

# Inverse Depth Monocular SLAM

Javier Civera, Andrew J. Davison, JMM Montiel

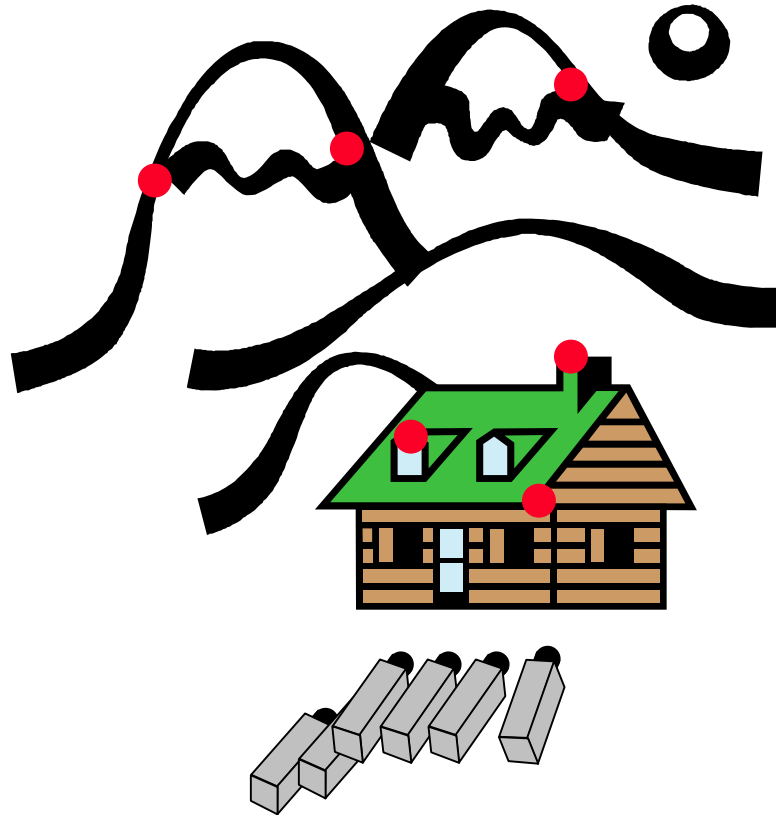


Javier Civera  
JMM Montiel

Imperial College  
London

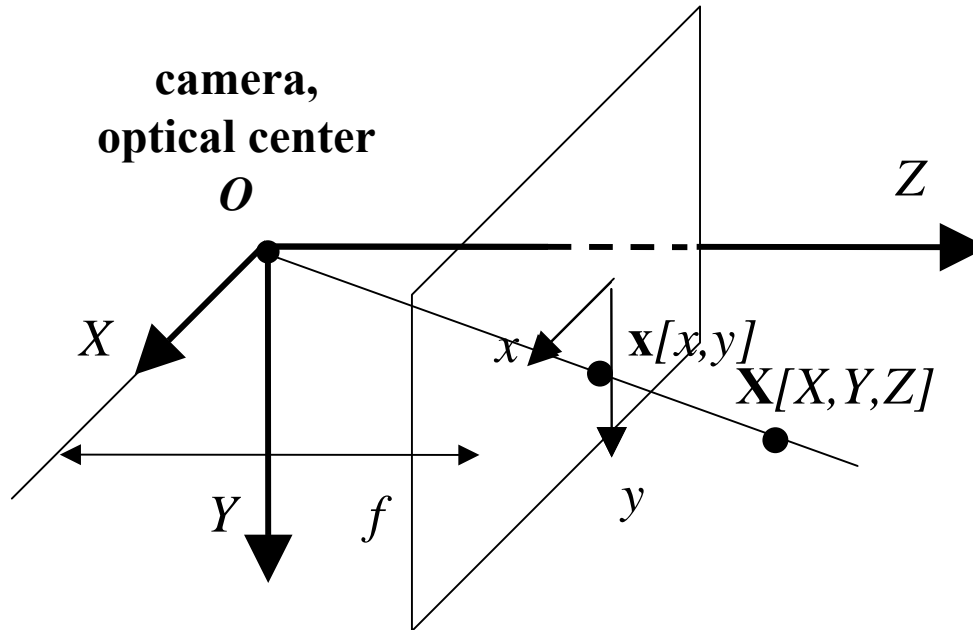
A.J. Davison

# 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**

# Camera Geometry: Pure Bearing-only Sensor



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

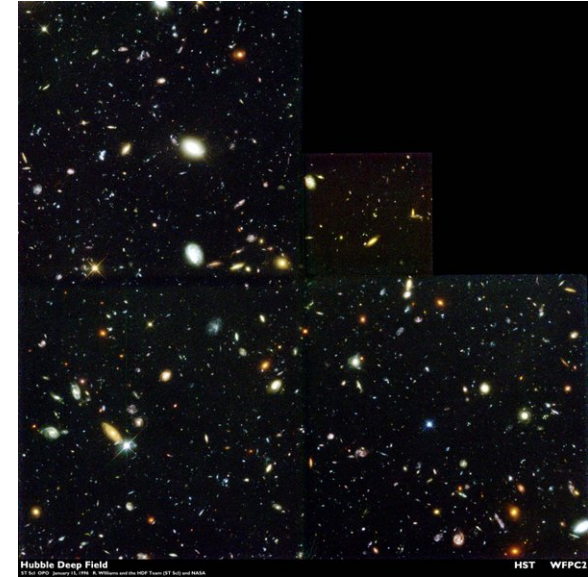


image of galaxies  
at  $10^9$  years light far from  
the Earth

# 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



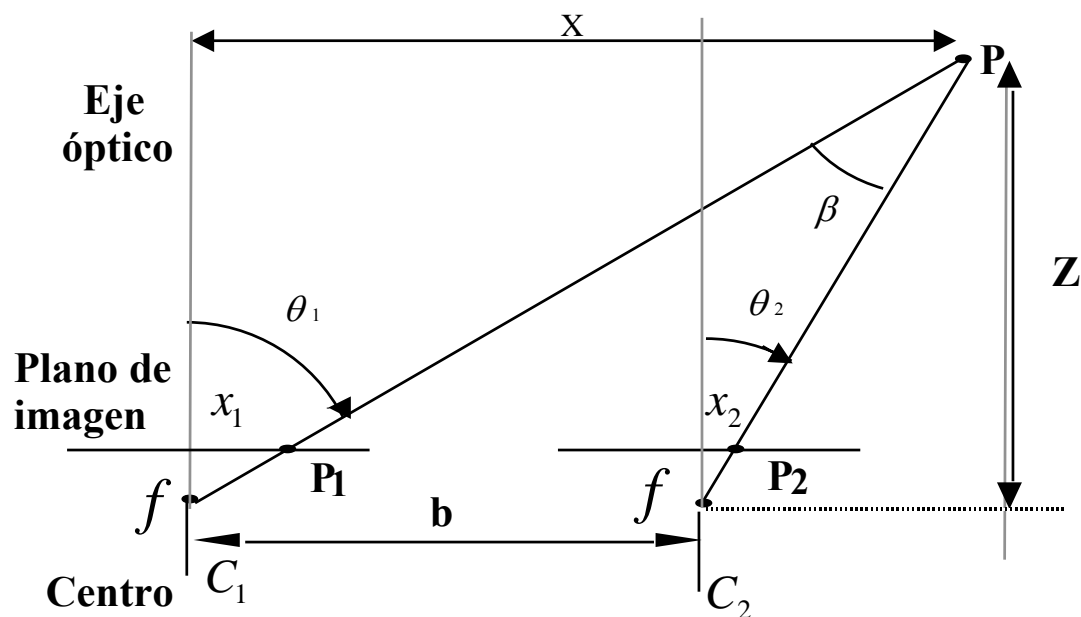
# Parallax



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



# Stereo Vision. Sensitivity Analysis



Geometry, depending only on:

- relative camera location
- observed point location

Parallax, angle defined by the two rays corresponding to the two cameras for a scene point.

$$\text{ray 1, } x = z \tan \theta_1$$

$$\text{ray 2, } x = z \tan \theta_2 + b$$

$$z = \frac{b}{\tan \theta_1 - \tan \theta_2} \quad \tan \theta_1 \approx \theta_1 \quad \tan \theta_2 \approx \theta_2 \quad z \approx \frac{b}{\theta_1 - \theta_2}$$

