An Efficient Data Association Approach to Simultaneous Localization and Map Building

Sen Zhang, Lihua Xie and Martin Adams School of Electrical and Electronic Engineering Nanyang Technological University, Singapore 639798 {ps7184247i, elhxie, eadams}@ntu.edu.sg

Abstract-In this paper we present an efficient integer programming (IP) based data association approach to simultaneous localization and mapping (SLAM). In this approach, the feature based SLAM data association problem is formulated as a 0-1 IP problem. The IP problem is approached by first solving a relaxed linear programming (LP) problem. Based on the optimal LP solution, a suboptimal solution to the IP problem is then obtained by applying an iterative heuristic greedy rounding (IHGR) procedure. Unlike the traditional nearest-neighbor (NN) algorithm, the proposed algorithm deals with a global matching between existing features and measurements of each scan and is more robust for an environment of high density features which is usually the case in outdoor environments. We provide a simulation study where the NN algorithm fails whereas our proposed algorithm performs satisfactorily. Experimental results also demonstrate the effectiveness and efficiency of our approach.

Index Terms- Data association, Integer programming

I. INTRODUCTION

A feature based approach to simultaneous localization and mapping (SLAM) is to use the information obtained by sensors mounted on a vehicle to build and update a map of the environment and compute the vehicle location in that map [1], [2], [3], [4]. One of the critical problems in obtaining a robust SLAM solution is data association, i.e. relating sensor measurements to features in the map that has been built thus far [5], [6], [7], [8], [9]. Correct correspondences between the sensed feature observations and map landmarks are essential for consistent map construction since any single false matching may invalidate the entire process [10]. Simple localization may be able to recover from a minor mis-association, because only the vehicle pose estimate is affected, but with SLAM the map is also altered and these inconsistencies tend to be self-propagating, causing divergence. Hence, data association failure is a much more serious problem for SLAM than for any other localization.

There have been some approaches to data association. In stochastic mapping, the simplest method is the nearest neighbor (NN) algorithm which is a classical technique in tracking problems [11], [12], [13], [14], [15], [16]. The great advantage of NN is its O(mn) computational complexity except its conceptual simplicity. It is reliable for features where clutter density is low and sensor precision is high. However, during the process of SLAM, especially in complex outdoor environments, clutter level is high and the innovations

in the matchings of different observations obtained from the same vehicle position are correlated. In this situation, the NN algorithm may accept a wrong matching, which leads to divergence in state estimation. In order to improve the robustness of data association, Neira and Tardós [6] presented an approach using joint compatibility test based on the branch and bound search with a computational cost which is acceptable in indoor environments. Juan Nieto et al. [17] gives a fast SLAM algorithm for data association by applying multiple hypotheses tracking method in a variety of outdoor environments. But these two algorithms did not give a theoretical analysis of the algorithm complexity and experimental results demonstrated that if the observation feature number in one scan is large, the algorithm will not be fast for real time implementation. In other approaches, Baley et al. [18] considered relative distances and angles between points and lines in two laser scans and used graph theory to find the largest number of compatible pairings between the measurements and existing features. The work of Lim and Leonard [19] applies a hypotheses test to implement data association of the relocation in SLAM using geometrical constraints. Castellanos and Tardós [1] uses binary constraints to localize the robot with an *a priori* map using an interpretation tree. In these methods, geometric constraints among features are used to obtain hypotheses with pairwise compatible parings. However, pairwise compatibility doesn't guarantee joint compatibility [6], and additional validations are required.

In our work, we propose a single-frame 0-1 integer programming (IP) approach to data association. Firstly, we formulate the data association of SLAM as an integer programming problem. In order to reduce the computational burden, a validation gate is applied to reduce the size of the solution space. An iterative heuristic greedy rounding process based on linear programming techniques instead of the traditional Lagrangian relaxation algorithm [20], [21] is proposed to obtain a suboptimal solution to the integer programming problem. The algorithm has moderate computational requirement. Simulation results show that the proposed method is more efficient than the NN algorithm. In fact, for the given simulation study, the NN algorithm leads to a diverged state estimation of vehicle pose whereas the proposed algorithm performs satisfactorily. Experimental results further demonstrate the real-time applicability of the algorithm. As compared to the other existing methods aforementioned, our approach has a lower computational complexity by using a modified suboptimal optimization algorithm. Further, it can be easily extended to multi-scan cases when considering joint compatibility of matching with a much lower computational requirement than the existing methods [6] [17].

