Learning Decision Trees for Mapping the Local Environment in Mobile Robot Navigation

Ian Sillitoe
Department of Electronic and Electrical Engineering
Loughborough University of Technology
Loughborough, Leicestershire, LE11 3TU, U.K.
i.p.sillitoe@lut.ac.uk

Tapio Elomaa
Department of Computer Science
P. O. Box 26 (Teollisuuskatu 23)
FIN-00014 University of Helsinki, Finland
elomaa@cs.helsinki.fi

Abstract

This paper describes the use of the C4.5 decision tree learning algorithm in the design of a classifier for a new approach to the mapping of a mobile robot's local environment. The decision tree uses the features from the echoes of an ultrasonic array mounted on the robot to classify the contours of its local environment. The contours are classified into a finite number of two dimensional shapes to form a primitive map which is to be used for navigation. The nature of the problem, noise and the practical timing constraints, distinguishes it from those typically used in machine learning applications and highlights some of the advantages of decision tree learning in robotic applications.

1 INTRODUCTION

Ultrasonic sensors are often used to measure distances and in the majority of cases the *time of flight* of an ultrasonic pulse of energy is utilized. That is, the total time taken for the pulse to leave the transmitter, reach the object in question, and then on the final leg of its journey, the time for the reflections wave front to reach the receiver. When the total time is divided by the speed of sound an estimate of the shortest distance travelled by the pulse can be obtained. There are many variants of this technique in which a single transducer is used as both the transmitter and receiver or where there are arrays of both receivers and transmitters. Much attention has been placed upon the accurate determination of the beginning of the received pulse, since this plays an important role in maximising the accuracy of such a system (Barshan and Kuc, 1992).

These approaches essentially give a *spot reading*; i.e., they return the distance to a given spot on the object. In order to allow a time-of-flight ultrasonic system to produce a more detailed map of the surface of the object the sensor spot must be moved across the object. This can be done in either of two ways: *mechanically*, where the sensor is moved relative to the object (Kuc and Bozma, 1991), or through the use of a *phased array technique*, whereby the relative phase of a number of transmitters is altered so that the combined wave front is shifted. The mechanical procedure requires that the sensor is moved and reading taken many times during a single scan. Hence, the procedure is limited to the mapping of environments which remain static during this long sampling period. Whilst, the phased array approach is based upon electronic scanning and so its response time is dominated by the much shorter time-of-flight measurement. The approach, however, places stringent requirements on the complex hardware used to both generate and process the transmitted and received waveforms.

This paper describes an alternative approach to the use of ultrasonic sensors for the classification of surfaces, one in which both the hardware and software requirements are extremely modest and which only requires the analysis of a single pulse. The technique uses two receivers and a single transmitter which are arranged so that their active volumes almost completely overlap and, hence, the two receivers 'see' the same 'illuminated' volume in front of the sensor array (see Figure 1). Thus, this system is similar in its physical arrangement to the 'Bat Sonar' system (Barshan and Kuc, 1992), which used time of flight to identify the position of objects within its working volume. However, it differs in three significant features: firstly, the whole echo returned by the object is processed, not only the time-of-flight measurement, secondly, the objective of this system is to classify the shape of the object as well as determining its position, and thirdly, this approach constraints the objects to be 'close' to the sensor. These characteristics make the approach useful for rapid mapping of the local environment as required in mobile robot applications.

In addition to introducing the new approach to the use of ultrasonic sensors for the classification of surfaces, we also describe, in this paper, how machine learning techniques have been used to facilitate feature extraction for the classifier. We report some of our preliminary experiences on using the C4.5 algorithm (Quinlan, 1993) to construct decision tree classifiers from feature vectors recorded for example objects exposed to the sensor system. In our experience, decision tree learning seems to suit the purpose very well.