The three angles in triangle \$C\_1PC\_2\$ add up to \$\pi\$ rad

$$z \approx \frac{b}{\beta} \quad \beta \text{ Parallax angle}$$



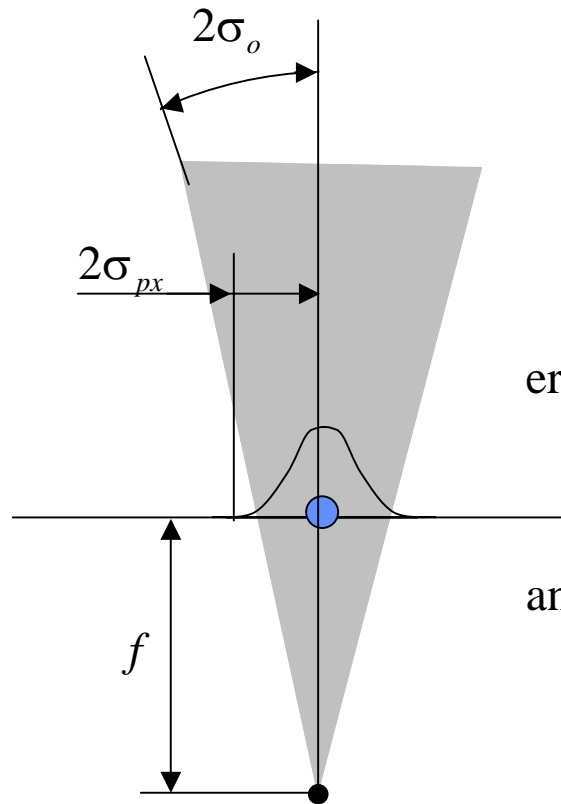
# Image errors

$\sigma_{px}$ , error standard deviation in pixels

as a rule of thumb error range is  $\pm 2\sigma_{px}$

example, for a clicking error  $\pm 1$  pixel,

corresponding standard deviation  $\sigma_{px} = 0.5$  pixels



error analysis is based on the ray orientation error

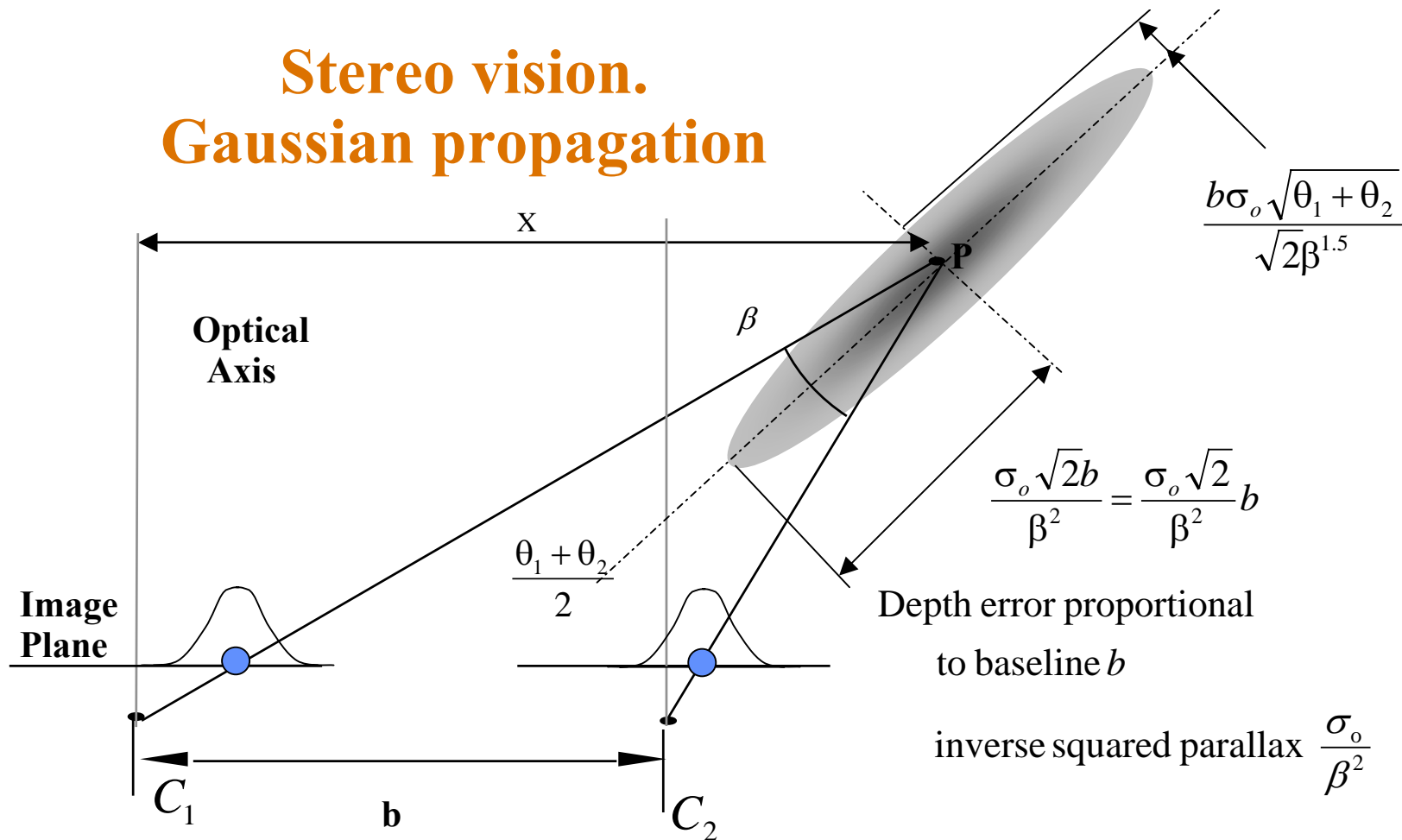
the standard deviation for the orientation error,  $\sigma_o$ , in radians

an approximate relation between them

$$\sigma_o \cong \frac{d}{f} \sigma_{px}, \quad d, \text{ pixel size (mm.)}$$

$f$ , lens focal length (mm.)

# Stereo vision. Gaussian propagation



Ellipsoid with major axis oriented in the two rays bisectric direction.

Error propagation on X, Z direction depends on two rays direction.

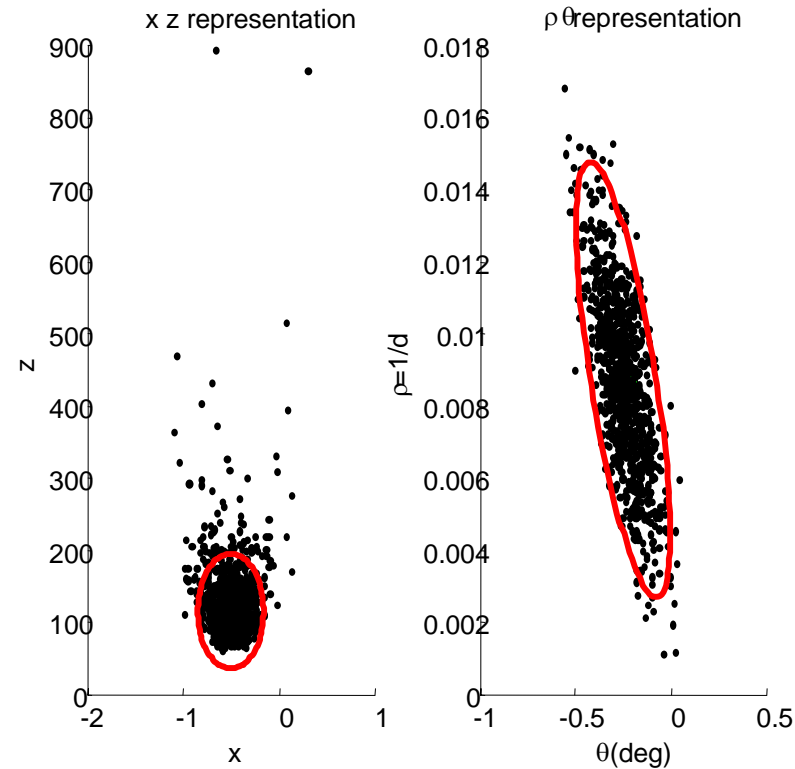
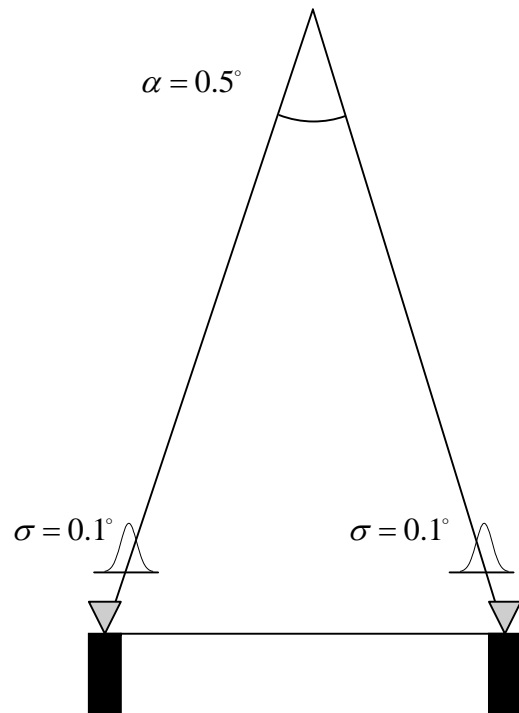
Depth error proportional to baseline  $b$

inverse squared parallax  $\frac{\sigma_o}{\beta^2}$

Lateral error smaller than depth error, proportional to baseline  $b$

inverse parallax to the 1.5th power  $\frac{\sigma_o}{\beta^{1.5}}$

# Linearity limits of the Gaussian propagation. 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  
JMM Montiel

Imperial College  
London

A.J. Davison

# State of the art I

- **SLAM, initially proposed by Smith and Cheesman, 1986, widespread usage in robotics for multisensor fusion [Castellanos 1999], [Feder 1999] [Thrun et al. 2005]**
  - 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

