In 19th Portuguese Conference on Pattern Recognition (RecPad 2013)
Lisboa, Portugal, November 2013

Mosaicing the Interior of Tubular Structures

D. Pereira, J. Tomaz, R. Ferreira, J. Gaspar
http://www.isr.ist.utl.pt

Institute for Systems and Robotics / IST
Lisbon, Portugal

Abstract

This paper addresses a *Simultaneous Mapping and Localization* SLAM methodology for a system capable of performing visual inspections in an unknown environment domain assumed to be a tubular shaped structure (*TSS*), using a monocular camera. Under the *TSS* assumption, a geometrical model was developed, which directly maps visual features from planar images onto a 3D representation, through a backprojection procedure, thus enabling a scenario reconstruction composed of cylindrical segments. The EKF framework allows reconstructing the path described by the camera and its visualized structure, based in visual landmarks described by feature detectors.

1 Introduction

Estimating the motion of a camera moving inside a *Tubular Shaped Structure* (*TSS*) involves considering various aspects. Two important aspects are the number of degrees of freedom of the camera motion and the shape of the environment structure. In this paper, the *TSS* includes straight and curved sections, allowing free movement of the camera. Motion can be estimated by registering the texture, retrieving distinctive visual features, and dewarping into a mosaic. Hence we start from the standard idea of reconstructing points of the scene and then focus on fitting a simple 3D cylindrical model to the various tube sections, which makes simple the dewarping step [4].

2 Estimation of Tubular Structure and Camera Motion

Assuming the world as a *TSS* domain navigated by a forward hand-held like monocular camera (see Fig 3), smoothly moving with a constant speed inside a tube without revisiting previous positions, one can determine the features 3D locations, which are strapped to a cylindrical section wall with radius ρ. By thoroughly defining a state vector $\mathbf{s}$ composed of both cameras and the *TSS* geometrical parameters estimates, and incorporating a geometrical observation model reliable enough to produce relevant estimations of the cylindrical section geometrical parameters, the simultaneous camera localization and mapping can be achieved.

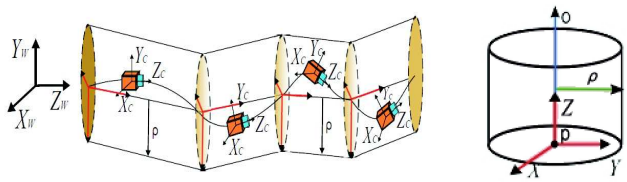


Figure 3: TSS Model (left) and tube parameters (right).

2.1 Extended Kalman Filter

In this section we define a filtering methodology that allows estimating both the TSS shape and the camera motion. The first step consists of defining the state vector of the filter, $\mathbf{s}$, as a joint composition of camera pose $\mathbf{x}$ and cylinders geometrical parameters $\mathbf{y}$:

$$\mathbf{s}_k = [\mathbf{x}_k \; ; \; \mathbf{y}_k] \tag{1}$$

where the semicolon denotes vertical stacking of vectors.

Camera motion behavior is modeled with a constant velocity dynamic model. The state vector $\mathbf{x}_k$ provides the position $\mathbf{r}_k^{WC}$, orientation $\mathbf{q}_k^{WC}$ and both linear $\mathbf{v}_k^W$ and angular ω_k^C velocities at every instant, i.e.

$$\mathbf{x}_k = \left[\mathbf{r}_k^{WC} \; ; \; \mathbf{q}_k^{WC} \; ; \; \mathbf{v}_k^W \; ; \; \omega_k^C\right]. \tag{2}$$

A cylindrical section is characterized by a 7 dimension $\mathbf{y}_{nk}$ state vector as an array of the cylinder parameters centre position $\mathbf{p}_{nk}^W$, orientation $\mathbf{o}_{nk}$, and radius $\log(\rho_{nk})$ expressed in a logarithm form

$$\mathbf{y}_{nk} = \left[\mathbf{p}_{nk}^W \; ; \; \mathbf{o}_{nk} \; ; \; \log(\rho_{nk})\right]. \tag{3}$$