The paper is organized as follows: Section 2 is devoted to an IP formulation for the data association of SLAM. Section 3 presents an iterative heuristic greedy rounding algorithm for the IP problem. Section 4 shows the simulation and experimental results. Some conclusions are drawn in Section 5.

II. PROBLEM FORMULATION

In this section we formulate the data association of SLAM as a 0-1 integer programming problem similar to [22]. To this end, a mathematical framework of SLAM which is based on the extended Kalman filter should be firstly understood. The details can be seen in [23].

Data association of SLAM is a decision process of associating measurements (observations) with existing features in the stochastic map. It should be noted that the term "measurements" (observations) in this paper refers to the observed features after the feature extraction rather than the raw sensor measurements. Generally, the number of the measurements obtained in each scan is not equal to the number of features whose positions are estimated by the EKF. Each measurement may either (1) belong to a previously known geometric feature or (2) be a new geometric feature or (3) be a spurious measurement (also called a false alarm). On the other hand, there also exist features that do not have associated measurements in the current scan. A dummy element is applied to denote the case of a false alarm or a new start feature or a feature that does not have an associated measurement [21].

Assume that there are M measurements from the latest scan which are to be assigned to N existing features in the map built based on the previous scans. Typically, $M \neq N$. Define the binary assignment variable

$$x_{nm} = \begin{cases} 1 & \text{if measurement } m \text{ is assigned to feature } n \\ 0 & \text{otherwise} \end{cases}$$
(1)

Note that $x_{n0} = 1$ implies that the feature *n* has no associated measurement in the current scan, and $x_{0m} = 1$ implies that measurement *m* is not assigned to any of the existing *N* features, but instead, assigned to a dummy feature–false alarm or newly initialized feature. In the data association process, one measurement originates from at most one feature, and one feature can produce at most one measurement. Therefore, the following constraints can be imposed to the association variables:

$$\sum_{m=0}^{M} x_{nm} = 1, \qquad n = 1, 2, \dots, N$$
 (2)

$$\sum_{n=0}^{N} x_{nm} = 1, \qquad m = 1, 2, \dots, M$$
(3)

Our goal is to match the sensor's observations with the features by providing estimates of features' positions at the time of current scan. In order to formulate the 2-D assignment problem, a generalized likelihood ratio which involves feature state estimates for the candidate associations is used to assign costs to each association. Similarly to the multitarget tracking problem [21], we maximize a likelihood function LH as follows:

$$LH = \prod_{\{n,m\in E_{nm}\}} \Lambda(z_m, f_n) \tag{4}$$

$$\Lambda(z_m, f_n) = \frac{1}{|2\pi \mathbf{s}|^{1/2}} \exp\{\frac{1}{2} [z_m - \hat{z}_n]^T \mathbf{s}^{-1} [z_m - \hat{z}_n]\} m \neq 0, n \neq 0$$
(5)

where the likelihood ratio $\Lambda(z_m, f_n)$ denotes the probability that the *m*th measurement matches the *n*th feature in the current sensor scan. z_m means the *m*th measurement of the scan and \hat{z}_n can be calculated from the EKF estimation process. **s** is the covariance of $z_m - \hat{z}_n$. E_{nm} means all possible assignment pairs. In order to constitute the format of the 2D-assignment optimization problem, instead of maximizing the product of matching probabilities, we can minimize the negative loglikelihood ratio. To this end, define:

$$C_{nm} = -\ln \Lambda(z_m, f_n) \quad m \neq 0, n \neq 0$$

Then, an equivalent cost function for (4) can be written as follows:

$$Minimize \sum_{\{n,m \in E_{nm}\}} C_{nm} x_{nm}$$

where $C_{nm} = 0$, when m = 0 or n = 0. Thus, the data association of SLAM can be formulated as the following 0-1 integer programming problem:

$$\min\sum_{\{n,m\in E_{nm}\}}C_{nm}x_{nm}\tag{6}$$

subject to

$$\sum_{n=0}^{M} x_{nm} = 1, \quad n = 1, 2, \dots, N$$
 (7)

$$\sum_{n=0}^{N} x_{nm} = 1, \qquad m = 1, 2, \dots, M$$
(8)

where $x_{nm} \in \{0, 1\}$ and

$$C_{nm} = \begin{cases} 0 & \text{if } m = 0 \text{ or } n = 0 \\ -\ln \Lambda(z_m, f_n) & \text{otherwise} \end{cases}$$
(9)