SFM, computer vision methods



## State of the art II

### 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 1998, 1999 ]**
  - No parallax, homography model
  - Parallax epipolar geometry model
  - Increasing the frame rate, the interframe motion closes to a homography



# State of the art III

## Feature initialization in Monocular SLAM

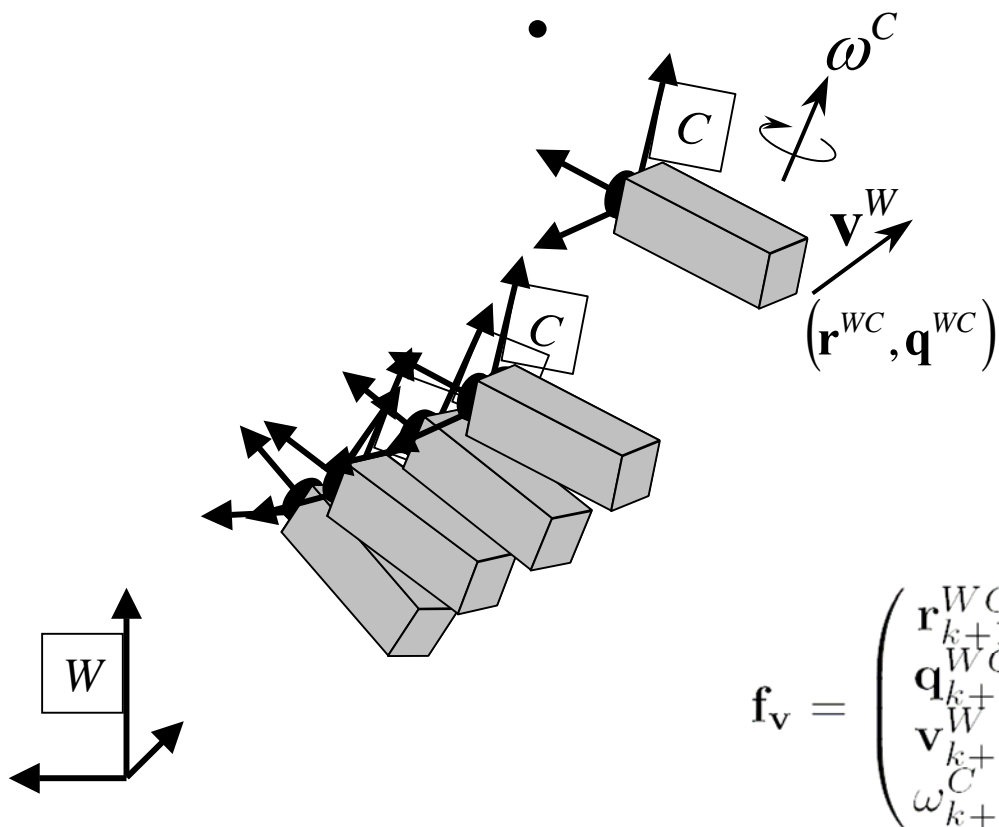
- **Delayed approach: Image tracking until safe Gaussian triangulation**  
Bailey 2003, Davison 2003
- **Undelayed initialization**
  - Multiple hypotheses in depth, Kwok Dissanayake 2004, Sola et al. 2005,

## Inverse depth usage

- **Concept used in different domains**
  - Parallax with respect plane at infinity, Zisserman 2003
  - Optical flow Heeger & Jepson 1992
  - Modified polar coordinates in bearing only TMA Aidala & Hammel 1983
  - Sequential structure estimation from known motion Okutomi&Kanade 1993.
  - Pairwise first and individual EKF's Chowdhury&Chellappa 2003
- **Recently used for Monocular SLAM**
  - Trwany & Roumeliotis 2005
  - Eade & Drummond 2006, FastSLAM, a distinguished initialization stage.



# Camera motion priors



Constant velocity motion model:

- Smooth camera motion
- Impulse acceleration noise

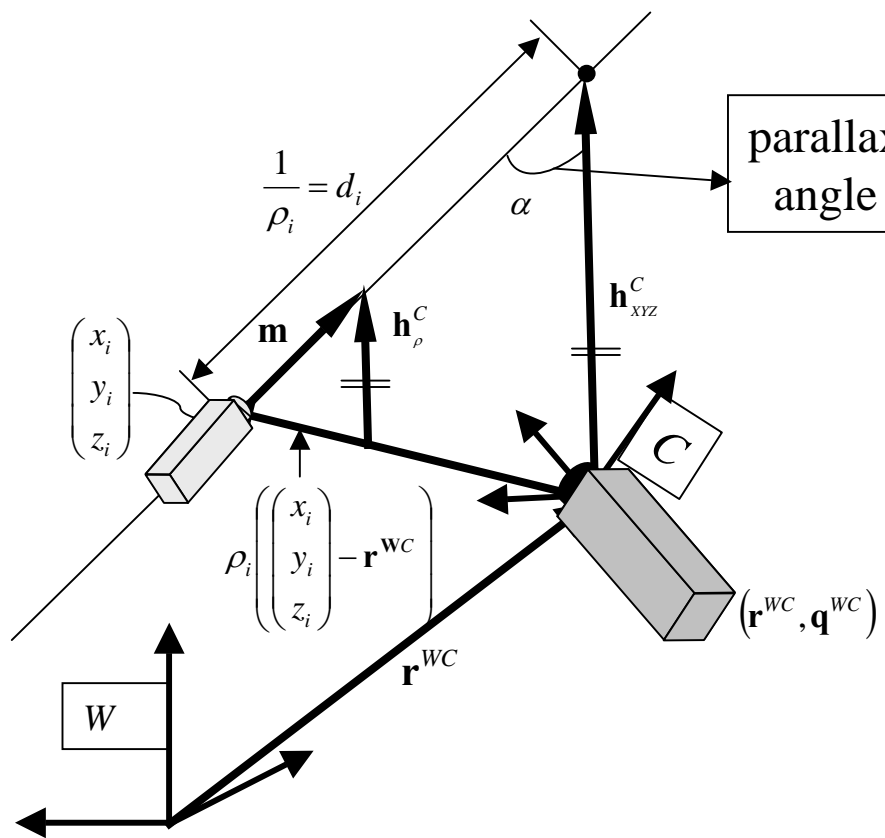
$$\mathbf{n} = \begin{pmatrix} \mathbf{V}^W \\ \Omega^W \end{pmatrix} = \begin{pmatrix} \mathbf{a}^W \Delta t \\ \alpha^W \Delta t \end{pmatrix}$$

$$\mathbf{f}_v = \begin{pmatrix} \mathbf{r}_{k+1}^{WC} \\ \mathbf{q}_{k+1}^{WC} \\ \mathbf{v}_{k+1}^W \\ \omega_{k+1}^C \end{pmatrix} = \begin{pmatrix} \mathbf{r}_k^{WC} + v_k^W \Delta t + a_k^W \Delta t^2 \\ \mathbf{q}_k^{WC} \times \mathbf{q}(\omega_k^C \Delta t + \alpha_k^C \Delta t^2) \\ \mathbf{v}_k^W + a_k^W \Delta t \\ \omega_k^C + \alpha_k^C \Delta t \end{pmatrix}$$