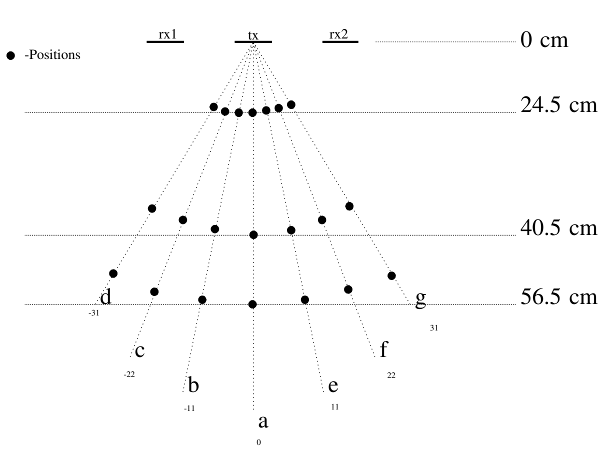

Figure 1: The sensor and the training positions for the objects within the active volume.

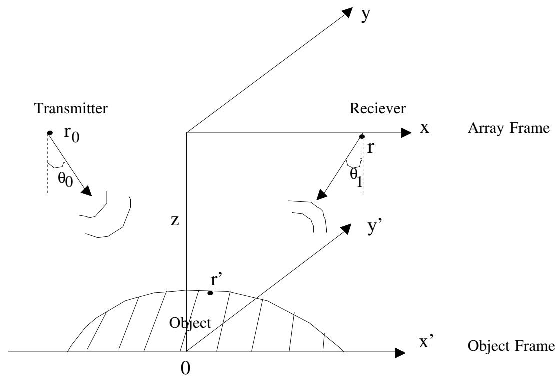

Figure 2: Single transmitter/receiver set-up.

2 THE PROBLEM DOMAIN

Using Huygens approximation for wave behaviour a single wave front can be regarded as a large number of smaller spherical waves. Each of these wavelets can be traced during a reflection and the superposition of the resulting wave front can be used to determine the effect of the reflection upon the original complete waveform. The further a wavelet must travel, the longer is its time of flight and, thus, wavelets, which are the result of refections from different parts of the surface, will appear to be shifted in time at the receiver. A second effect attenuates amplitude of the reflected pulse as a function of the distance. Watanabe and Yoneyama (1992) have formalised these for a single transmitter and receiver as shown in Figure 2, where the incident wave is transmitted at $t = 0$ and the pressure of the wave $P_{in}(\mathbf{r}, t)$ at $\mathbf{r}$, when the position of the transmitter is given by $\mathbf{r}_0 = (x_0, y_0, z_0)$ and where $|\mathbf{r} - \mathbf{r}_0|$ is large in comparison with the wavelength, is

$$P_{in}(\mathbf{r}, t) = \Theta(ct - |\mathbf{r} - \mathbf{r}_0|) \exp(j\mathbf{k}_{in} \cdot (\mathbf{r} - \mathbf{r}_0) - j\omega t),$$

where $\mathbf{k}_{in} = (k \sin\theta_0, 0, -k\cos\theta_0)$, and where c is the velocity of sound, ω the angular frequency, k the wave number, and θ_0 the illuminating angle. $\Theta(x)$ is a step function that is used to represent the time of flight of the pulse. The value of the function is 1, when $x \geq 0$, and 0, otherwise. The reflection coefficient of the surface is $\xi(x', y')$ and surface function is

$$z' = \zeta(x', y')$$

The distance $L(\mathbf{r}, \mathbf{r}')$ from the transmitter $\mathbf{r}_0$ through the surface point $\mathbf{r}' = (x', y', \xi(x', y'))$ to the receiver at $\mathbf{r} = (x, y, z)$ is given by

$$L(\mathbf{r}, \mathbf{r}') = |\mathbf{r}_0 - \mathbf{r}'| + |\mathbf{r}' - \mathbf{r}|$$

Finally, using the Kirchhoff approximation we obtain the complex pressure at the receiver

$$P(\mathbf{r}, t) = \frac{j\exp(jkr)F(\mathbf{r})}{4\pi r} \int_{-\infty}^{\infty} dx' \int_{-\infty}^{\infty} dy' \exp(j\mathbf{V} \cdot \mathbf{r}')$$
$$\cdot \xi(x, y)\Theta(ct - L(\mathbf{r}, \mathbf{r}')),$$

where

$$\begin{aligned}
\mathbf{r} &= (x, y, z) \\
r &= |\mathbf{r}| \\
\mathbf{r}' &= (x', y', \zeta(x', y')) \\
\mathbf{V} &= (V_x, V_y, V_z) \\
V_x &= -k\left(\frac{x}{r} - \sin\theta_0\right) \\
V_y &= -k\left(\frac{y}{r}\right) \\
V_z &= -k\left(\frac{z}{r} + \cos\theta_0\right) \\
F(\mathbf{r}) &= \frac{|\mathbf{V}|^2}{V_z}
\end{aligned}$$

