Scripsie: State Estimation & Observation

Aidan Landsberg

February 27, 2015

Dr. Van Daalen,

This section, for the moment, provides initially, a general background to explain the content that follows. It will later be integrated into another, larger section in the final report, while other information will still be added to this particular section. The main goal is to set up the necessary theoretical basis in order to describe, amongst other thing, the motion model (prior to your approval). Also, I have purposefully omitted all relevant references as I am yet to completely study all the literature I have obtained. I am though, well aware of which sections of this write up is to be referenced and will proceed to do so at a later stage.

1 Single Camera Simultaneous Localisation and Mapping

The ultimate goal of the approach presented here, is to obtain a probabilistic three dimensional (3D) map of features, representing at every time instance, the estimates of both the state of the camera as well as the positions of every feature observed. These features of interest are more commonly referred to as *landmarks* and the aforementioned terms will, from hereon in, be used synonymously. Most importantly though, the map is to contain the *uncertainty* associated with each of the aforementioned estimates.
The process regarding the construction of this map of features is to be implemented through the use of an (Extended) Kalman filter. The map initially, completely void of any landmarks, is recursively updated according to the subsequent fusions of both predictions and measurements presented to the Kalman filter. As new (potentially interesting) features are observed, the state estimates of both the camera as well as the landmarks are both updated - augmenting the state vector with additional features (if indeed they are observed) while deleting any landmarks that are of no longer of interest. In order to obtain the best possible result, the algorithm should strive to obtain a sparse set of higher-quality landmarks rather than a dense set of ordinary landmarks within the environment.

2 Extended Kalman Filter

The following section will provide a brief introduction to the *Extended Kalman Filter* (EKF), the fundamental algorithm that enables the realisation of the SLAM problem. With regard to SLAM, a Kalman Filter can be briefly decried as an algorithm that **optimally** estimates the state of the robots pose as well as the position of the landmarks within the map, given process and measurement noise. The EKF, an extension of the general Kalman filter, aims to enable the modelling of non-linear systems.
The Kalman filter is a popular, well studied technique for filtering and prediction of a *continuous* linear system that contains uncertainty. It is realised utilising a recursive Gaussian filter in order to estimate the state of the system accordingly, provided that the state vector $\hat{\mathbf{x}}_k$ is modelled by a single multivariate Gaussian distribution.

The solution to this specific implementation of the Simultaneous Localisation and Mapping (SLAM) problem, takes the following probabilistic form:

$$P\big(\hat{\mathbf{x}}_k, \hat{\mathbf{y}}_k \mid \mathbf{z}_{0:k}, \mathbf{u}_{0:k}, \mathbf{x}_0\big). \tag{2.1}$$

with the aforementioned distribution described at every time instance k. A brief description would yield that the distribution above, describes a joint density - at given time instance k - of the robot state as well as the landmark locations **given** all of the previously recorded observations and control inputs. It is assumed that a measurement function h is used to obtain the measurements $\mathbf{z}_k$ - these measurements also incorporate an uncertainty. Using the Kalman filter, each time-step k can be operated into two (sequential) steps: the *prediction step* and the *update step*. Each of the aforementioned steps are discussed in more detail later in this section, but the general idea is that the Kalman filter firstly estimates the new state of the system according to the previous state as well as the control inputs, after which a measurement is predicted if indeed the system were to find itself within the estimated state. An actual measurement is then obtained through the system sensors to determine the actual state of the system and ultimately, the actual state of the system is then compared to the predicted state. It is important to note that both prediction and measurement influence the state of the system.

The general implementation above, is only valid when considering linear systems. Considering that most practical systems of interest yield non-linear behaviour, the Kalman filter in its purest form is of little essence. Instead, linearisation can be used in order to accommodate non-linear behaviour. The linearisation process (to be explained later in this section) can briefly be described by using first order Taylor expansion to create a linear approximation of a non-linear function. The Taylor expansion creates an approximation that is linear and dependent on the properties of the functions derivative yielded from a single - generally that of the most likelihood - point (mean). *Jacobians* are commonly used to linearise non-linear functions. It is therefore important - in order to successfully implement a Kalman filter on a non-linear system - to linearise both the state transition model as well as the measurement model; both of which are generally non-linear functions. Once achieved, the EKF; which behaves otherwise identically in terms of operation to the general Kalman filter, can be implemented upon non-linear systems.