III. THE IHGR ALGORITHM FOR INTEGER PROGRAMMING

In this section, we propose a method to solve the data association problem formulated in the last section. The method is a combined iterative heuristic greedy rounding and linear programming. In order to reduce the computational burden, a validation gate is applied first to reduce the above global association into several local associations.

A. Gating

In order to reduce the solution space, a gating is applied before applying the LP. Only measurements that are close enough to the prediction state of an existing feature are considered possible candidates of association with the feature. The criterion of gating is given by:

$$\tau_{ij} = \nu_{ij}^T s_i^{-1} \nu_{ij} \le \varepsilon$$
$$i = 1, 2, \cdots, N; \quad j = 1, 2, \cdots, M$$

where

$$\nu_{ij}(k+1) = \mathbf{z}_j(k+1) - \hat{\mathbf{z}}_i(k+1)$$

and s_i is the covariance of the innovation ν_{ij} .

Note that since ν_{ij} is a Gaussian random variable, τ_{ij} is a random variable following the χ^2 distribution. Thus, a validation gate, ε , is used to decide whether the measurement $z_j(k+1)$ is a close enough match to the predicted feature position. From the χ^2 distribution table, we know that $\tau_{ij} < 6.63$ with a probability of 0.99. Here we set $\varepsilon = 6.63$.

B. Iterative Heuristic Greedy Rounding

In order to improve the computation speed, here we use an iterative heuristic greedy rounding procedure to solve the 0-1 IP problem.

By changing the integer constraint $x_{nm} \in \{0,1\}$ to $0 \leq 0$ $x_{nm} \leq 1$, the IP problem is relaxed to a linear programming (LP) one. The LP problem can be solved by basic LP algorithms, such as the Simplex algorithm [24]. If the optimal solution x_{op} of the LP-relaxation is fully integer-valued (in this case all decision variables will have the value of either 0 or 1) then the solution x_{op} is optimal for the 0-1 IP problem (6) [25]. Otherwise, we apply a procedure called iterative heuristic greedy rounding (IHGR) (see, e.g. [26]). Observe that the larger the decision variable x_{nm} , the higher the probability that the m-th measurement associates with the n-th feature. Hence, the algorithm starts with setting the maximum decision variable (with a value close to 1) to 1 and all other entries in the same row and column to zero to meet the constraints (7) and (8). Then, solve the LP problem for the rest of the assignment matrix and repeat the IHGR procedure to decide the next pairing of measurement and feature. The process is carried on until all measurements have been assigned. In this manner, a feasible (but not necessarily optimal) solution for the original IP is constructed.

If there is no predominant variable that is near to 1, i.e. a tie occurs, then we will calculate all the possible combination cost of (6), and select the lowest cost. In the IHGR procedure, when x_{nm} is set to 1, all variables in the column and row associated with the specific set in E_{nm} must be set to 0. Once a variable is forced to a certain value, it is not allowed to change any more. To achieve this, all rounded variables and all implicated variables are discarded from the IHGR procedure. In this way, the IHGR will never set the value of a variable twice. This deletion of variables also applies to the initial LP solution, i.e. all variables with value 1 and all

zero-valued variables implicated by them, are removed. The IHGR algorithm repeats the actions of selection, rounding and deletion until there are no variables left. The outcome will then be a feasible solution to (6). Observe that the IHGR can be implemented efficiently, but it is clear that there is no guarantee that this heuristic procedure yields the optimal solution of the IP problem. However, the experiments to be discussed in the following section show that the IHGR does result in acceptable feature-measurement assignments of which the achieved cost is close to the optimal cost.

C. Algorithm Complexity

Due to the application of the gating process that is affected by random factors, we cannot give an exact description of the complexity. However, we know that in any fixed dimension, linear programming can be solved in strongly polynomial linear time (linear in the input size) [27]. For our case, the input size is $M \times N$. Thus, we can roughly know the worstcase complexity of the proposed algorithm is $O(MN + (M - 1)(N-1) + \ldots + (|M-N|+1) \times 1)$ for the IHGR process.