In summary, the variation in amplitude of the received waveform holds information about the three dimensional contour of surface, whilst the delay in the pulses reception indicates the closest point of the object. However, the information explicitly describing the contour is not easily extracted. Watanabe and Yoneyama (1992) describe a successful attempt to extract this information for a different application; the approach used acoustic holographic techniques, which required the calculation of inverse Fourier transform. Neural networks were used to compensate for non-linear effects and to the final recognition of particular taught objects. This method was confined to recognising a particular set of taught example objects, which were further constrained to remain a fixed distance from the sensor.

The objective of the sensor system described here is to be able to classify untaught contours into a finite number of classes, which are chosen to be suitable for decision making by a mobile robot when navigating through a semi-structured environment (i.e., laboratory or office). Thus, the system must be efficient and robust in the face of environmental noise (such as temperature variation etc.) as well as the variety of surfaces it will encounter. The approach taken by the system is to calculate the range and azimuth to the nearest point within the sensors working volume, (to the degree of accuracy required for use in navigation these can be determined efficiently from simple time of flight measurements), independently classify the amalgamated contour of the objects within the scene into one of the taught classes, and then form a triplet which describes the scene immediately in front of the sensor. The triplet and the output of other sensor systems is then to be used in a behavioural navigational system.

The expression for $P(\mathbf{r}, t)$, as it stands, does not express a number of processes which have their effect in the shape of the echo (see Hickling and Marin (1986) for a detailed description of some of these effects). These include multiple reflections between the object and the robot, the dominant effect upon the shape of the echo by corner reflections between the object and floor as the object is moved further from the sensor, from fluctuations due to air velocity and temperature, and reverberation within the array. It is not possible to efficiency model these effects without *a priori* knowledge or assumptions about the environment.

Hence, rather than take a model-based approach we use a signature analysis method, which classifies the echoes directly as the result of a reflection from one of the primitive group of contours. This has the additional advantage in bypassing the time consuming pre-processing stages required to transform the echoes into a spatial representation, which is then re-interpreted in terms of the goals of navigation algorithm.

Thus, we have swapped the problems of the model-based approach for those of feature selection and classification learning in an ill-posed problem. It is well recognised that the selection of relevant features is often crucial to the tractability of classification problems; too few, irrelevant, or coupled features can have similar effects as overspecification, leading to ill-conditioning of the problem and difficulty in convergence. The following sections describe the rationales for the approaches taken in learning and feature selection.

3 APPROACHES TO LEARNING

Two types of inductive learning programs dominate the literature. *Connectionist* approaches (i.e. neural networks) which work at the sub-symbolic level and compile the interconnections observed in training data into weights of inter-node links in the pre-specified network architecture. Using this approach, it is practically impossible to articulate the function computed by a neural network; it can only be observed by the network's action. *Symbolic* approaches, on the other hand, emphasize the intelligibility of the data representation at every stage. Several different representation formalisms are used to describe the knowledge extracted; for example, production rules, decision trees, lists, and graphs are generally used.

Contrasting the performance of symbolic and connectionist approaches is not a new theme and has been covered extensively. Atlas *et al*. (1990), Dietterich, Hild, and Bakiri (1990), Fisher and McKusick (1989), Fisher *et al*. (1989), Mooney *et al*. (1989), Shavlik, Mooney, and Towell (1991), and Weiss and Kapouleas (1989) are seven examples of early empirical studies that have set the tone of the general understanding of the approaches relative strengths and weaknesses.

The consensus of these seven studies is that the two approaches are more or less equivalent on prediction accuracy, but when noise is added the connectionist approaches perform significantly better. The symbolic approach turns out to be superior in training time, while in classification both approaches are equally efficient. Moreover, neural networks typically require much more examples than decision trees before converging.