2.1 State Representation

All relevant state estimates are embedded within the state vector $\hat{\mathbf{x}}_t$ which is comprised of two parts, the camera state $\hat{\mathbf{x}}_v$ and the feature estimates $\hat{\mathbf{y}}$ respectively. The camera state provides the estimate for its pose at each tilmestep and the landmark estimates provide the landmark's feature description as well its estimated position within the map.
Mathematically, the probabilistic map can be represented through the state vector $\hat{\mathbf{x}}_t$ and a covariance matrix P. $\hat{\mathbf{x}}_t$, as previously mentioned, is a a single column vector containing the estimates of the camera as well as the landmarks, and P is a square matrix. These quantities can be mathematically shown as follows:

$$\hat{\mathbf{x}}_t = \begin{pmatrix} \hat{\mathbf{x}}_v \\ \hat{\mathbf{y}}_1 \\ \hat{\mathbf{y}}_2 \\ \vdots \\ \hat{\mathbf{y}}_N \end{pmatrix}, \quad P_{NN} = \begin{bmatrix} P_{x,x} & P_{x,y_1} & P_{x,y_2} & \cdots & P_{x,y_N} \\ P_{y_1,x} & P_{y_1,y_1} & P_{y_1,y_2} & \cdots & P_{y_1,y_N} \\ P_{y_2,x} & P_{y_2,y_1} & P_{y_2,y_2} & \cdots & P_{y_2,y_N} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ P_{y_N,x} & P_{y_N,y_1} & P_{y_N,y_2} & \cdots & P_{y_N,y_N} \end{bmatrix}. \tag{2.2}$$

These quantities then, allow us to approximate the uncertainty regarding the generated feature map as a N-dimensional single multi-variate Gaussian distribution, where N, as stated above, is the total number of state estimates within the state vector.

Before continuing, it is important to consider and understand the notation used in the sections that follow. Two separate coordinate systems are to be considered, namely the *fixed* inertial reference fame system W and the cameras free coordinate frame system, more commonly referred to as the body frame C. System variables defined within either of the aforementioned coordinate systems, are from here on in, to be designated a superscript to establish in which coordinate system it may be relevant (e.g. x^W). Derivatives of parameters are denoted through a prime symbol, second derivates are denoted through a double dot symbol and so forth; for instance the derivative of position x will be denoted as $\dot{x}$ and its second derivative denoted as $\ddot{x}$. Vectors will be printed in bold and non-italics to better distinguish them from scalers. An example can be shown regarding the variable x: $\mathbf{x}$ denotes a vector while its scalar counterpart would be represented as x.

2.1.1 Camera Position State Representation

The following concept describes a suitable method to represent all relevant information regarding the cameras position and orientation in a 3D space. According to most implementations of robot localisation, there exists no concern to contrast between the concepts of a camera state $\hat{\mathbf{x}}_v$ and a camera position state $\mathbf{x}_p$: it is therefore important to note that a position state - containing the required information regarding a robots position - is merely an element of the camera state vector. The state camera vector - comprising of 13 individual states - is mathematically described as follows:

$$\hat{\mathbf{x}}_v = \begin{pmatrix} \mathbf{r}^W \\ \mathbf{q}^{WC} \\ \mathbf{V}^W \\ \omega^C \end{pmatrix}. \tag{2.3}$$

where $\mathbf{r}^W = (x\ y\ z)^T$ indicates the 3D cartesian position of the camera, $\mathbf{q}^{WC}$ the unit orientation *quarternion* - to be mathematically defined and described in the appendix - indicating the camera orientation (represented in the body frame) *relative to the inertial reference frame* W, $\mathbf{V}^W$ indicating the linear velocities of the camera relative to the inertial reference frame W and ω^C indicating the angular velocities of the camera relative to its own body frame C. Often, the modelling of dynamic systems require that additional parameters - separate to those describing the position and orientation of the robot - be included in the state vector along with the position state vector. This is illustrated in the description above, with the position state vector $\mathbf{x}_p$ comprising of the 3D position vector, $\mathbf{r}^W$ and the unit orientation *quarternion*, $\mathbf{q}^{WC}$. The linear and angular velocities, $\mathbf{V}^W$ and ω^C, form the additional information required for system modelling.

