Combining PMHT with classifications to perform SLAM

Brian Cheung†,⋆
brian.cheung@dsto.defence.gov.au
Samuel Davey†
samuel.davey@dsto.defence.gov.au
Douglas Gray⋆
dgray@eleceng.adelaide.edu.au
†Defence Science and Technology Organisation, Australia
⋆School of Electrical and Electronic Engineering, The University of Adelaide, Australia

Abstract – The problem referred to as Simultaneous Localisation and Mapping (SLAM) requires estimation of unknown target locations when the platform location knowledge is unreliable. It is a technique often associated with autonomous platforms that carry a variety of complementary sensors. Besides target detection and platform positional information, these sensors, such as laser ranging and cameras, can often provide perceived classification information that is generally not utilised by the data association algorithm. This paper demonstrates how classification information can be used to assist the data association technique known as the Probabilistic Multi-Hypothesis Tracker (PMHT) when applied to the feature-based SLAM problem. Some example results are given to illustrate the performance improvement that can result from this approach.

Keywords: Data association, probabilistic multi-hypothesis tracker (PMHT), classification, simultaneous localisation and map building (SLAM).

1 Introduction

Simultaneous localisation and mapping (SLAM) is the problem of creating a map of an unknown area while concurrently estimating the location of the sensing platform within the map. This paper is concerned with feature-based SLAM, where the map is represented by a collection of landmarks at (unknown) fixed locations. This problem may be difficult because the unknown landmark locations and the unknown platform trajectory are coupled through the (typically non-linear) measurement process. Many SLAM solutions use the statistical assumption that the landmark locations and the platform states are random variables, as introduced by Smith et al. [1]. One method of solution is then to stack the unknowns into a single state vector and employ an Extended Kalman Filter (EKF) [1].

The two main problems with the EKF approach are that the state vector may be very large, and hence the covariance matrix of its estimate may be unwieldy, and that the EKF may be an inadequate approximation for the non-linear problem. Various approaches have been developed to address these shortcomings. For example, FastSLAM [2, 3] uses particle filtering to address non-linearity and factorisation to avoid large state vectors; the Unscented Kalman Filter (UKF) [4] has been used as an alternate non-linear filter; and the Extended Information Filter (EIF) [5] uses a recursion for the inverse of the covariance matrix which has been shown to be approximately sparse.

Much of the SLAM literature is based on state estimation methods. As with target tracking, data association may also be important, but this is an issue not usually considered. In most cases, it is explicitly assumed that the association of measurements to landmarks is known [6]. In practice, solutions such as nearest neighbour association are used. It is important that the platform identify and associate landmarks with current observations accurately to minimise errors in both the estimation of localisation and mapping.

The Probabilistic Multi-Hypothesis Tracker (PMHT), developed by Streit and Luginbuhl [7], is a data association algorithm derived from the application of the Expectation Maximisation (EM) algorithm [8] to target tracking. The PMHT uses EM to model the assignment of measurements to targets as hidden variables and estimates target states by taking the expectation over the assignments. The advantage of the PMHT over other conventional data association techniques is that it has linear complexity in the number of landmarks it is tracking. Davey applied the PMHT algorithm to data association within the SLAM context [9]. It was shown to result in state estimation performance superior to various hard assignment alternatives.

SLAM is generally performed by a mobile platform that has a variety of different sensors onboard. In addition to landmark feature detection, other information may be available, typically classification measurements from automatic target recognition algorithms that help associate measurements with particular landmarks. Landmark classification and recognition can be used to aid the data association process. This may assist by recognising that a landmark has been revisited in cases where the position measurements are ambiguous. Classification may also provide a means to
 recognising moving targets and landmarks that can dynamically change. For example, by classifying a particular landmark as having certain properties (such as a door being opened or closed), appropriate models can be applied.