Subsequent comparative studies (e.g., Thrun *et al*., 1991) have stressed the connectionist approach's additional advantage of being able to construct, if not express, new (sub-symbolic) features from the primitive ones. This property is lacking from the basic symbolic approaches, even though specific algorithms with feature construction capabilities now exist (see e.g., Pagallo and Haussler, 1990). However, a recent study by Craven and Shavlik (1993) demonstrates that even back-propagation fails in constructing quite simple new features.

In connectionist approaches it is often difficult to determine the relative significance of individual features and the use to which they are put. Hence, it is possible to generate features which are redundant or detrimental to the learning process. In turn, this has an effect upon the efficiency of the final classifier, since it is possible that, under these circumstance, it will spend additional time processing unnecessary features. As the processing of sensory information is a precursor and tightly bound to the majority of the decision making processes in robotic systems, this additional time can limit the performance of the whole system. This phenomenon can take on extreme forms, e.g., in stereoscopic image processing, where the interval taken to process images can be of the order of tens of seconds, during which time the robot must remain stationary. Thus, it is often necessary to tailor the form of the classification process to the timing constraints imposed by the system of which it is a part.

The extent to which a particular feature is used in classification and hence its relative usefulness, is a more subtle problem. It will be seen that topologically dissimilar obstacles at particular poses can give rise to similar echo shapes and so after limited feature extraction can be classified identically. The obvious solution to this problem is to extend the feature set so that the additional features represent the

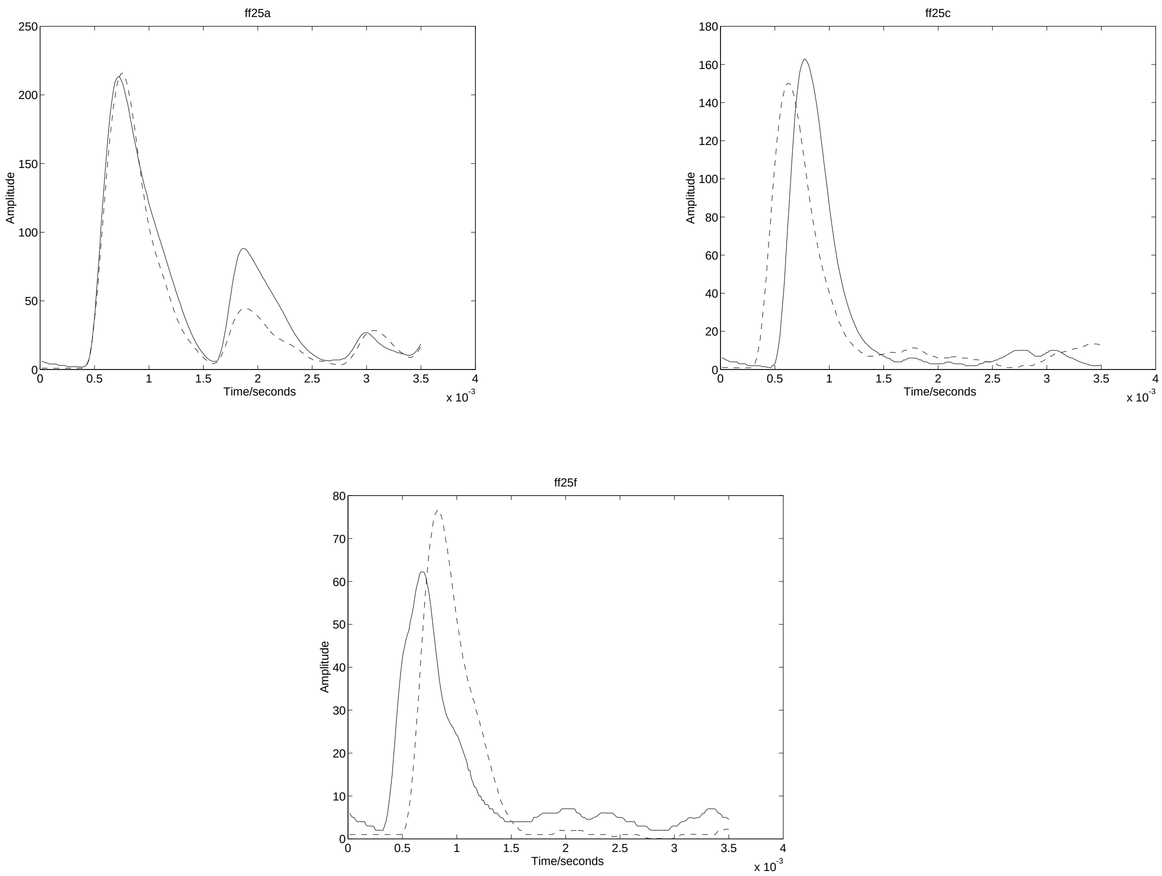

Figure 3: Typical echoes taken from a flat surface at positions a, c, f and at distance of 25 cm.