2.1.2 Cartesian Feature Representation

As previously discussed, the aim is to describe a set of high-quality, well defined landmarks within the map. The map itself is to contain a 3D position of *each* observed landmark as well as a combined uncertainty. The feature estimates $\hat{\mathbf{y}}$ - comprising of N landmarks - is mathematically described through three individual cartesian coordinates - x, y and z respectively:

$$\hat{\mathbf{y}}_n = (x_n\ \ y_n\ \ z_n)^T. \tag{2.4}$$

where n represents a specific single landmark.

With reference to the theory on image processing, it can be discussed that the depth of a given landmark (in this case the z-coordinate) cannot be immediately determined, but rather approximated via triangulation given the landmark is observed over a sequence of (minimally) two known camera positions. The x and y measurements however, can be immediately determined from the image plane.

2.1.3 Control Inputs

2.2 Prediction Step

With reference to the probabilistic form of the solution to the SLAM problem, the prediction step requires a description in terms of a probability distribution. The description of the aforementioned motion model can then, in terms of the probability distribution on the state transitions, take the following form:

$$\begin{aligned} &P(\hat{\mathbf{x}}_k \mid \hat{\mathbf{x}}_{k-1}, \mathbf{u}_k) \\ &\quad = \frac{1}{\sqrt{|2\pi \mathbf{R}_w|}} exp\left(\frac{1}{2}(\hat{\mathbf{x}}_k - \hat{\mathbf{x}}_{k\to k+1})^T \mathbf{R}_w^{-1}(\hat{\mathbf{x}}_k - \hat{\mathbf{x}}_{k\to k+1})\right). \end{aligned} \tag{2.5}$$

where $\hat{\mathbf{x}}_{k\to k+1}$ represents the motion model.

The motion model is assumed to take the form of a Markov process, yielding that the next state $\hat{\mathbf{x}}_k$ is only dependent upon the state immediately preceding it - $\hat{\mathbf{x}}_{k-1}$ - as well as the input control $\mathbf{u}_k$. Additionally, it is important to note that the uncertainty regarding the motion model is independent of the uncertainty regarding both the observation model as well as that of the probabilistic map itself.

2.2.1 State Space Model

As previously discussed, the Kalman Filter requires a state transition (motion) model in order to estimate the current state of the system. In short, the motion model describes the transition from the previous state to the following state with regard to the robots kinematic motion as well as the control inputs. The motion model in this particular instance can be described through a **linear** differential equation of the following form:

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t) + \mathbf{w}(t). \tag{2.6}$$

where the state matrix $\mathbf{A}$, describes the manner in which state evolves from time $t-1$ to t without the influence of noise and controls, the input matrix $\mathbf{B}$, describes how the control vector $\mathbf{u}(t)$ evolves from time step $t-1$ to t and $\mathbf{w}(t)$ is a **zero-mean**

Gaussian random variable representing the process noise with a covariance matrix $\mathbf{R}_w$.

Considering that the Kalman Filter is a recursive, numerical evaluation, it is necessary to convert the previously defined continuous model into its discrete counterpart. Various methods of discretisation exist, though this specific implementation makes use of the forward difference/Eulers method. This method *approximates* the derivative for a state for a sampling period ΔT as follows:

$$\begin{aligned} \dot{\mathbf{x}}[k] &= \lim_{\Delta T \to 0} \frac{\mathbf{x}[k+1] - \mathbf{x}[k]}{\Delta T} \\ \Delta T \dot{\mathbf{x}}[k] &\approx \mathbf{x}[k+1] - \mathbf{x}[k]. \end{aligned} \tag{2.7}$$