PMHT has been extended to make use of classification information to improve data association when the kinematic information is ambiguous, and this extension is referred to as the PMHT-c [10]. An important trait of PMHT-c is that it allows the use of classification estimates which may be inaccurate, rather than assuming that the classifier has perfect accuracy. This paper extends the existing PMHT SLAM method of Davey [9] to incorporate classification information and thereby improve data association accuracy. It follows the PMHT-c approach of assuming that the classifier provides imperfect observations of the true landmark class.

This paper focuses on the use of classification measurements to improve data association while performing SLAM. The source of these classification measurements is application dependent and will not be explored here. The paper assumes that for each of the sensor measurements, there is a classification observation available. The set of classes is known, and the statistical accuracy of the classifier is known through knowledge of the classifier’s confusion matrix. As an example, classes may discriminate landmarks based on size or colour. The selection of appropriate features for the classification input is an active field for outdoor environments, e.g. [11, 12], indoor environments, e.g. [13], and multi-environments, e.g. [14].

The remainder of the paper is arranged as follows. Section 2 outlines the SLAM problem. Section 3 derives the PMHT-c for SLAM. Section 4 demonstrates some of the results achieved on simulated data and on an experimental data set collected in Victoria Park, Sydney, Australia. Section 5 presents a summary of the results of the paper.

2 Problem Formulation

In SLAM, the problem is to jointly estimate the location of the platform(s) and to estimate a map by which the platform may navigate. This paper addresses feature based SLAM, where the map consists of a set of discrete landmarks. Assume that there are \( N \) sensors observing \( M \) landmarks over \( T \) observation times, and that not all of these landmarks may be visible to the sensors at any one time.

Let the state of sensor \( n \) at scan \( t \) be denoted by \( y^n_t \) and the set of all sensor states by \( Y^n \). Similarly, let the state of landmark \( m \) at scan \( t \) be denoted by \( x^m_t \) and the set of all landmark states by \( X^m \). It is assumed that the dimension of the state vector for each landmark and sensor is fixed, but may vary between landmarks and sensors. Some (or all) of the sensors may be situated on common platforms and it is assumed that redundant states are not estimated. The sensor state may consist of the position, velocity and orientation of the sensor, whereas the landmark state may only contain the landmark position since landmarks are stationary.

It is assumed that the prior distribution of the state of each sensor is known and is given by \( \phi^n_0(y^n_0) \) for sensor \( n \). The sensor dynamics are also assumed to be known and can be described by the evolution probability density function (pdf) \( \phi^n_t(y^n_t|y^n_{t-1}) \).

Similarly, the prior distribution of each landmark is given by \( \psi^m_0(x^m_0) \) for landmark \( m \). In practice, this prior is not known, but it may be estimated from data. As the landmarks are assumed to be stationary, the evolution of the pdf is known.

Let the \( r \)-th measurement at time \( t \) for sensor \( n \) be \( z^{(r)}_{tn} \) and let the source of the measurement be denoted by \( k_{tnr} \). The assignment variable \( k_{tnr} \) is often assumed to be known in SLAM problems, but for this paper it is unknown and its estimation is the data association process. At time \( t \) there are \( \eta_{tn} \) measurements from sensor \( n \). \( \eta_{tn} \) may be zero. Let \( Z^{(r)} \) denote the set of all measurements and \( K \) denote the set of all assignments.

The observation process is described by a known measurement pdf that is denoted by \( p_{zn}^{(r)}(z^{(r)}_{tn}|y^n_t, x^m_t, k_{tnr} = m) \). The particular form of the measurement pdf may be dependent on the sensors used in the application. In the case where the sensor may produce false detections, the pdf of the false detections is \( p_{zn}^{(r)}(z^{(r)}_{tn}|y^n_t, x^m_t, k_{tnr} = 0) \) and may be assumed to be uniform over the observation space in a simple case.

Let there be a known set of classes of landmarks. Each landmark has an associated class, \( \theta^m \), which is a static parameter of the landmark and will be assumed known here. In practice, the landmark class must be learned from data, e.g. [12].