Neira and Tardos [6] presented a data association approach using joint compatibility test based on the branch and bound search (JCBB). JCBB performs incremental construction and search of an interpretation tree of joint association hypotheses. The gating determines acceptable hypotheses and performs branch and bound pruning of the search space. The discussion in [6] does not provide any theoretical bound, but gives an empirical complexity estimate of $O(1.53^N)$, where N is the number of observed features. But for the IHGR algorithm, the worst-case computational burden is $O(MN \times \min\{M,N\})$ without considering the gating. Therefore, when the observed feature number is large (such as more than 30), our algorithm can work much more efficiently than the JCBB.

IV. EXPERIMENTAL RESULTS

The algorithm presented was tested in two different environments, an artificial environment and a real outdoor environment. In these two environments, the SLAM was implemented by using the data association algorithm proposed in the last section. The experimental results show that the algorithm is efficient in SLAM.

A. Simulation environment

The first test environment is established by randomly generating some features and assuming the vehicle trajectory along a circle whose radius is 5 meters. The robot moves at a constant speed and the heading angle changes 1 degree at each sampling instant. We choose 12 features and their locations (the x,y coordinates) are random numbers between 0 and 20 meters in global coordinates.

For 200 times we change the features' positions randomly at each time and apply the NN data association algorithm and the IHGR data association method to perform the SLAM process, respectively. The successful data association rate when using the IHGR algorithm is 97.5% (195/200) while it is only 80.5% (162/200) for the NN algorithm. When the features



Fig. 1. The unsuccessful mapping when applying NN data association algorithm in the SLAM process.



Fig. 2. The mapping and vehicle path when applying IHGR data association method in the SLAM process.

are closely located, the NN algorithm fails. Figure 1 and Figure 2 show the SLAM results for the NN algorithm and the proposed IHGR algorithm. In this case, the positions of 5 features in the environment are fixed and they are located at (2.9, 2.3); (2.6, 1.9); (2.4, 1.9); (2.5, 2.6); (2.7, 2.8), respectively. The remaining 7 features are randomly distributed. It can be observed from Figure 1 that the NN algorithm leads to diverged estimates of vehicle pose and feature positions. On the other hand, our IHGR method performs very well as observed from Figure 2. In fact, the vehicle's truth path is almost overlapped with the estimated one during the SLAM process.

A comparison on execution time between the NN algorithm and the IHGR algorithm versus the number of features is shown in Figure 3 (the algorithms are run on Pentium IV PC, 1.7GHz) for the cases when the NN algorithm is able to give a successful SLAM. In the simulation, we extract data at each scan under different feature number and assume that the number of measurements is the same as that of the existing features (this is the most time consuming case). The result shows that our IHGR is implementable in real-time application.

B. Real outdoor environment

In order to implement the IHGR data association algorithm in a SLAM process for a real environment, we use the experimental data set from [28] which is obtained by Guivant and Nebot. The testing site is a car park at Sydney University. The vehicle is equipped with GPS, laser sensor and encoders. A kinematic GPS system of 2 cm accuracy was used to evaluate the ground truth. Thus, the true navigation map was available for comparison purpose. Wheel encoders give an odometric measurement of the vehicle location. The dead reckoning sensors and laser range sensor are combined together to predict the vehicle's trajectory using extended Kalman filter and to build up the map at the same time. Note that the most commonly encountered feature in an outdoor environment is tree trunks. The feature detection was done by using a geometric analysis of the range measurement to obtain the most likely center of the tree trunk. The laser scans are processed using Guivant's algorithm to detect tree trunks's center and estimate their radii.

We run continuous SLAM for about 5000 time steps and obtained a map shown in Figure 4 using our proposed IHGR data association during the SLAM process. It can be seen that the IHGR method performs well for the SLAM implementation in real time. Figure 5 and Figure 6 gives the measurement (range and angle) innovation and their 2σ confidence bounds during the SLAM process, respectively. The results show that the IHGR algorithm works well during the SLAM process.