The state estimate of the discrete counterpart at the following sampling instance, namely $k+1$, is then presented as follows (given a small enough sampling instance ΔT):

$$\mathbf{x}[k+1] = \big(\mathbf{I} + \mathbf{A}\Delta T\big)\mathbf{x}[k] + \mathbf{B}\mathbf{u}[k]\Delta T + \mathbf{w}[k]\Delta T. \tag{2.8}$$

where $\big(\mathbf{I} + \mathbf{A}\Delta T\big) = \mathbf{A}_d$ is the discrete state matrix, $\mathbf{B}\Delta T = \mathbf{B}_d$ is the discrete input matrix and $\mathbf{w}[k]\Delta T = \mathbf{w}_d[k]$ is the discrete input process noise.

Ultimately, the form of the final difference equation describing the system at each individual sampling instance is given as follows:

$$\mathbf{x}[k+1] = \mathbf{A}_d\mathbf{x}[k] + \mathbf{B}_d\mathbf{u}[k] + \mathbf{w}_d[k]. \tag{2.9}$$

2.2.2 State Transition

In order to derive the motion model for the system at hand, it is vital that the certain characteristics of the system be understood. Firstly, the robot system - from here on in to be referred to as the **camera** - is comprised of a monocular camera and an attached Inertial Measurement Unit (IMU) package. Secondly, the camera is to be considered as a six degree of freedom (DOF) rigid body. Briefly the six DOF describe the camera's three *translational* and three *rotational* degrees of freedom.
We therefore set out to define a kinematic motion model - using Newton's laws of motion - to describe the cameras movement through the environment as a result of initially unknown, external inputs to the system. Lastly, it should be stressed that embedded within the motion model, should be the impacts of uncertainty through both internal and external factors. It is assumed in this instance, that at each

time-step, an unknown angular acceleration $\boldsymbol{\Omega}^R$ acts upon the system. This input is modelled as a zero-mean Gaussian process that causes an impulse of angular velocity:

$$\mathbf{w}_d[k] = \mathbf{w}[k]\Delta T = \left[\boldsymbol{\Omega}^R\right] = \begin{pmatrix} \alpha_x \Delta T \\ \alpha_y \Delta T \\ \alpha_z \Delta T \end{pmatrix}. \tag{2.10}$$

with a covariance matrix $\mathbf{R}_w$ that is assumed as a diagonal initially, to represent uncorrelated noise in all of the rotational components.
It must also be stressed that initially, a stochastic, linear discrete-time model is adopted to approximate the motion model. Using the kinematic equations of linear and angular motion, it is aimed to ultimately and complete the previously defined state space model. We begin by describing all relevant states and inputs:

$$\begin{aligned} \mathbf{x}[k] &= \begin{bmatrix} x_k & y_k & z_k & q_{0,k} & q_{1,k} & q_{2,k} & q_{3,k} & \dot{x}_k & \dot{y}_k & \dot{z}_k & \alpha_{x,k} & \alpha_{y,k} & \alpha_{z,k} \end{bmatrix}^T \\ \mathbf{u}[k] &= \begin{bmatrix} \ddot{x}_k & \ddot{y}_k & \ddot{z}_k & \dot{q}_{0,k} & \dot{q}_{1,k} & \dot{q}_{2,k} & \dot{q}_{3,k} \end{bmatrix}^T \end{aligned} \tag{2.11}$$

and extend the discrete-time difference equation describing the system to incorporate the motion model,

$$\mathbf{x}[k+1] = \mathbf{A}_d\mathbf{x}[k] + \mathbf{B}_d\mathbf{u}[k] + \mathbf{w}_d[k]$$

$$\mathbf{A}_d = \begin{bmatrix} 1 & 0 & 0 & \Delta T & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & \Delta T & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & \Delta T & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} = \left(\mathbf{I} + \mathbf{A}\Delta T\right)$$

$$\mathbf{B}_d = \begin{bmatrix} \Delta T & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \Delta T & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \Delta T & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \Delta T & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \Delta T & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \Delta T & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \Delta T \end{bmatrix} = \mathbf{B}\Delta T \tag{2.12}$$

$$\mathbf{w}_d[k] = \mathcal{N}(0, \mathbf{R}_w) = \left[\boldsymbol{\Omega}^R\right] = \mathbf{w}[k]\Delta T$$

it can be observed from the model above that the motion model adheres to the forward method of discretisation derived in (2.8). The motion model also adheres to the Markov process assumption, in that it can be completely described through only its transition from the previous state as well as the control inputs.