Assume that each measurement has an associated classification measurement, \( z_{tnr}^{(k)} \), which is a discrete random variable taken from the known set of landmark classes and an observation of the true landmark class, \( \theta^m \). Let \( Z^{(k)} \) denote the set of all classification measurements and \( Z \) be the union of \( Z^{(r)} \) and \( Z^{(k)} \), i.e. all of the observed data.

The probability mass function of the classification measurement is referred to as the confusion matrix [15]. The elements of the confusion matrix are denoted as \( c(i, j) = P(z_{tnr}^{(k)} = i|\theta^m = j) \) with \( C = \{c(i, j)\} \). The confusion matrix is assumed to be independent of the measurement index \( r \) and time independent.

3 The PMHT with Classification in SLAM

The standard PMHT algorithm is derived in detail in [7] and [16]. PMHT is based on the application of expectation-maximisation (EM) [8] to multi-target data association. The major advantage of using PMHT is that the computational complexity increases linearly with the number of targets, measurements and time steps unlike other association algorithms which can grow exponentially. This allows the algorithm to be implemented without approximation and allows for efficient smoothing over time batches when the application requires.
The PMHT algorithm is a method for finding the best estimate of the target states, $X$, when the measurement source $K$ is unknown. It does this by treating the assignments as missing data using EM. The state estimate is derived iteratively by maximising an auxiliary function

$$Q(X|\hat{X}(i)) = \sum_K P(K|\hat{X}(i), Z) \log P(X, K, Z),$$  

where $i$ is an iteration index. Upon convergence, the algorithm’s output is the state estimate. This auxiliary function $Q(.)$ can be maximised using any appropriate estimator. It can be shown that the auxiliary function is equivalent to the log-likelihood of a known assignment problem with synthetic measurements determined by the expectation step [7]. Thus for linear Gaussian cases, the Kalman filter may be used to solve the equivalent problem. For nonlinear problems such as SLAM, a nonlinear filter must be used, such as the Extended Kalman Filter (EKF).

The following independence assumptions are made:

- All state sequences are independent of each other;
- The (unknown) true assignments are independent, identically distributed random variables with a prior probability mass $\pi^m_{tnr} \equiv P(k_{tnr} = m)$. The collection of these probabilities is denoted as $\Pi_T$;
- All measurements are conditionally independent given the assignments, the states of the landmarks and sensors, and the landmark classes.

The derivation of the PMHT SLAM with classification measurements follows the same development as the standard PMHT SLAM given in detail in [9]. The extension of classification to PMHT was introduced by [10].

In EM terminology, the complete data are $(X, Y, K, Z)$, the incomplete data are $(X, Y, Z)$ and $K$ are the missing data. The auxiliary function is the expectation of the complete data log-likelihood over the missing data, which now takes the form:

$$Q(X, Y, K, Z) = \sum_K P(X, Y, K, Z) \log P(X, Y, K, Z),$$

where the summation is over all permutations of the assignment variable $K$. Note that the confusion matrix has implicitly been assumed to be known.

For compactness, let

$$\prod_{t,n,r}(\cdot) \equiv \prod_{t=1}^{T} \prod_{n=1}^{N} \prod_{r=1}^{M}(\cdot),$$

where $i$ is an iteration index. Upon convergence, the algorithm’s output is the state estimate. This auxiliary function $Q(.)$ can be maximised using any appropriate estimator. It can be shown that the auxiliary function is equivalent to the log-likelihood of a known assignment problem with synthetic measurements determined by the expectation step [7]. Thus for linear Gaussian cases, the Kalman filter may be used to solve the equivalent problem. For nonlinear problems such as SLAM, a nonlinear filter must be used, such as the Extended Kalman Filter (EKF).

The following independence assumptions are made:

- All state sequences are independent of each other;
- The (unknown) true assignments are independent, identically distributed random variables with a prior probability mass $\pi^m_{tnr} \equiv P(k_{tnr} = m)$. The collection of these probabilities is denoted as $\Pi_T$;
- All measurements are conditionally independent given the assignments, the states of the landmarks and sensors, and the landmark classes.

