Inverse Depth Monocular SLAM

J. Civera, A.J. Davison and J.M. Montiel
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 images is used as the method to determined 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
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]
State of the Art

- 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
  - 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
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 galaxies at $10^9$ years light far from the Earth
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
Camera Motion Priors

\[
\mathbf{x}_v = \begin{pmatrix}
\mathbf{r}^W_C \\
\mathbf{q}^W_C \\
\mathbf{V}^W \\
\mathbf{\omega}^C
\end{pmatrix}
\]

Constant Velocity Motion Model
smooth camera motion
impulse acceleration noise

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

\[
\mathbf{f}_v = \begin{pmatrix}
\mathbf{r}^W_{C,k+1} \\
\mathbf{q}^W_{C,k+1} \\
\mathbf{v}^W_{C,k+1} \\
\omega^C_{k+1}
\end{pmatrix} = \begin{pmatrix}
\mathbf{r}^W_{C,k} + \mathbf{v}^W_{C,k} \Delta t + \mathbf{a}^W_{k} \Delta t^2 \\
\mathbf{q}^W_{C,k} \times \mathbf{q}(\mathbf{\omega}^C_{k} \Delta t + \alpha^C_{k} \Delta t^2) \\
\mathbf{v}^W_{C,k} + \mathbf{a}^W_{k} \Delta t \\
\omega^C_{k} + \alpha^C_{k} \Delta t
\end{pmatrix}
\]
Scene Point Coding in Inverse Depth. Measurement Equation

Inverse dept point coding

\[ y_i = \begin{pmatrix} x_i & y_i & z_i & \theta_i & \phi_i & \rho_i \end{pmatrix}^T \]

Bearing only camera measurement

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

Measurement equation

\[ h^C = \begin{pmatrix} h_x \\ h_y \\ h_z \end{pmatrix} = R^{CW} \begin{pmatrix} \rho_i \begin{pmatrix} x_i \\ y_i \\ z_i \end{pmatrix} - r^{WC} \end{pmatrix} + m(\theta_i, \phi_i) \]
Scene Point Coding in Inverse Depth. Measurement Equation

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

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

\[ R^{CW} \begin{pmatrix} x_i \\ y_i \\ z_i \end{pmatrix} = m(\theta_i, \phi_i) \]

Those of the first time the feature is observed

\[ \rho_i \begin{pmatrix} x_i \\ y_i \\ z_i \end{pmatrix} - r^{WC} \]

Parallax distant point, \( \rho_i \to 0 \Rightarrow \) parallax goes to zero

Close camera locations, low baseline, \( \begin{pmatrix} x_i \\ y_i \\ z_i \end{pmatrix} - r^{WC} \) goes to zero

Javier Civera
J.M.M. Montiel
A.J. Davison
Scene Point Coding in Inverse Depth. Measurement Equation

\[
\begin{aligned}
\frac{1}{\rho_i} = d_i,
\end{aligned}
\]

\[
\begin{pmatrix}
x_i \\
y_i \\
z_i
\end{pmatrix} + \frac{1}{\rho_i} m(\theta_i, \phi_i)
\]

scene point

parallax angle

\[
\begin{pmatrix}
x_i \\
y_i \\
z_i
\end{pmatrix} - r^{WC}
\]

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

low parallax, only point at infinity is observed:

\[
h^C \approx R^{CW} \left( m(\theta_i, \phi_i) \right)
\]

\[
\begin{aligned}
u &= u_0 - f \frac{h_x}{d_x} h_z \\
v &= v_0 - f \frac{h_y}{d_y} h_z
\end{aligned}
\]
EKF sequential processing

\[
\begin{align*}
x_{k+1|k} & = F(x_{k,k}, 0) \\
\hat{y}_{k+1|k} & = H(x_{k+1|k}) + \text{diag}(R_1, \ldots, R_n)
\end{align*}
\]

stored patch

Estimation at step k

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

Prediction at step k

Update at step k

\[
\begin{align*}
P_{k+1|k} & = F P_{k+1|k} F^T + G \left( \frac{(\Delta t)^2}{2} P_{\alpha} 0 \right) G^T \\
\hat{h}_{k+1|k} & = h(\hat{y}_{k+1|k}, \hat{x}_{k+1|k}) \\
S_{k+1|k} & = H P_{k+1|k} H^T + \text{diag}(R_1, \ldots, R_n)
\end{align*}
\]