# Scene point coding in inverse depth.

## Measurement equation

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



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

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

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

distant point,  $\rho_i \rightarrow 0, \Rightarrow$  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

at low parallax, a point is observed as

$$\mathbf{h}_\rho^C = \mathbf{R}^{CW} \mathbf{m}(\theta_i, \phi_i)$$

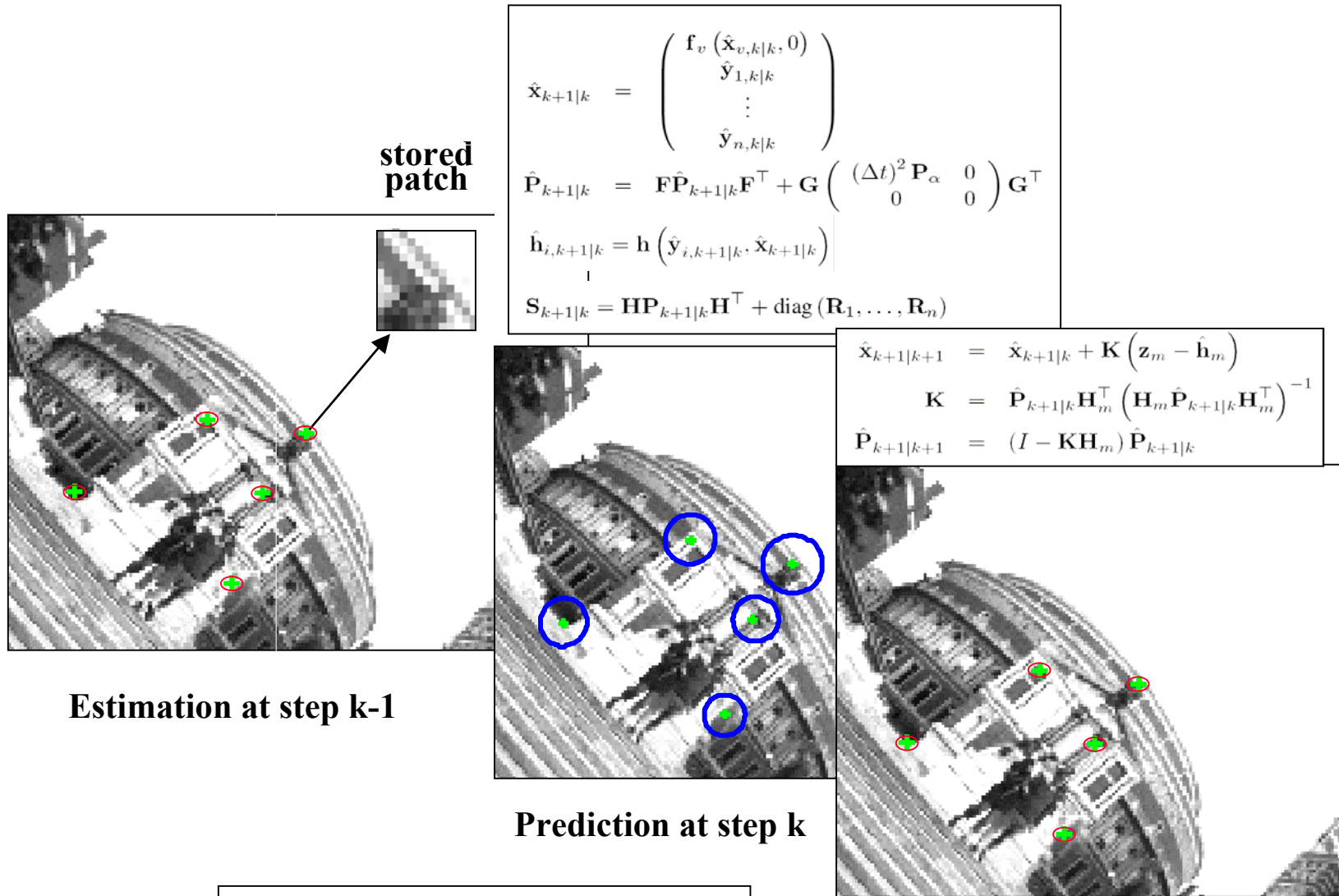
## SLAM joint map+camera state vector

The full estimate coded in a unique Gaussian distribution

1. Camera state vector.
2. *ALL* the map features.
3. The joint Gaussian has proven to be useful in coding strong correlation between the observations.

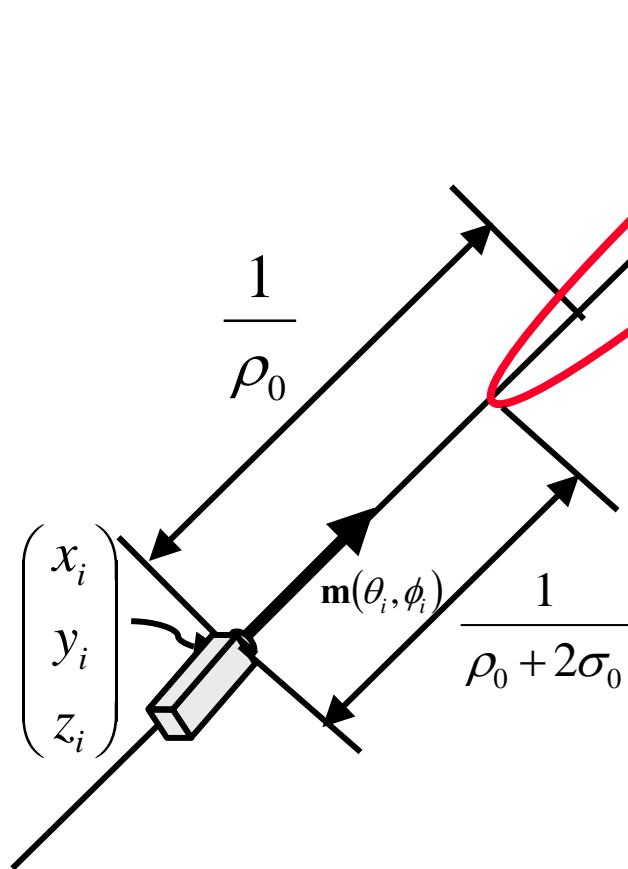
$$\mathbf{x} = \left( \mathbf{x}_v^T \quad \mathbf{y}_1^T \quad \cdots \quad \mathbf{y}_n^T \right)^T$$
$$\mathbf{P} = \begin{bmatrix} \mathbf{P}_{\mathbf{x}_v \mathbf{x}_v} & \mathbf{P}_{\mathbf{x}_v \mathbf{y}_1} & \cdots & \mathbf{P}_{\mathbf{x}_v \mathbf{y}_n} \\ \mathbf{P}_{\mathbf{x}_v \mathbf{y}_1}^T & \mathbf{P}_{\mathbf{y}_1 \mathbf{y}_1} & \cdots & \mathbf{P}_{\mathbf{y}_1 \mathbf{y}_n} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{P}_{\mathbf{x}_v \mathbf{y}_n}^T & \mathbf{P}_{\mathbf{y}_1 \mathbf{y}_n}^T & \cdots & \mathbf{P}_{\mathbf{y}_n \mathbf{y}_n} \end{bmatrix}$$