The derivation of the PMHT SLAM with classification measurements follows the same development as the standard PMHT SLAM given in detail in [9]. The extension of classification to PMHT was introduced by [10].

In EM terminology, the complete data are $(X, Y, K, Z)$, the incomplete data are $(X, Y, Z)$ and $K$ are the missing data. The auxiliary function is the expectation of the complete data log-likelihood over the missing data, which now takes the form:

$$Q(X, Y, K, Z) = \sum_K P(X, Y, K, Z) \log P(X, Y, K, Z),$$

where the summation is over all permutations of the assignment variable $K$. Note that the confusion matrix has implicitly been assumed to be known.

For compactness, let

$$\prod_{t,n,r}(\cdot) \equiv \prod_{t=1}^{T} \prod_{n=1}^{N} \prod_{r=1}^{M}(\cdot),$$

if the confusion matrix is unknown, then it may be included as an extra parameter to estimate [10]. However, it has been shown that good performance may be obtained by assuming a fixed confusion matrix, even when there is a large error between the assumed confusion matrix and the truth. [17].
Thus the conditional probability of the assignments is given by the product of individual per measurement \( w_{tnr}^m \), the normalised likelihood of the \( r \)th measurement from sensor \( n \) at time \( t \) being due to landmark \( m \). The numerator of the weight is simply the product of the assignment prior, the positional measurement likelihood and the classification measurement likelihood, the last of which is an element of the confusion matrix.

Combining the two equations (4) and (9) leads to the auxiliary function to be maximised:

\[
Q(X, Y, \Pi | X^i, Y^i, \Pi^i) = \log P(X) + \log P(Y) + \sum_{m=0}^{M} \sum_{t,n,r} w_{tnr}^m \log \pi_{tn}^m
+ \sum_{m=0}^{M} \sum_{t,n,r} w_{tnr}^m \log \zeta_t^{(c)}(\tilde{x}_{tnr}^m | y_t^n, x_t^m)
+ \sum_{m=0}^{M} \sum_{t,n,r} w_{tnr}^m \log c\left(z_{tnr}^m, \theta^m\right)
\equiv Q_{XY} + Q_{\Pi} + Q_{C}
\] (10)

The term \( Q_{C} \) in (11) is given by

\[
Q_{C} = \sum_{m=0}^{M} \sum_{t,n,r} w_{tnr}^m \log c\left(z_{tnr}^m, \theta^m\right),
\]

and depends only on the classification measurements, the confusion matrix and the landmark classes. These are all known quantities, so \( Q_{C} \) is constant and can be ignored when maximising. If the confusion matrix were unknown, maximising this term would lead to its estimate.

The term \( Q_{\Pi} \) in (11) is given by

\[
Q_{\Pi} = \sum_{m=0}^{M} \sum_{t,n,r} w_{tnr}^m \log \pi_{tn}^m,
\]

and is found in the standard multi-sensor PMHT [18]. It is maximised subject to the constraint that \( \sum_m \pi_{tn}^m = 1 \) using a Lagrangian, resulting in the updated prior estimate

\[
\pi_{tn}^{m(\text{+1})} = \frac{1}{\eta_{tn}} \sum_{r=1}^{n_r} w_{tnr}^m,
\] (12)

i.e. the weights’ relative frequency.

The remaining term, \( Q_{XY} \), couples the landmark states, the sensor states and the state measurements and is given by

\[
Q_{XY} = \log P(X) + \log P(Y)
+ \sum_{m=0}^{M} \sum_{t,n,r} w_{tnr}^m \log \zeta_t^{(c)}(\tilde{x}_{tnr}^m | y_t^n, x_t^m).
\] (13)