A Kalman Filter estimates optimal states over time, given observations in the presence of noise of a dynamic system driven by noise inputs. This estimation is computed recursively as a Markov chain model, i.e. it only requires the previous estimate $k-1$ of a given state at instance k. An assumption required by this framework is to use linear systems and observations models with zero-mean multivariate Gaussian distributed noise. The Extended Kalman Filter (EKF) allows handling non-linear systems functions and measurement models

$$\begin{aligned} \mathbf{x}_k &= \mathbf{f}(\mathbf{x}_{k-1}, \mathbf{u}_k) + \varepsilon_k \\ \mathbf{z}_k &= \mathbf{h}(\mathbf{x}_k, \mathbf{u}_k) + \delta_k \end{aligned} \tag{4}$$

where ε_k and δ_k, denote system and measurement noise. Propagating the uncertainties and updating the system vector are handled through linearization, with a first order Taylor series expansion, of the state transition and measurement functions.

2.2 System Dynamics

The non-linear state transition model function $\mathbf{f}$, is defined as the stacking of two independent state transition processes, $\mathbf{f}_x$ and $\mathbf{f}_y$, for both camera and cylinder state vector $\mathbf{x}$ and $\mathbf{y}$, which are influenced by an additive zero-mean Gaussian transition noise $\varepsilon_k \sim \mathcal{N}(0, \mathbf{Q}_k)$:

$$\mathbf{f} = \left[\mathbf{f}_x \; ; \; \mathbf{f}_y\right]. \tag{5}$$

The constant velocity model allows smooth velocity variations with zero-mean Gaussian distributed acceleration noise $\mathbf{n}_x \sim \mathcal{N}(0, \mathbf{N}_{xk})$. The acceleration noise has linear and angular components $\mathbf{n}_x = [\mathbf{a}^W; \vartheta^C]$. Under this model assumption, and defining $\mathbf{V}^W = \mathbf{a}^W \Delta t$ and $\Omega^C = \vartheta^C \Delta t$ as the linear and angular velocities impulse between a transition step Δt, one can express the camera state transition, $\mathbf{x}_{k+1} = \mathbf{f}_x(\mathbf{x}_k, \mathbf{n}_{xk})$ [1, 2]:

$$\begin{bmatrix} \mathbf{r}_{k+1}^{WC} \\ \mathbf{q}_{k+1}^{WC} \\ \mathbf{v}_{k+1}^W \\ \omega_{\omega+1}^C \end{bmatrix} = \begin{bmatrix} \mathbf{r}_k^{WC} + \left(\mathbf{v}_k^W + \mathbf{V}^W\right)\Delta t \\ \mathbf{q}_k^{WC} \times \mathbf{q}\left(\left(\omega_k^C + \Omega^C\right)\Delta t\right) \\ \mathbf{v}_k^W + V^W \\ \omega_k^C + \Omega^C \end{bmatrix}. \tag{6}$$

2.3 Observation Model

In this work, we use the pin-hole model to describe geometrically the 2-D imaging of a 3-D point:

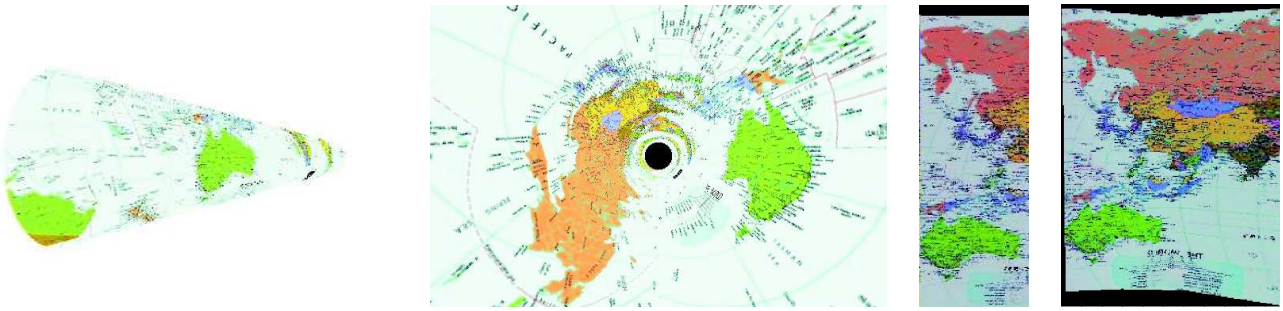
$$\tilde{\mathbf{m}} \sim \mathbf{P}\tilde{\mathbf{M}} \tag{7}$$