# Active search

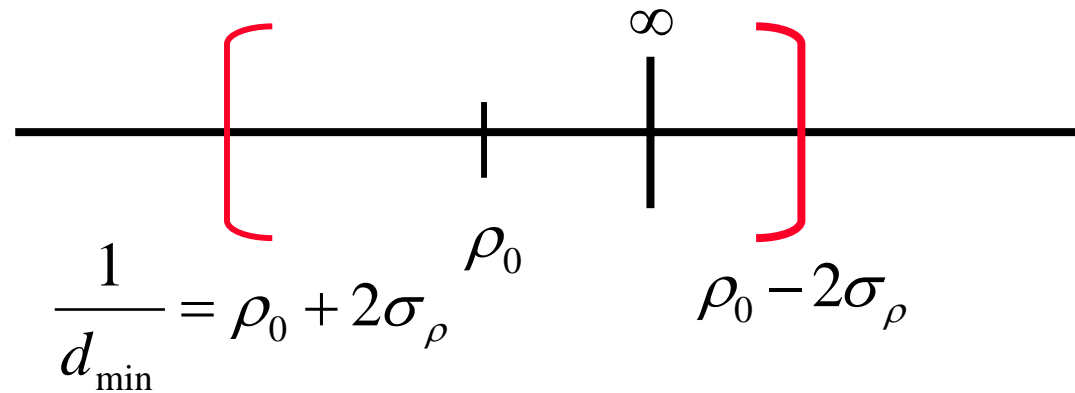


stored patch is searched in all acceptance the region

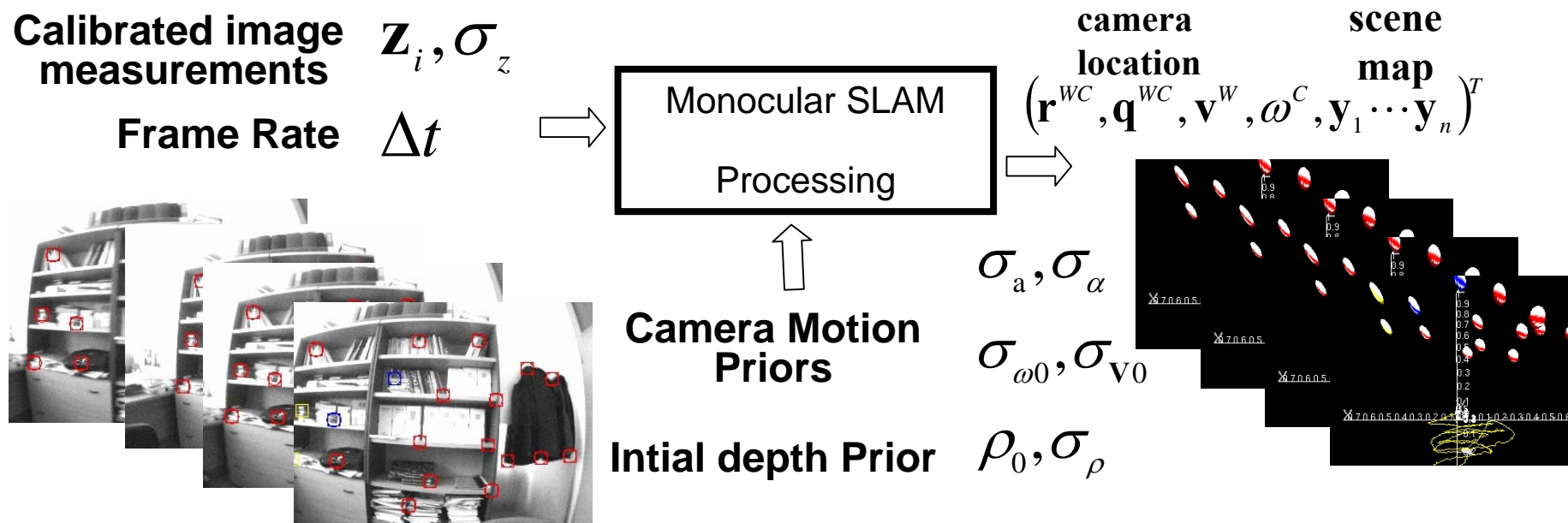
## Inverse depth feature initialization prior



New points are observed from just an observation  $x, y, z, \theta, \phi$  and the corresponding covariance are initialized from the first feature observation  $\rho$  at  $\rho_0$  and its covariance  $\sigma_\rho$  is initialized so that the interval  $[\rho - 2\sigma_\rho, \rho + 2\sigma_\rho]$  covers a region from  $d_{\min}$  and including infinite



# Dimensional Monocular SLAM



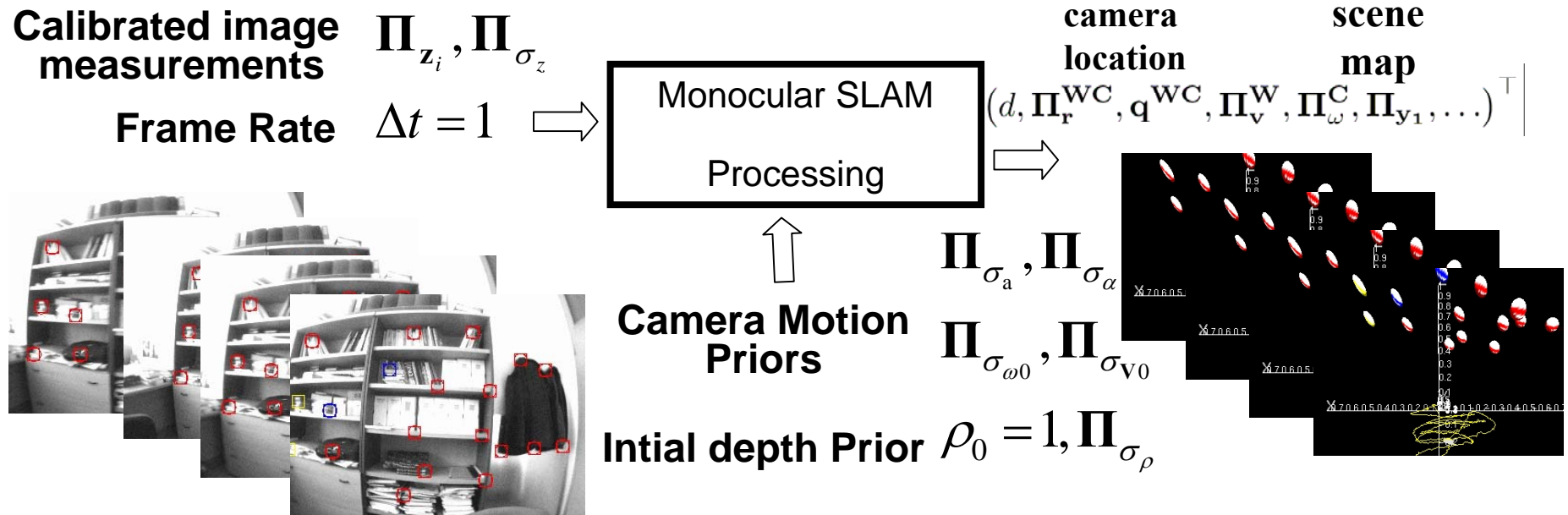
So estimation process can be considered as a function

$$(\mathbf{r}^{WC}, \mathbf{q}^{WC}, \mathbf{v}^W, \omega^W, \mathbf{y}_1, \dots, \mathbf{y}_n)^T = \mathbf{f}(\sigma_a, \sigma_\alpha, \sigma_z, \mathbf{z}, \Delta t, \rho_0, \sigma_{\rho_0}, \sigma_{v0}, \sigma_{\omega 0})$$

Being the dimensions involved, time, and length:

$\mathbf{r}$	$\mathbf{q}$	$\mathbf{v}, \sigma_{v0}$	$\omega, \sigma_{\omega 0}$	$\mathbf{z}, \sigma_z$	$\sigma_a$	$\sigma_\alpha$	$x_i, y_i, z_i$	$\theta_i, \phi_i$	$\rho_i, \rho_0, \sigma_{\rho 0}$
$l$	1	$lt^{-1}$	$t^{-1}$	1	$lt^{-2}$	$t^{-2}$	$l$	1	$l^{-1}$

# Dimensionless Monocular SLAM



$\Delta t, \rho_0$  are selected to define the dimensionless coefficients [Buckingham 1914]:

$\Pi_r$	$\Pi_q$	$\Pi_v$	$\Pi_\omega$	$\Pi_{\rho_i}$	$\Pi_{\sigma_{v0}}$	$\Pi_{\sigma_{\omega 0}}$	$\Pi_{\sigma_{\rho 0}}$	$\Pi_z$	$\Pi_{\sigma_z}$	$\Pi_{\sigma_a}$	$\Pi_{\sigma_\alpha}$
$r\rho_0$	$q$	$v\rho_0\Delta t$	$\omega\Delta t$	$\frac{\rho_i}{\rho_0}$	$\sigma_{v0}\rho_0\Delta t$	$\sigma_{\omega 0}\Delta t$	$\frac{\sigma_{\rho 0}}{\rho_0}$	$z$	$\sigma_z$	$\sigma_a\rho_0\Delta t^2$	$\sigma_\alpha\rho_0\Delta t^2$

state vector split in scale,  $d$ , and dimensionless coefficients

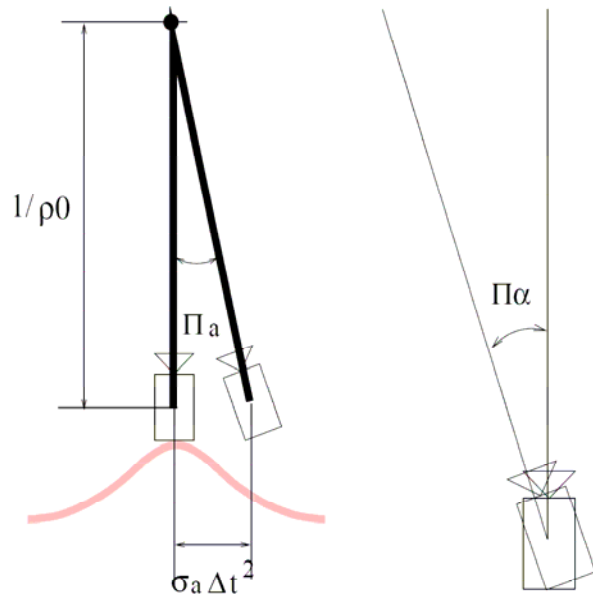
$$(d, \Pi_r^{WC}, q^{WC}, \Pi_v^W, \Pi_\omega^C, \Pi_{y_1}, \dots)^T$$

monocular camera cannot observe scale.  $d$ . If scale were known:

$$r^{WC} = d\Pi_r^{WC}, \quad v^W = d\Pi_v^W \Delta t, \quad \omega^W = d\Pi_\omega^W \Delta t$$