For Gaussian measurement noise, it can be shown that this function is equivalent to the log likelihood of a SLAM problem with known data association [9],

\[
Q_{XY} = \log P(X) + \log P(Y)
+ \sum_{m=0}^{M} \sum_{t=1}^{T} \sum_{n=1}^{N} \log \zeta_t^{(c)}(\tilde{z}_{tn}^m | y_t^n, x_t^m),
\] (14)

where the synthetic measurement, \( \tilde{z}_{tn}^m \), is given by

\[
\tilde{z}_{tn}^m = \frac{1}{\eta_{tn}} \sum_{r=1}^{n_r} w_{tnr}^m \tilde{z}_{tnr}^m,
\] (15)

and the synthetic measurement function, \( \tilde{c}_{\eta}(\cdot) \), is a Gaussian distributed random variable with the same mean as the true measurement function and a variance that is a scaled version of the sensor measurement variance, \( R \),

\[
\tilde{R}_{tn} = \frac{1}{\eta_{tn}} \sum_{r=1}^{n_r} R_{tnr}.
\] (16)

Having now arrived at a known-association SLAM problem, any suitable SLAM estimation algorithm may be employed to find the state estimates, using the synthetic measurement and measurement variance as inputs. In the case where the measurement function is Gaussian but nonlinear, the EKF may provide adequate accuracy.

4 Results

The performance of PMHT-c SLAM was investigated through simulation and with real sensor data collected by the University of Sydney.

4.1 Simulated results

The effectiveness of the PMHT-c SLAM algorithm described above is now gauged through Monte Carlo simulations and compared with some alternative data association methods. These alternative data association methods include

1. Local Nearest Neighbour (LNN) : A simple association strategy where each landmark within the sensor’s range is associated with the closest measurement within a gated distance.

2. Nearest Neighbour - Joint Probabilistic Data Association (NN-JPDA) [19] also known as “cheap JPDA” : Approximate data-association probabilities are computed for each measurement and landmark, and these are used to iteratively associate landmark-measurement pairs. On each iteration, the most likely association is found, and the corresponding landmark and measurement are removed from consideration for following iterations. This approach results in a hard assignment which can then be used by the SLAM estimator.

3. PMHT : The version of the PMHT SLAM without classification is used so that direct comparisons with the improvement that classification provides can be seen.
The simulation scenario consisted of a single sensor platform moving through a field of randomly placed landmarks. To compare the data association effectiveness, the state vector is initialised with the landmark positions. The platform’s speed and turn rate were constant with Gaussian process noise with covariance $Q$, and the position and orientation were obtained by integrating them. The sensor observed landmarks with a field of view limited to 80m in range and $\pm (\pi/2)$ radians in angle. Figure 1 shows a realisation of the platform motion and the randomly generated landmarks around the platform. In the simulation scenario, the landmarks were randomly classified into two types. The circles and pluses in figure 1 represent the landmark class. 200 scans were simulated and statistics were averaged over 100 Monte Carlo realisations.

The classification measurement used for the PMHT-c SLAM simulations had a confusion matrix of the form

$$C = \begin{bmatrix} \alpha & 1 - \alpha \\ 1 - \alpha & \alpha \end{bmatrix},$$

(17)

with $\alpha = 0.9$.

The effectiveness of data association was quantified with two estimation accuracy metrics. The first was the percentage of divergent trials. Here, a trial is declared divergent if the instantaneous platform position error was more than 3m at any time during the trial. This indicates a failure of the SLAM algorithm to localise the platform. The second metric was the root mean square (RMS) position error averaged over time and over the 100 Monte Carlo trials. Divergent trials were not used for this metric.

The metrics were calculated for varying sensor accuracy, which was progressively degraded to increase data association difficulty. The sensor measurement noise variance was $(0.5\kappa)^2$ m$^2$ in range, where $\kappa$ was varied, and $(0.01\kappa)^2$ rad$^2$ in bearing. The percentage of divergent trials and the RMS estimation accuracy are plotted as a function of the measurement accuracy parameter $\kappa$ in figures 2 and 3, respectively.