where $\tilde{\mathbf{M}} = [X \; Y \; Z \; 1]^T$ denotes a 3D point, $\tilde{\mathbf{m}} = [u \; v \; 1]^T$ is the respective image, and $\mathbf{P}$ is the projection matrix [3]. Conversely, to describe the 3-D location of points $\mathbf{M}$ in space, imaged as $\mathbf{m}$ in the image plane, one can relate the 3-D ray leaving the camera optical centre $\mathbf{C}$ and piercing the image plane at $\mathbf{m}$ as a line which intersects the plane where $\mathbf{M}$ is lying. $\mathbf{C}$ and $\mathbf{D}$ define a line in 3-D space, where every point in the line is thus given by the following equation $\mathbf{M} = \mathbf{C} + \alpha\mathbf{D}$, with a scaling factor $\alpha \in [-\infty, +\infty]$. The full expression for representing a point $\mathbf{m}$ back-projection into $\mathbf{M}$ is

$$\mathbf{M} = -\mathbf{p}_{(123)}{}^{-1}\mathbf{p}_4 + \alpha\mathbf{p}_{(123)}{}^{-1}\mathbf{m} \tag{8}$$

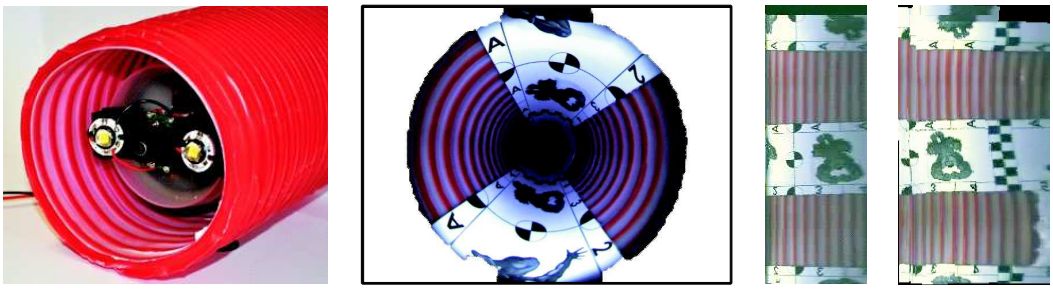
where $\mathbf{p}_{(123)}$ is a 3×3 matrix representing the first three columns of the projection matrix $\mathbf{P}$, and $\mathbf{p}_4$ is the fourth column.

The observation model describes the process of implicitly representing observed features Λ in a 3-D parameterization as a function of the systems state $\mathbf{s}$ parameters by inferring constraints to the environment structure and consequently to the features 3-D locations. At an instant k



(a) (b) (c) (d)

Figure 1: VRML setup (a). View of the camera inside the tube (b). Dewarping of the tube for a single step (c). Result of dewarping several steps (d).



(a) (b) (c) (d)

Figure 2: Real setup (a). Inside view of the tube (b). Dewarping of the image for one step (c). Result of dewarping several steps (d).