In order to check the effectiveness of the IHGR data association, we randomly choose two scans (scan 68 and scan 87) and show the matching matrix x_{nm} after the IP problem is solved. In scan 68, the laser sensor obtained 2 measurements and the existing feature number has accumulated to 11. As mentioned, the term "measurement" means the extracted features. Measurement 1 is associated with feature 3 and measurement 2 is associated with feature 2. The rest of the features are all undetected in this scan. In table 2, for scan 87, measurement 1 is matched with a dummy element which



Fig. 3. The mean execution time of the IHGR algorithm and the NN algorithm. The mean execution time of IHGR is nearly linear with respect to the observed feature number by repeated experiments



Fig. 5. The range innovation and its 2σ confidence bounds during SLAM process with IHGR data association



Fig. 4. The SLAM path and the feature map during SLAM process with IHGR data association



Fig. 6. The observation angle innovation and its 2σ confidence bounds during the SLAM process with IHGR data association

TABLE I

THE MATCHING MATRIX IN SCAN 68

	Existed feature number and dummy element													
		1	2	3	4	5	6	7	8	9	10	11	0	
scan 68	1	0	0	1	0	0	0	0	0	0	0	0	0	
	2	0	1	0	0	0	0	0	0	0	0	0	0	
	0	1	0	0	1	1	1	1	1	1	1	1	0	

TABLE II The matching matrix in scan 87

		Existed feature number and dummy element														
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	0
Scan 87	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
	2	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
	3	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
	0	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0

means this is a new feature or false alarm. From the property of laser sensor, we regard it as a new feature. The other measurements and features are similar to those in scan 68 which can be seen in the above table.

V. CONCLUSIONS

The popular NN algorithm for data association is very sensitive to the high density feature situation and the increase in vehicle and sensor error. This paper presented a new data association algorithm for SLAM which is more effective than the NN algorithm in complex case. We first formulated the data association problem in SLAM as a 0-1 IP problem. In order to obtain a fast solution, the 0-1 IP problem is firstly relaxed to a LP. Then we proposed to use the IHGR procedure in conjunction with basic LP algorithms. The proposed algorithm has been shown to be robust and performed very well as compared to the commonly used NN algorithm. It has also been demonstrated that the IHGR algorithm has low computational complexity and is suitable for real time SLAM. Simulation and experimental results have proven its efficiency and effectiveness.

ACKNOWLEDGMENT

The authors would like to thank Professor S. Wijesoma, Mr. Fan Tang and Mr. P. Lochana for helpful discussion on data association.

REFERENCES

- J. A. Castellanos, J. M. Montiel, J. Neira, and J. D. Tardós, "The spmap: A probabilistic framework for simultaneous localization and map building," *IEEE Trans. Robot. Automat.*, vol. 15, pp. 948–953, 1999.
- [2] J. J. Leonard and H. F. Durrant-Whyte, "Simultaneous map building and localization for an autonomous mobile robot," in *Proc. IEEE Int. Wkshp Intelligent Robots and Syst.*, Osaka, November 1991, pp. 1442–1447.
- [3] D. Fox, "Kld-sampling: Adaptive particle filters and mobile robot localization," Advances in Neural Information Processing Syst., vol. 14, pp. 26–32, 2001.
- [4] S. Zhang, L. Xie, and M. D. Adams, "Geometrical feature extraction using 2d range scanner," in *Proc. of the Fourth Int. Conf. Contr. and Automat.*, Montreal, Canada, June 2003, pp. 901–905.