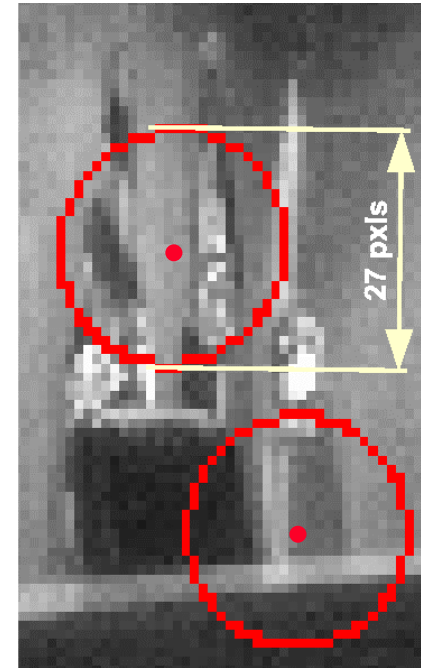
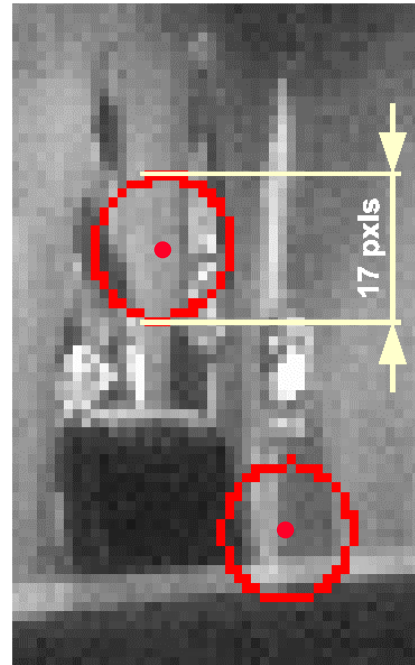
$$y_i = (d\Pi_{x_i}, d\Pi_{y_i}, d\Pi_{z_i}, \theta_i, \phi_i, \Pi_{\rho_i}/d)$$

# Interpreting Dimensionless Parameters as Image Quantities



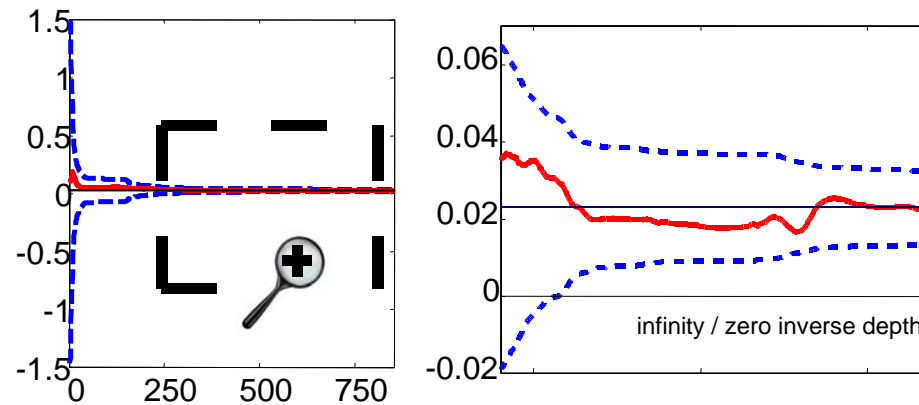
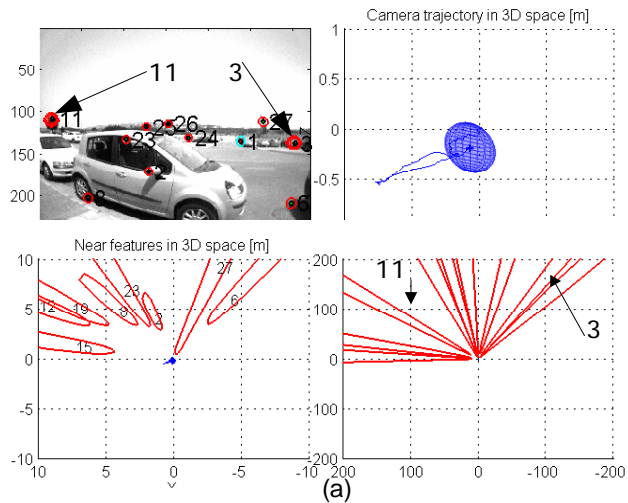
dimensionless  
camera linear  
acceleration  
standard  
deviation

dimensionless  
camera  
angular  
acceleration  
standard  
deviation

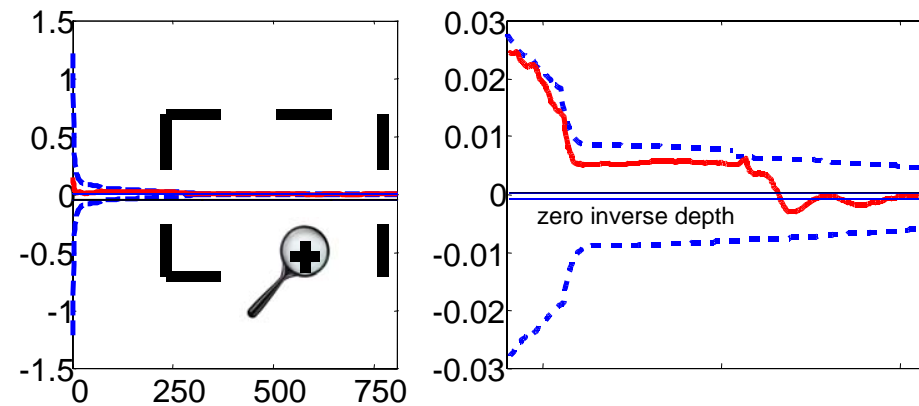
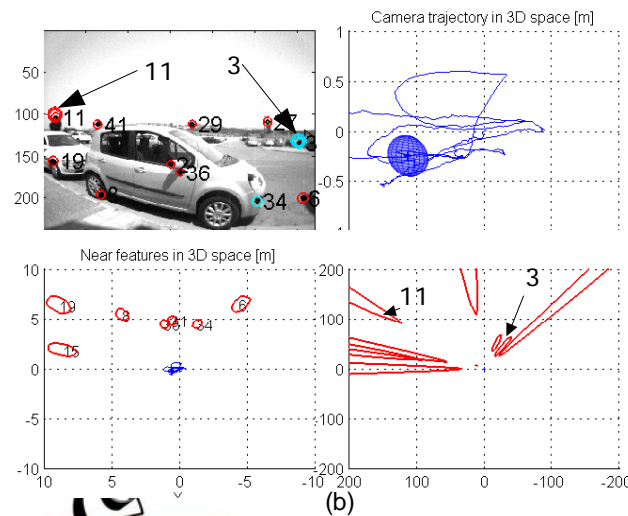


