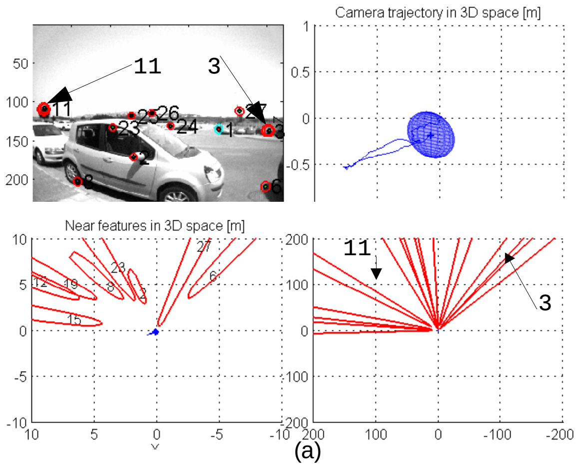


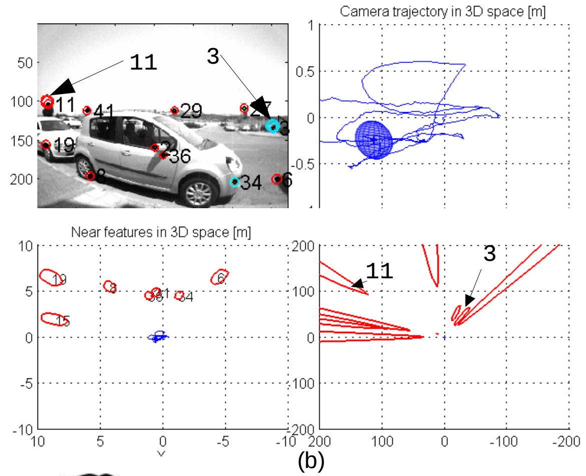
Javier Civera

J.M.M. Montiel

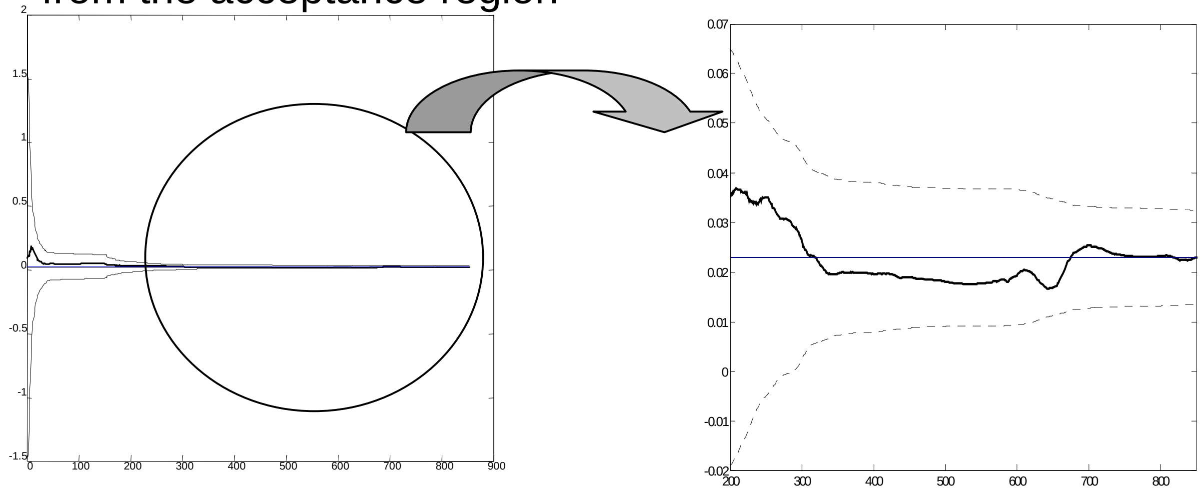
A.J. Davison

Inverse Depth Estimation History

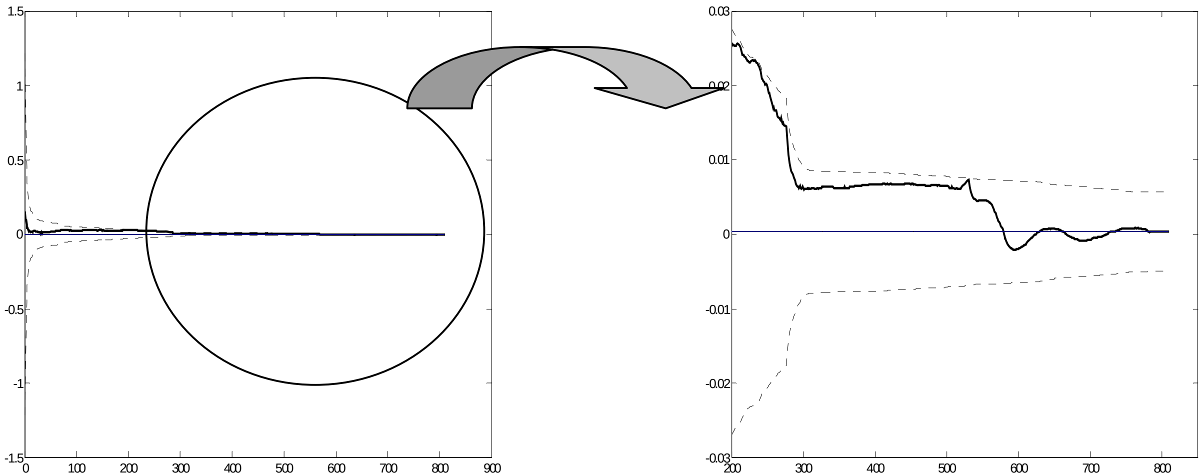




near feature (3), eventually excludes infinite
from the acceptance region



distant feature (11), infinite
always included in the acceptance region

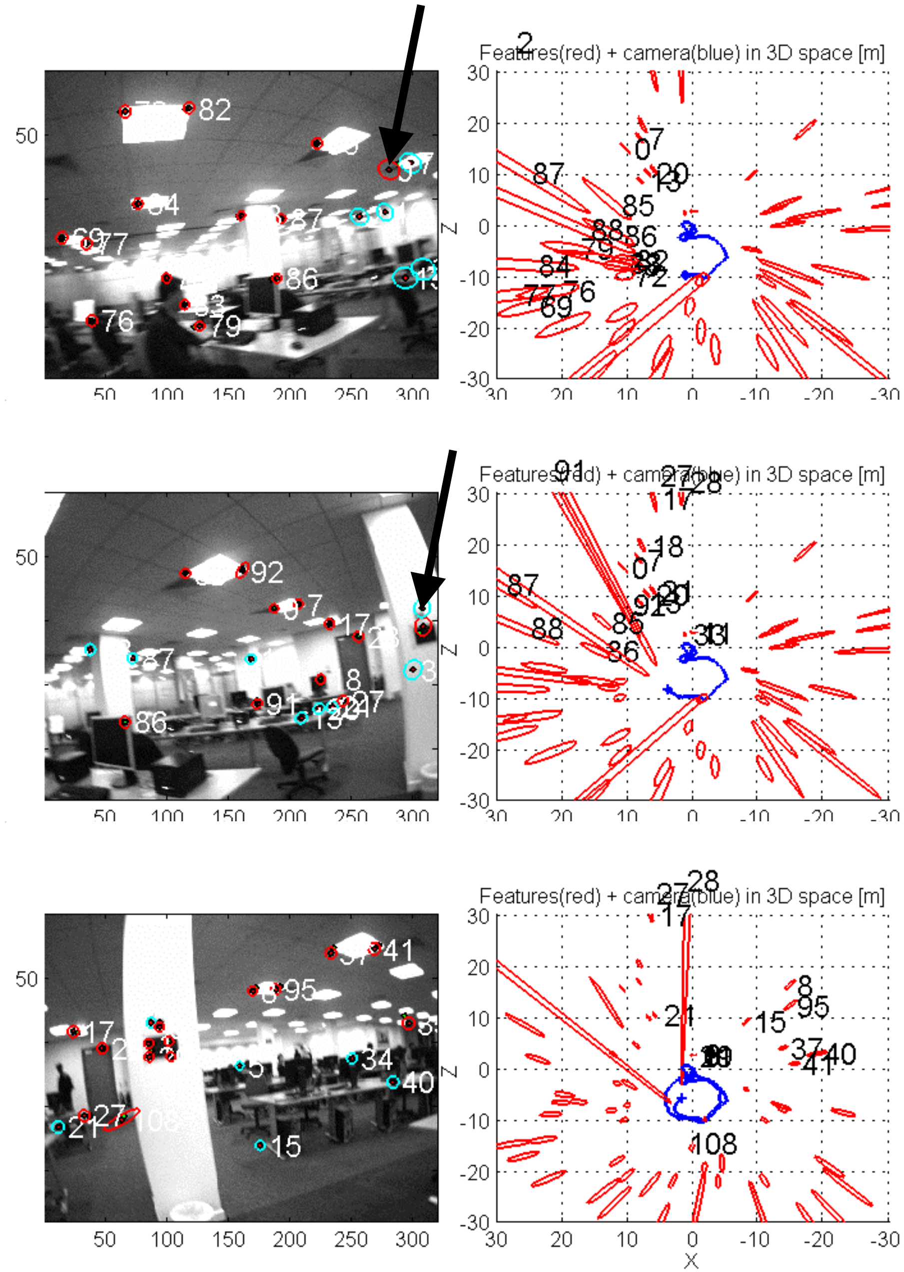


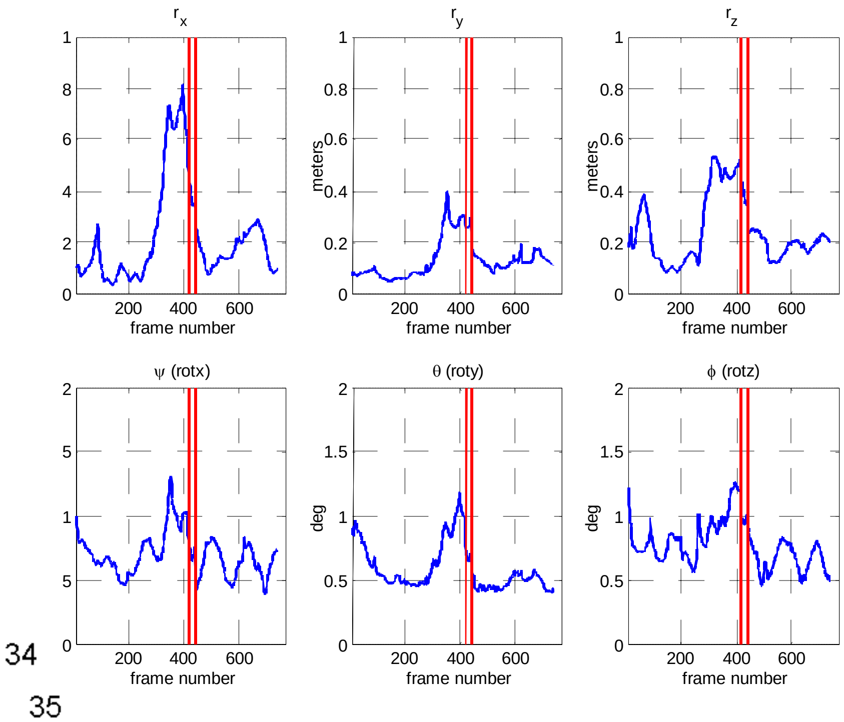
Javier Civera

J.M.M. Montiel