- [5] J. J. Leonard and H. F. Durrant-Whyte, "Application of multi-target tracking to sonar based mobile robot navigation," in *Proc. IEEE Int. Conf. Decision Contr.*, Chicago, December 1990, pp. 3118–3123.
- [6] J. Neira and J. D. Tardós, "Data association in stochastic mapping using the joint compatibility test," *IEEE Trans. Robot. Automat.*, vol. 17, pp. 890–897, 2001.
- [7] J. Guivant, E. Nebot, and H. F. Durrant-Whyte, "Simultaneous localization and map building using natural features in outdoor environments," in *Proc. IEEE Int. Conf. on Intelligent Autonomous Syst.*, Venice, Italy, November 2000, pp. 581–588.
- [8] T. Bailey and E. M. Nebot, "Localisation in large-scale environments," *Robotics and Autonomous Systems*, vol. 37, pp. 261–281, 2001.
- [9] J. A. Castellanos, J. Neira, and J. D. Tardós, "Multisensor fusion for simultaneous localization and map building," *IEEE Trans. Robot. Automat.*, vol. 17, pp. 908–914, 2001.
- [10] H. F. Durrant-Whyte, S. Majunder, M. Batista, and S. Scheding, "A bayesian algorithm for simultaneous localisation and map building," in *Int. Symposium of Robotics Research (ISRR01)*, Lorne Victoria, Australia, November 2001, pp. 3118–3123.
- [11] J. Guivant and E. M. Nebot, "Optimization of the simultaneous localization and map building algorithm for real time implementation," *IEEE Trans. Robot. Automat.*, vol. 17, pp. 242–257, 2001.
- [12] J. Guivant, E. M. Nebot, and S. Baiker, "Localisation and map building using laser range sensors in outdoor applications," *J. Robot. Syst.*, vol. 17, pp. 565–583, 2000.
- [13] J. Guivant, E. Nebot, and S. Baiker, "Simultaneous localization and map building using natural features and absolute information," *Robotics and Autonomous Systems*, vol. 40, pp. 79–90, 2002.
- [14] J. J. Leonard and H. F. Durrant-Whyte, "Dynamic map building for an autonomous mobile robot," *Int. J. Robot. Res.*, vol. 11, pp. 286–298, 1992.
- [15] M. D. Adams, S. Zhang, and L. Xie, "Paticle filter based outdoor robot localization using natural features extracted from laser scanner," in *Proc.* of *IEEE Int. Conf. Robot. Automat.*, New Orleans, USA, April 2004, in press.
- [16] J. J. Leonard and H. S. Feder, "Decoupled stochastic mapping," *IEEE J. Oceanic Engineering*, vol. 26, pp. 561–571, 2001.
- [17] J. Nieto, J. Guivant, E. Nebot, and S. Thrun, "Real time data association for fastslam," in *Proc. IEEE Int. Conf. Robot. Automat.*, Taiwan, September 2003, pp. 2512–2517.
- [18] T. Bailey, E. M. Nebot, J. K. Rosenblatt, and H. F. Durrant-Whyte, "Data association for mobile robot navigation: a graph theoretic approach," in *Proc. IEEE Int. Conf. Robot. Automat.*, San Francisco, California, April 2000, pp. 2512–2517.
- [19] J. H. Lim and J. J. Leonard, "Mobile robot relocation from echolocation constraints," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 22, pp. 1035–1041, 2000.
- [20] A. B. Poore and A. J. Robertson, "A new class of langrangian relaxation based algorithms for a class of multidimensional assignment problems," *Computational Optimization and Applications*, vol. 8, pp. 129–150, 1997.
- [21] A. B. Poore, "Multidimensional assignment formulation of data association problems arising from multitarget tracking and multisensor data fusion," *Computational Optimization and Applications*, vol. 3, pp. 27– 57, 1994.
- [22] P. L. D. Lochana, S. Wijesoma, and M. Adams, "Simultaneous localization and map building," in *Proc. SCIRAS*, Singapore, December 2003, in press.
- [23] G. Dissanayake, P. Newman, S. Clark, H. F. Durrant-Whyte, and M. Csorba, "A solution to the simultaneous localization and map building (slam) problem," *IEEE Trans. Robot. Automat.*, vol. 17, pp. 229–241, 2001.
- [24] S. Vajda, *Linear programming: algorithms and applications*. Chapman and Hall, 1981.
- [25] R. G. Parker and R. L. Rardin, *Discrete optimization*. Acdamic Press, 1988.
- [26] J. L. Miller and L. S. Franz, "A binary rounding heuristic for multiperiod variable-task duration assignment problems," *Computers and Operations Research*, vol. 23, pp. 819–828, 1996.
- [27] N. Karmarkar, "A new polynomial-time algorithm for linear programming," *Combinatorica*, vol. 4, pp. 373–395, 1984.
- [28] J. Guivant and E. Nebot, ACFR, Experimental outdoor dataset. http://www.acfr.usyd.edu.au/homepages/academic/enebot/dataset.htm, 2003.