# Inverse depth estimation history

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



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

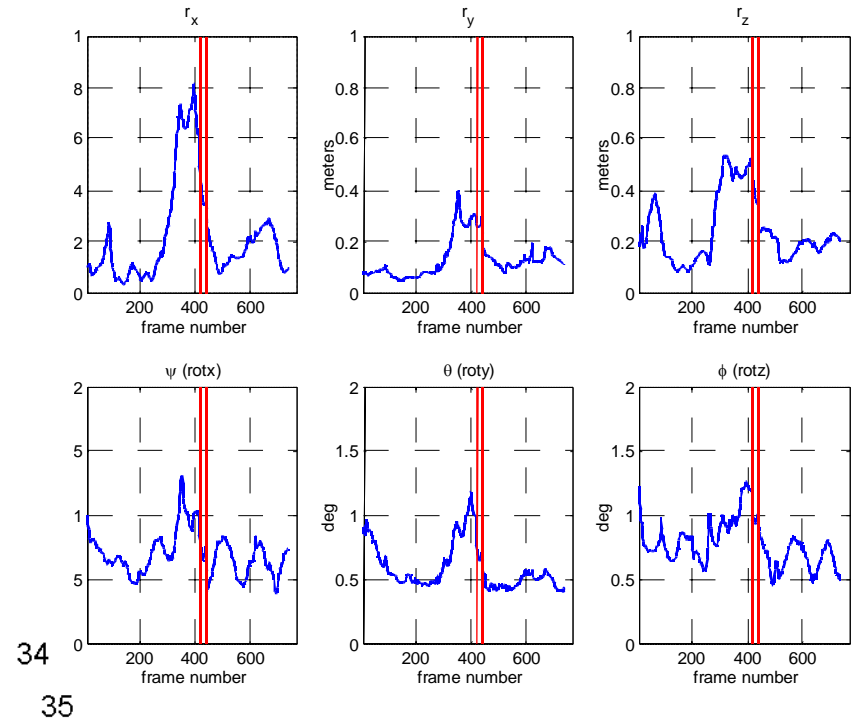
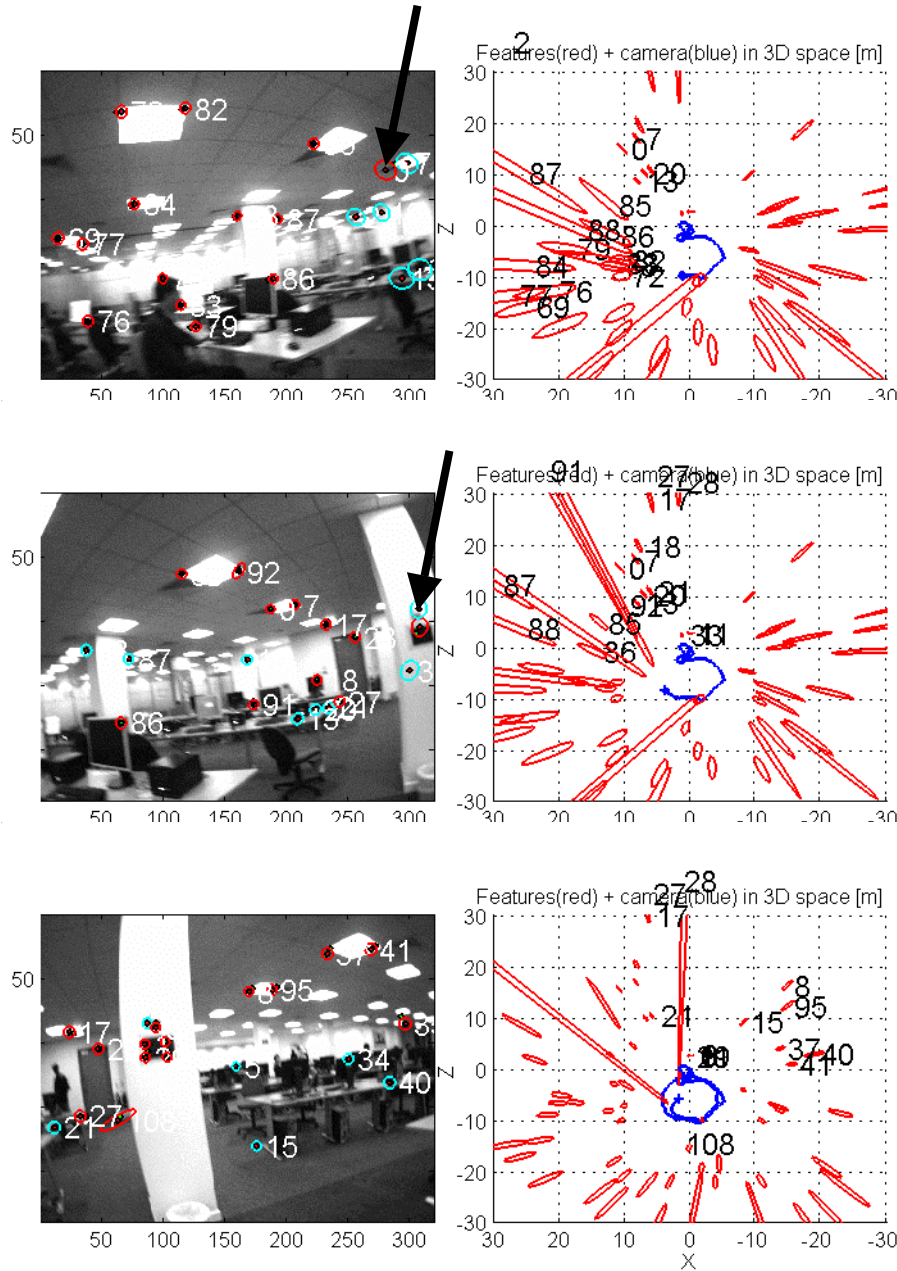