2.2.3 Non-Linear Modelling

2.3 Correction Step

With reference again to the probabilistic form of the solution to the SLAM problem, the measurement step too, requires a description in terms of a probability distribution. The observation model however, models the uncertainty regarding a measurement taken at an instance $\hat{\mathbf{z}}_k$ given that the locations of both the robot as well as the landmarks are known. This uncertainty can be described in the following form:

$$\begin{aligned} &P(\hat{\mathbf{z}}_k \mid \hat{\mathbf{x}}_k, \hat{\mathbf{y}}_k) \\ &\quad = \frac{1}{\sqrt{|2\pi \mathbf{Q}_v|}} exp\left(\frac{1}{2}(\hat{\mathbf{z}}_k - \hat{\mathbf{z}}_{k\to k+1})^T \mathbf{Q}_v^{-1}(\hat{\mathbf{z}}_k - \hat{\mathbf{z}}_{k\to k+1})\right). \end{aligned} \tag{2.13}$$

where $\hat{\mathbf{z}}_{k\to k+1}$ represents the measurement model.

It can be (reasonably) assumed that the uncertainty regarding the measurements are conditionally independent given the uncertainty regarding the robot and landmark locations if indeed they are completely defined. Also, the correction step seeks to obtain the difference between the actual measurements $\hat{\mathbf{z}}_k$ and the predicted measurements. These predicted measurements are to be obtained through an observation model that we from hereon in refer to as the measurement function, denoted as $\mathbf{h}_i$.

2.3.1 Measurement Function

The correction step of the Kalman filter aims to ultimately correct the previously estimated robot pose and landmark position through exterior sensor measurements. The measurement process generally involves a measurement estimate that incorporates an uncertainty. With regard to the implantation proposed in this paper, landmarks are required to be observed and measured through the use of a camera. To mathematically describe this process, the previously mentioned measurement function is used to effectively model the measurement estimation. It is essential that the measurement function, like the motion model, be **linear** in nature and additionally, the

measurement function must describe the position of a **point** feature with regard to the previously estimated states - namely the robot pose and the landmark positions.

Considering that the camera observations are obtained with regard to its own reference frame C, the definition of the measurement function is adapted in order to be described with regard to the inertial reference frame. The measurement function $\mathbf{h}_i^W$ that describes a directional vector in relation to the cameras body frame is thus mathematically defined as follows:

$$\mathbf{h}_i^W = \mathbf{R}^{CW}\left(\mathbf{y}_i^W - \mathbf{r}^W\right) = \left(\begin{pmatrix} x_i \\ y_i \\ z_i \end{pmatrix} - \mathbf{r}^W\right) \tag{2.14}$$

where the subscript i corresponds a directional vector $\mathbf{h}^C$ to its cartesian point $\mathbf{y}^W$, $\mathbf{r}^W$ describes the cartesian position of the camera, $\mathbf{y}^W$ describes the cartesian position of a given landmark and $\mathbf{R}^{CW}$ represents the rotational matrix that is required to transform the aforementioned positional vectors from the inertial reference frame into the cameras body frame coordinate system.

With reference to Chapter 2 regarding perspective cameras, it can be recalled that a given features position is described by a 2-dimensional position of the image frame of the camera. Recalling, the standard pinhole camera model defines this position mathematically as follows:

$$\mathbf{h}_i = \begin{pmatrix} u_i \\ v_i \end{pmatrix} = \begin{pmatrix} u_0 - fk_u \frac{h_{i,x}^R}{h_{i,z}^R} \\ v_0 - fk_v \frac{h_{i,y}^R}{h_{i,z}^R} \end{pmatrix} \tag{2.15}$$

where fk_u, fk_v, u_0 and v_0 are the previously described camera calibration parameters.

2.3.2 Feature Tracking

2.3.3 System Update