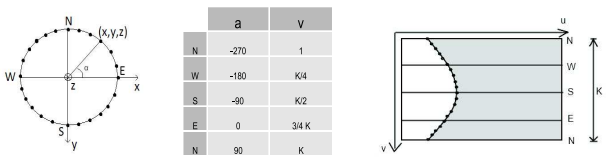
the state transition function $\mathbf{f}$ is applied to $\mathbf{s}_k$, retrieving $\mathbf{s}_{k+1} = \mathbf{f}(\mathbf{s}_k, \mathbf{n}_k)$. Λ_k is the input set of features in pixel coordinates acquired at the instant k, whereas $\widehat{\Lambda}_{k+1}$ is the set of features in following instant $k+1$ and the observation model output. The observation model can thus be written as a function $\mathbf{h}$:

$$\widehat{\Lambda}_{k+1} = \mathbf{h}_{k+1}(\mathbf{s}_{k+1}, \Lambda_k) \tag{9}$$

Consider that the set of features Λ_k, 3-D locations on the *TSS* surface, are the intersections of rays leaving the camera centre $\mathbf{r}_k^{WC}$ with the cylinder wall, piercing the 2-D image plane where all features lie. With the information present in the state vector $\mathbf{s}_k$, i.e. camera pose $\mathbf{x}_k$ and cylinders section parameters $\mathbf{y}_k$, one can estimate the 3-D location of every feature in set Λ_k through a back-projection methodology, and compute the re-projection of this 3-D coordinates with the next instant $k+1$ state vector $\mathbf{s}_{k+1}$ to a 2-D image plane, thus acquiring $\widehat{\Lambda}_{k+1}$ features.

3 Dewarping

The process of dewarping the interior of the tube is done by finding the transformation between the pixels in the "open" images and the pixels in the "closed" scenes viewed by the camera inside the *TSS*. Figure 4 shows that relation. Knowing that relation, each pixel in the "open" image has a corresponding 3D position in reference to the camera's centre. The projection of that 3D point in the "closed" image is the respective pixel in the "open" image. A final mosaic is stitched showing the full inside texture of the tube.



	a	v
N	-270	1
W	-180	K/4
S	-90	K/2
E	0	3/4 K
N	90	K

(a) (b) (c)

Figure 4: (a) Coordinates of the points in the closed image (b) Relation between the angle of the coordinate in the closed image and the vertical pixel in the open image (c) Pixels in the open image

4 Experimental Results

Two experiments have been conducted to test the proposed methodology. In the first experiment the tubular shape is simulated in VRML. The VRML based system imply using image processing, namely features (e.g. SIFT or SURF) detection and matching, enhanced by the RANSAC methodology for outliers removal. Figure 1 shows various steps and results of the process. The filtered nature of the estimated tubular structure and camera motion implies a dewarping which has small variations, expansion or compression, along the tube length. In Figure 1(d) the small variations along tube length are compensated vertically (the small black regions in the mosaic indicate this compensation).

The second experiment consists of evaluating the developed algorithm with a set of images retrieved from a real camera navigation inside a textured *TSS*. The setup and the results can be seen in Figure 2. The resulting mosaic, Figure 2(d), shows some horizontal oscillation denoting under- or over-estimation of camera rotation. This is due to biased observations such as the ones associated to an unbalanced number of features around the camera.

5 Conclusion

This paper aimed at presenting a robust model to solve the *Simultaneous Localization and Mapping* problem with a priori structural environment knowledge, assumed to be a *TSS*. The cameras pose estimate and the camera trajectory can be recovered, up to a scale factor, assuming the constant speed dynamic model, allowing a posterior description of the environment structure and its reconstruction.

Acknowledgments

This work has been partially supported by the FCT project PEst-OE / EEI / LA0009 / 2013, by the FCT project PTDC / EEACRO / 105413 / 2008 DCCAL, and by the FCT project EXPL / EEI-AUT / 1560 / 2013 ACDC.

References

[1] Ted J. Broia and Rama Chellappa. Estimating the kinematics and structure of a rigid object from a sequence of monocular images. *IEEE T-PAMI*, 13(6):497–513, June 1991.

[2] Javier Civera, Andrew J. Davison, and J.M.M Montiel. Inverse depth parametrization for monocular slam. *IEEE Transactions on Robotics*, 24(5):932–945, October 2008.

[3] Olivier Faugeras. *Three-Dimensional Computer Vision, A Geometric Viewpoint*. The MIT Press, 1993.

[4] Luís Ruivo. Omni-sonda: Câmara de visão omnidireccional para inspecção visual. Master's thesis, DEEC/IST, 2007.