Javier Civera  
JMM Montiel

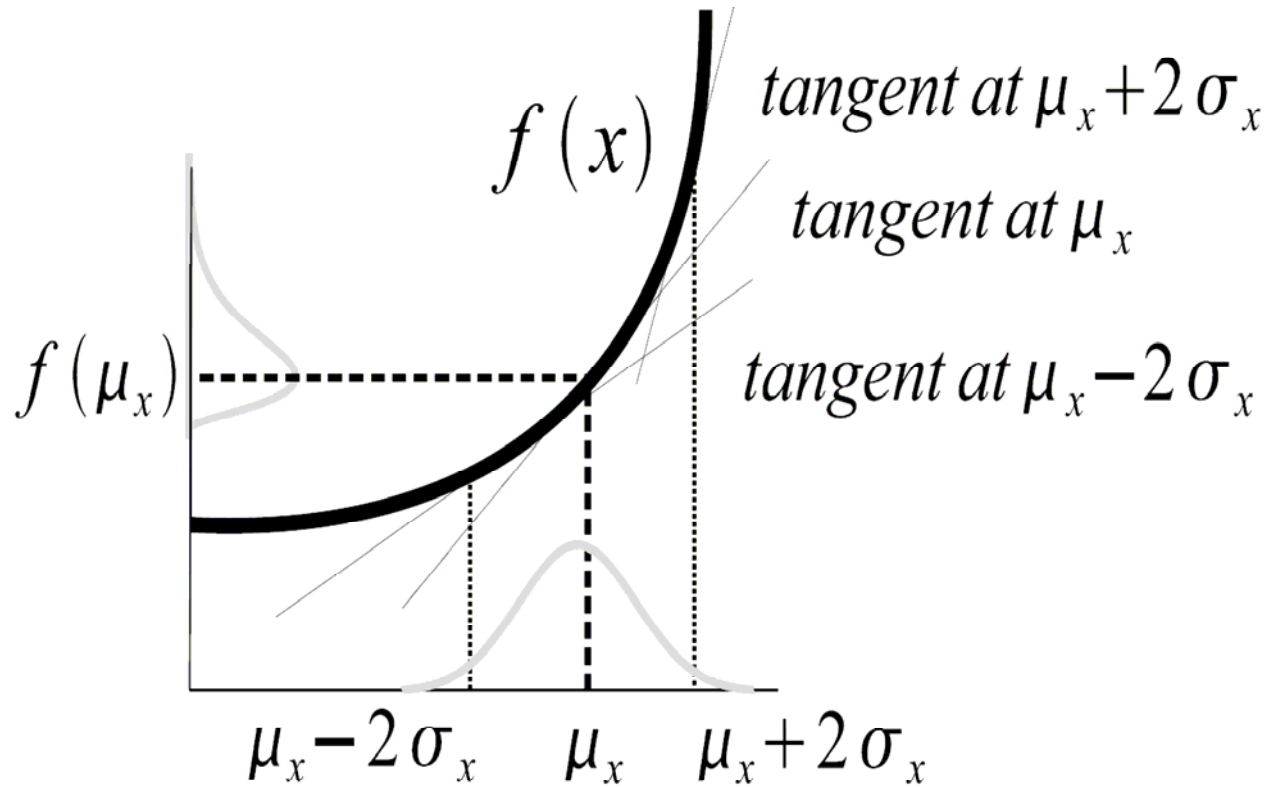
Imperial College  
London

A.J. Davison

# Loop Closing

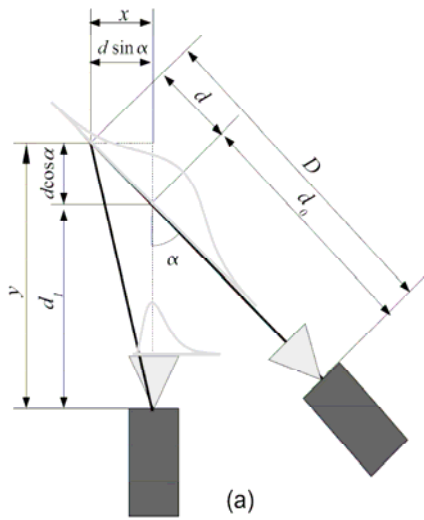


## Linearity Index



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

# Inverse depth linearity analysis

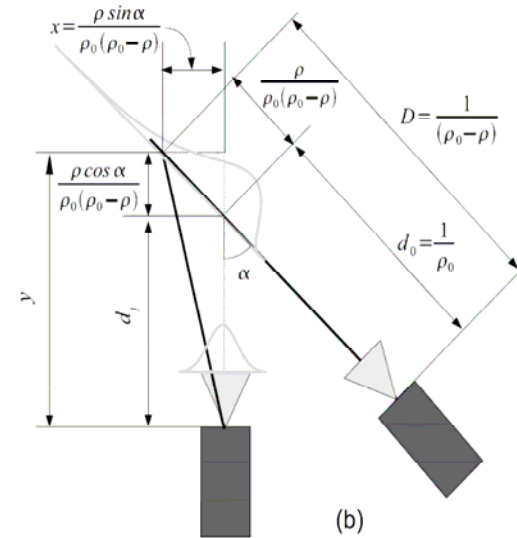


$$L_d = \frac{4\sigma_d}{d_1} |\cos\alpha|$$

at initialization

$$\alpha \approx 0 \Rightarrow \cos\alpha \approx 1, L_d \uparrow$$

poor linearity



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

at initialization

$$\alpha \approx 0 \Rightarrow 1 - \cos\alpha \approx 0, L_\rho \downarrow$$

linearity

after parallax gathering,

$$1 - \cos\alpha, \text{ but } \sigma_\rho$$

linearity

good performance along the whole estimation

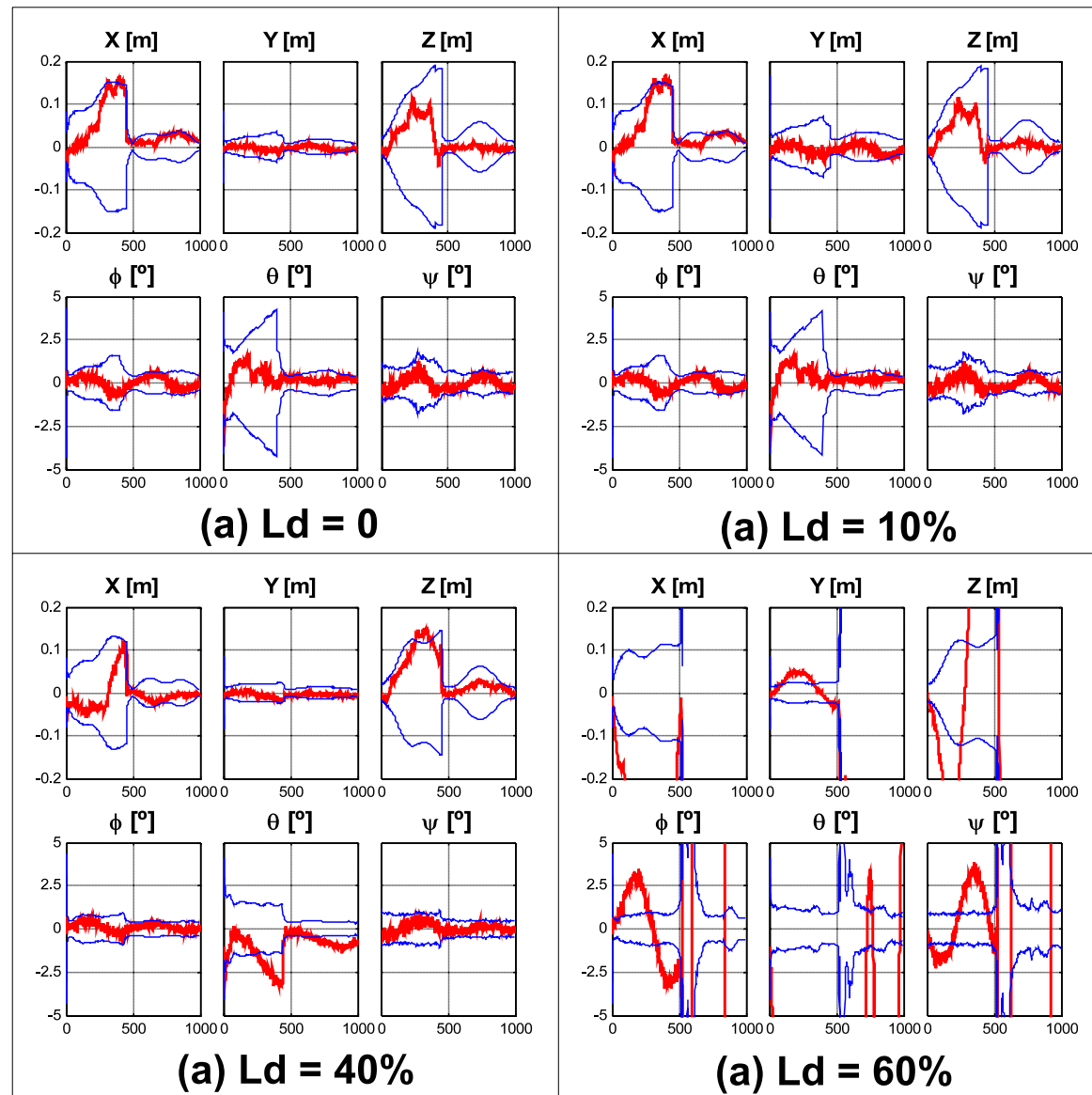
## Inverse depth to XYZ conversion

- 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

$$L_d = \frac{4\sigma_d}{d_1} |\cos \alpha| < 10\% \quad \left| \begin{array}{l} d_i = \|\mathbf{h}^C\|, \quad \mathbf{h}^C = \mathbf{x}_i - \mathbf{r}^{WC} \\ \sigma_d = \frac{\sigma_\rho}{\rho_i^2}, \quad \sigma_\rho = \sqrt{\mathbf{P}_{\mathbf{y}_i \mathbf{y}_i}(6,6)} \\ \cos \alpha = \mathbf{m}^\top \mathbf{h}^C \|\mathbf{h}^C\|^{-1} \end{array} \right.$$

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

# Inverse depth to XYZ conversion threshold



— Camera location error      — 95% acceptance error

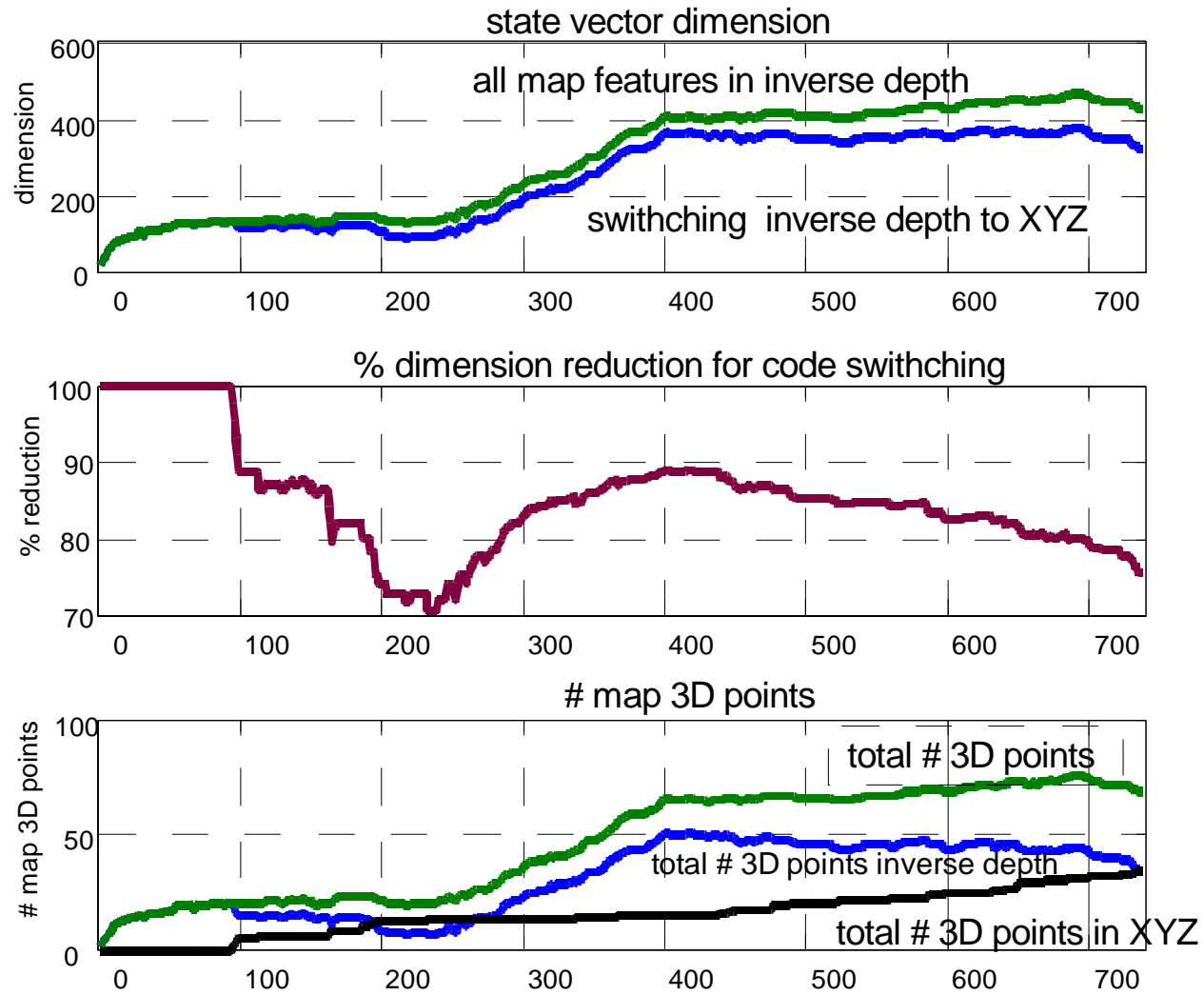


Javier Civera  
JMM Montiel

Imperial College  
London

A.J. Davison

# Switching evolution



# Switching evolution



★ *Inverse depth coding*

△ *Depth coding*



Javier Civera  
JMM Montiel

Imperial College  
London

A.J. Davison

## Bibliografía

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 and Andrew J. Davison: "A visual compass based on SLAM". In Proc. Intl. Conf. on Robotics and Automation, pages 1917--1922, 2006.

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

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