A.J. Davison

Loop Closing



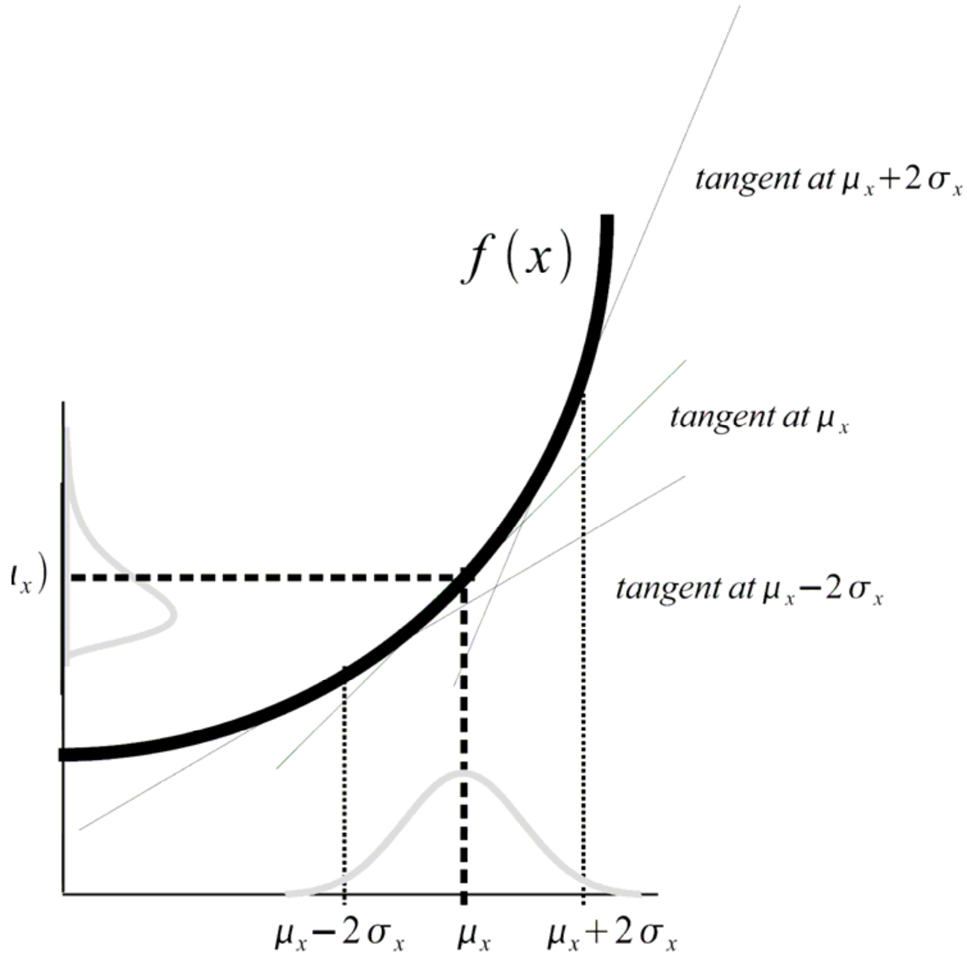


Imperial College London

J.M.M. Montiel

A.J. Davison

Linearity Index



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

ROBOTICS UNIVERSIDAD DE ZARAGOZA

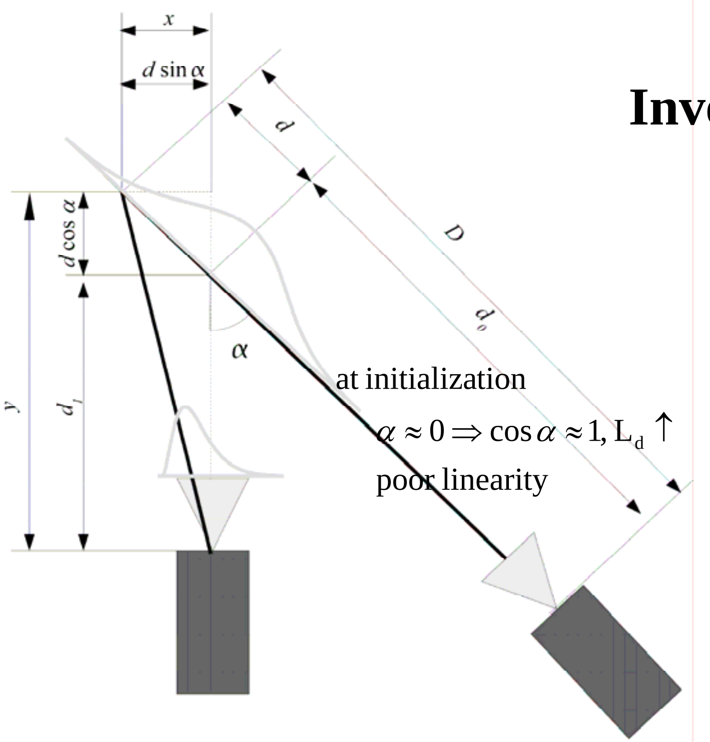
Javier Civera

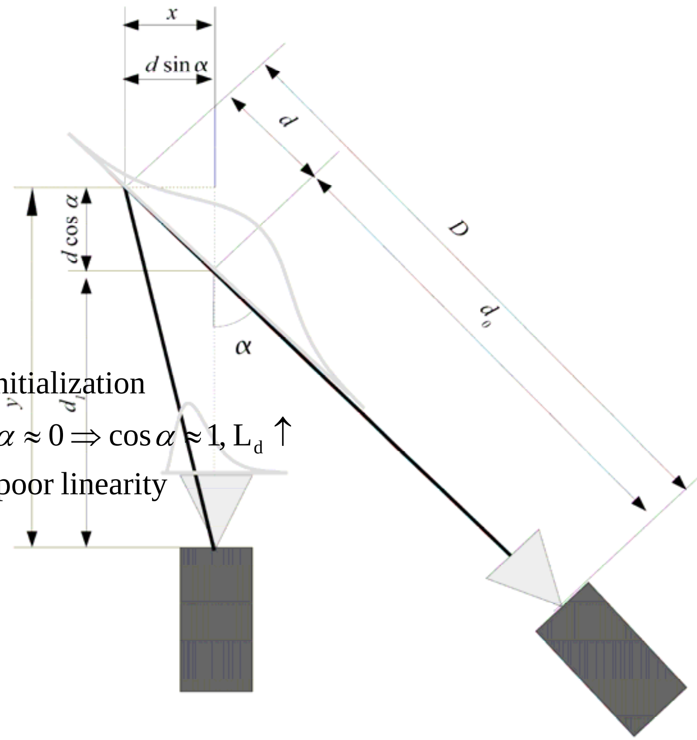
J.M.M. Montiel

Imperial College London

A.J. Davison

Inverse depth linearity analysis





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

at initialization