Javier Civera
J.M.M. Montiel

A.J. Davison
Inverse Depth Feature Initialization Priors

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

\[
\left[ \rho - 2\sigma_{\rho}, \rho + 2\sigma_{\rho} \right]
\]

covers a region from \( d_{\text{min}} \) and including infinite

\[
\frac{1}{d_{\text{min}}} = \rho_0 + 2\sigma_{\rho}, \quad \rho_0 - 2\sigma_{\rho}
\]
Inverse Depth Feature Initialization

\[
\begin{bmatrix}
\hat{x}_i \\
\hat{y}_i \\
\hat{z}_i \\
\hat{\theta}_i \\
\hat{\phi}_i \\
\hat{\rho}_i
\end{bmatrix}^T,
\]

\[
\begin{bmatrix}
\hat{x}_i \\
\hat{y}_i \\
\hat{z}_i
\end{bmatrix} = r^{WC} \rho_i = \rho_0
\]

\[
\begin{bmatrix}
\theta_i \\
\phi_i
\end{bmatrix} = \begin{bmatrix}
\arctan\left(-h_y^W, \sqrt{h_x^W h_z^W + h_z^W h_z^W}ight) \\
\arctan(h_x^W, h_z^W)
\end{bmatrix}
\]

\[
h^W = R_{WC} \begin{bmatrix}
q_w^W \\
v
\end{bmatrix}
\]

\[
\hat{P}_{k|k}^{new} = J \begin{bmatrix}
\hat{P}_{k|k} & 0 & 0 \\
0 & R_f & 0 \\
0 & 0 & \sigma\rho^2
\end{bmatrix} J^T
\]

\[
J = \begin{bmatrix}
I \\
\frac{\partial y}{\partial \mathbf{r}^{WC}}, \frac{\partial y}{\partial \mathbf{q}^{WC}}, 0, \ldots, 0, \\
\frac{\partial y}{\partial \mathbf{h}}, \frac{\partial y}{\partial \mathbf{\rho}}
\end{bmatrix}
\]

Current camera and map estimate

Javier Civera
J.M.M. Montiel

imperial College
London

A.J. Davison
Dimensional Monocular SLAM

Calibrated image measurements $z_i, \sigma_z$

Frame Rate $\Delta t$

Monocular SLAM Processing

Camera Motion Priors

Initial depth Prior

camera location $(r^{WC}, q^{WC}, v^w, \omega^C, y_1, \cdots, y_n)^T$

scene map

$\sigma_a, \sigma_\alpha$

$\sigma_{\omega_0}, \sigma_{v_0}$

$\rho_0, \sigma_\rho$
Inverse Depth Estimation History

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

distant feature (11), infinite always included in the acceptance region
Loop Closing
Linearity Index

\[ L = \left| \frac{\partial^2 f}{\partial x^2} \right|_{\mu_x}^{2\sigma_x} - \frac{\partial f}{\partial x} \right|_{\mu_x} \]
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

\[
L_d = \frac{A \sigma_d}{d_1} \left| \cos \alpha \right| < 10\%
\]

\[
d_i = \left\| h^C \right\|, \quad h^C = x_i - r^{WC}
\]

\[
\sigma_d = \frac{\sigma_p}{\rho_i}, \quad \sigma_p = \sqrt{P_{y_i,y_i}} (6, 6)
\]

\[
\cos \alpha = m^T h^C \left\| h^C \right\|^{-1}
\]

\[
x_i = \begin{pmatrix} X_i \\ Y_i \\ Z_i \end{pmatrix} = \begin{pmatrix} x_i \\ y_i \\ z_i \end{pmatrix} \quad \left| \frac{1}{\rho_i} m(\theta_i, \phi_i) \right|
\]

\[
P_{\text{new}} = JPJ^T, \quad J = \text{diag} \left( I, \frac{\partial x_i}{\partial y_i}, I \right)
\]
Inverse depth to XYZ conversion threshold

(a) $L_d = 0$

(a) $L_d = 10$

(a) $L_d = 40$

(a) $L_d = 60$

Camera location error

95% acceptance error
Switching evolution

- State vector dimension
- All map features in inverse depth
- Switching inverse depth to XYZ

- % dimension reduction for code switching

- # map 3D points
- Total # 3D points
- Total # 3D points inverse depth
- Total # 3D points in XYZ
Switching evolution

(a)

(b)

(c)

(d)

Inverse depth coding

Depth coding

Javier Civera
J.M.M. Montiel

Imperial College
London

A.J. Davison
Bibliografía


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

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