The results in figures 2 and 3 demonstrate that the addition of classification information improves the performance of data association in this problem. Figure 2 shows that the PMHT-c SLAM gave 5 to 20% fewer divergent tracks as the measurement accuracy was increased when compared with the PMHT SLAM, which was considerably better than the LNN and NN-JPDA association techniques. Figure 3 also shows slight improvement in the RMS error of the platform trajectories.

Figure 4 illustrates the percentage of divergent tracks for PMHT-c as the accuracy of the simulated classification tags was reduced. The PMHT-c algorithm assumed a fixed $\alpha = 0.9$, which was thus mismatched to the true classification measurement accuracy. As expected, the performance degraded as the probability of incorrect classification tags increased. However, the change was minor up to error probabilities of as much as 30%. Even with very poor quality classification information, with only 60% probability of correct classification, the PMHT-c improved performance over PMHT with no classification input.

4.2 Victoria Park Data

The Victoria Park data set is a benchmark data set recorded by the University of Sydney. In the experiment, a utility vehicle was fitted with various sensors and driven around Victoria Park (at the University of Sydney), which contains sparsely distributed trees. The sensors onboard included a laser range finder which was used to observe the trees, and inertial sensors that provided measurements of the vehicle’s speed and steering direction. GPS data were collected to
provide ground truth. Due to occlusion, the GPS signal was not available over the whole experiment. Details about this experiment can be found in [20].

Victoria Park was modelled as a 2-D world and the problem was to map out the tree landmarks and to estimate the location of the platform as it moved through this 2-D world. The landmarks were modelled as stationary objects using a simple 2-D position state vector. The single moving platform had a state vector consisting of its 2-D position and its motion, which was approximated using a constant speed and constant turn rate model

\[
y_t = \begin{cases} 
X \text{ position} \\
Y \text{ position} \\
\text{heading} \\
\text{speed} \\
\text{steering angle}
\end{cases}_t 
\]  

(18)

with white Gaussian process noise on the speed and steering angle to account for manoeuvres. Each sensor was assumed to have white Gaussian measurement noise.

The SLAM algorithm was initialised with only the platform’s state vector with the platform’s location set at origin. New landmarks were added to the stacked state vector as they were initiated using an ad-hoc landmark initialisation algorithm similar to appendix II in [21].

The measurements from the laser range finder not only included the range and bearing of the trees, but also gave an estimate of their widths. These widths were placed into a histogram, resulting in the distribution shown in figure 5. The histogram supports the modelling of the tree population as a mixture of narrow and wide trees. Expectation Maximisation was used to fit a mixture of two Gaussian pdfs to the histogram. These two components are also shown in the figure.

A simple classifier was designed to classify the tree landmarks as either “small” or “big” trees by applying a threshold to the observed trunk width. The threshold value was the
Figure 6: PMHT-c estimated trajectory.

point where the two fitted Gaussian components intersect, as shown in figure 5.

The PMHT-c SLAM algorithm was run using the laser data only and with the associated classification data based on the above threshold. The platform trajectory estimate is shown in figure 6. The solid line marks the trajectory of the PMHT-c SLAM algorithm and the GPS reports are marked with dots. It is evident that there is error in the GPS measurements due to fluctuations in the GPS reports, sometimes with jumps of several metres due to satellite occlusions.

The RMS error between the GPS reports and the platform estimate generated using the laser data is approximately 4.5 m for PMHT SLAM and 4.39 m for PMHT-c SLAM. This is within the nominal accuracy of the GPS receiver.

5 Conclusions

This paper used the classification extension to PMHT to improve upon conventional PMHT when solving the SLAM data association problem.

Simulation experiments were used to demonstrate the effectiveness of PMHT-c for data association. Improvements were seen when compared with PMHT without classification and more common, but simpler data association algorithms, such as nearest neighbour.

The Victoria Park data was used to demonstrate the use of PMHT-c on real data with classifications estimated using laser data.

The use of classification measurements can be very beneficial for SLAM as better data association leads to more accurate landmark estimates and localisation of the platform.

References