$\alpha \approx 0 \Rightarrow \cos\alpha \approx 1, \mathrm{L_d} \uparrow$

poor linearity

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

at initialization

$\alpha \approx 0 \Rightarrow 1 - \cos\alpha \approx 0, \mathrm{L}_\rho \downarrow$

linearity

after parallax gathering,

$1 - \cos\alpha$, but σ_ρ

linearity

good performance along the whole estimation

ROBOTICS UNIVERSIDAD DE ZARAGOZA

Javier Civera

J.M.M. Montiel

Imperial College London

A.J. Davison

Inverse depth to XYZ conversion

- Inverse depth good performance along the whole estimation process
- Inverse depth coding needs 6 parameters
- XYZ coding good performance for reduced depth uncertainty
- So, it is not mandatory to switch from inverse depth to XYZ, but computing overhead can be reduce
- Switching criteria based on the linearity index

$$L_d = \frac{4\sigma_d}{d_1} \left|\cos\alpha\right| < 10\%$$

$$\begin{aligned} d_i &= \left\| \mathbf{h}^C \right\|, \quad \mathbf{h}^C = \mathbf{x}_i - \mathbf{r}^{WC} \\ \sigma_d &= \frac{\sigma_\rho}{\rho_i^2}, \quad \sigma_\rho = \sqrt{\mathbf{P}_{\mathbf{y}_i \mathbf{y}_i}(6,6)} \\ \cos\alpha &= \mathbf{m}^\top \mathbf{h}^C \left\| \mathbf{h}^C \right\|^{-1}. \end{aligned}$$

$$\mathbf{x}_i = \begin{pmatrix} X_i \\ Y_i \\ Z_i \end{pmatrix} = \begin{pmatrix} x_i \\ y_i \\ z_i \end{pmatrix} + \frac{1}{\rho_i}\mathbf{m}(\theta_i, \phi_i) \qquad \mathbf{P}_{\text{new}} = \mathbf{J}\mathbf{P}\mathbf{J}^\top, \quad \mathbf{J} = \text{diag}\left(\mathbf{I}, \frac{\partial \mathbf{x}_i}{\partial \mathbf{y}_i}, \mathbf{I}\right)$$

ROBOTICS UNIVERSIDAD DE ZARAGOZA

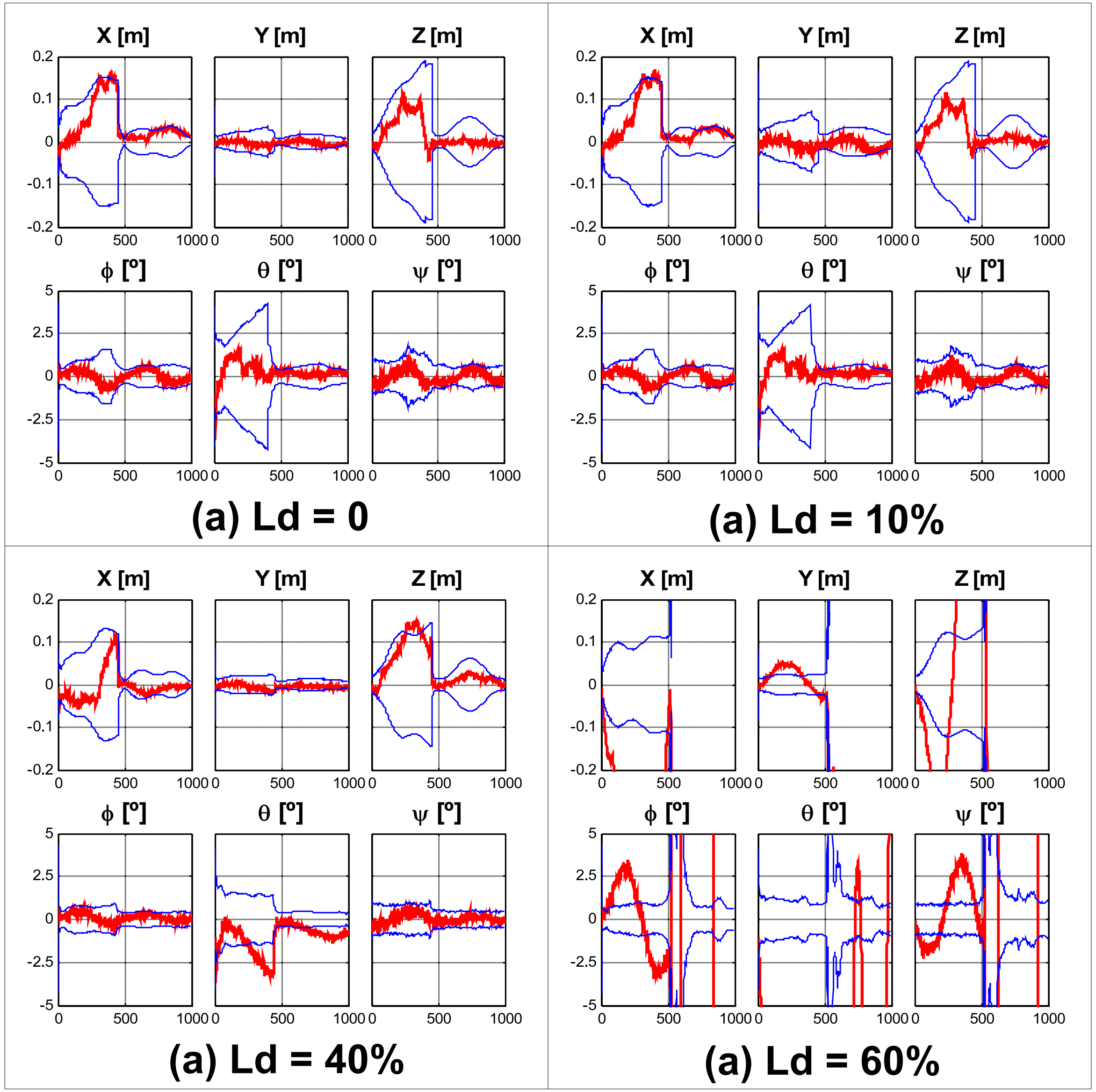
Javier Civera

J.M.M. Montiel

Imperial College London

A.J. Davison

Inverse depth to XYZ conversion threshold

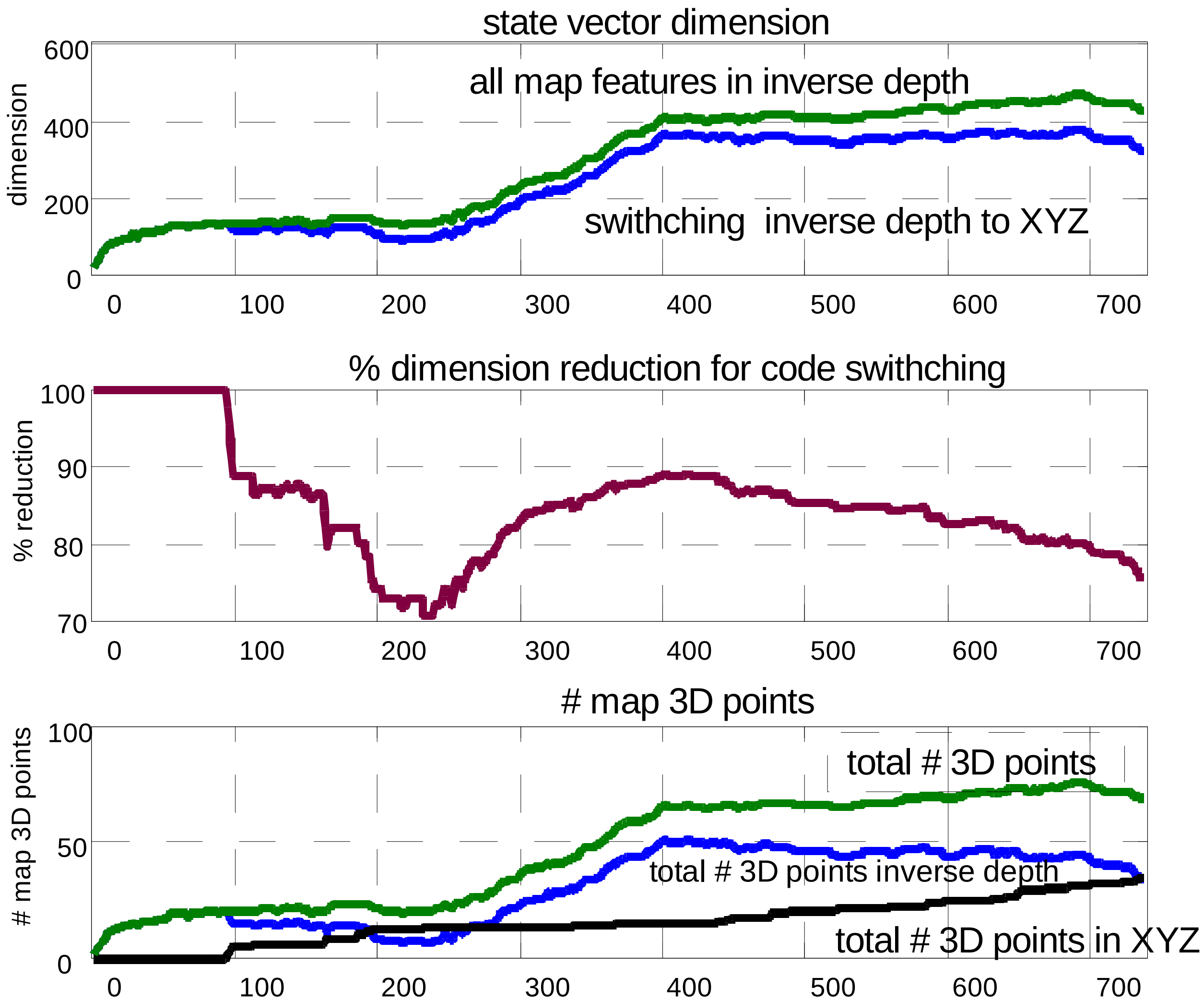


Camera location error   95% acceptance error

Javier Civera
J.M.M. Montiel

A.J. Davison

Switching evolution



Javier Civera
J.M.M. Montiel

Imperial College London

A.J. Davison

Switching evolution

☆ *Inverse depth coding*

△ *Depth coding*

ROBOTICS UNIVERSIDAD DE ZARAGOZA

Javier Civera

J.M.M. Montiel

Imperial College London

A.J. Davison

Bibliografía

Davison, AJ , Reid, ID, Molton, ND , Stasse, O :" MonoSLAM: Real-time single camera SLAM" IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, 29 (6): 1052-1067 JUN 2007

J.M.M. Montiel, Javier Civera y Andrew J. Davison: "Unified Inverse Depth Parametrization for Monocular SLAM". Robotics: Science and Systems Conference 2006.

Javier Civera , Andrew J. Davison and J.M.M. Montiel,: "Inverse Depth to Depth Conversion for Monocular SLAM". IEEE Int Conf on Robotics and Automation Rome, April 2007.

Javier Civera, Andrew J. Davison, J. M. M. Montiel. Dimensionless Monocular SLAM. 3rd Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA), Girona, 2007

J.M.M Montiel home page: http://webdiis.unizar.es/~josemari/

Anderw Davison home page http://www.doc.ic.ac.uk/~ajd/