finer dissimilarities between the pathological cases. However, sensor signals are often noisy and have significant legitimate variation due to the physical process which they sense. Thus the features which represent the finer detail of a signal often are most sensitive to its fluctuations (this effect is well known for features based upon high order statistics) and so there is a limit to the additional features which can be added to the set before classification performance declines.

Thus, the design of a classifier for sensor signals is in general iterative, a compromise, its operation is time constrained, and it is linked to the characteristics of the method by which the signal is obtained.

We chose to use a symbolic learning algorithm in the feature selection for the contour classifier in order to exploit the explicit nature of the generated knowledge representation. In particular, decision tree learning was selected because of its well-known advantages. Automatically producing decision trees that minimise the average cost of instance classification in cases where the evaluation of a feature has a natural cost has also been studied (Núñez, 1988; Tan and Schlimmer, 1990). Rather than trying these approaches or some separate feature selection technique (e.g., Kira and Rendell, 1992), we chose to use Quinlan's (1993) C4.5 and manual picking out of the final features in order to retain full control over this crucial process. C4.5 is known to be among the best decision tree learners and it allows transforming the often unconveniently large decision trees into more easily handled production rules.

4 FEATURE EXTRACTION

The selection of expressive continuous features is notoriously difficult. The approach taken here was to duplicate the features for each channel and to logically partition them into three groups each of which is dominated by particular properties of the signal. The first contains direct measurements which are influenced by the scale of the signal (`start`, `finish`, `maxvalue`, `maxtime`). `Start` is the time interval between the transmission of the ultrasonic pulse to the point at which the received pulse reaches 5% of its maximum value. `Finish` is the time interval taken for the echo to return to the 5% value. `Maxvalue` is the maximum amplitude of the signal and `maxtime` is the time interval taken for the signal to reach the `maxvalue`.

The second group consists of features, which are invariant to translation and scale. The group is formed from five normalised central moments of the waveform (n_{0q}) beginning with the third order moment. This group in effect describes the approximate shape of the waveform; it is calculated as below, where we assume that the amplitude of the wave-

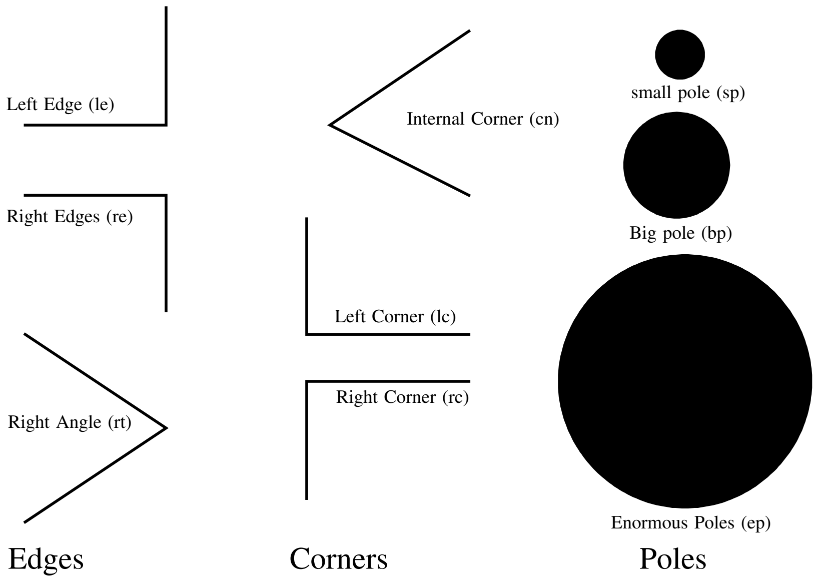

Figure 4: Training object types. The array faces into the diagram from the right.

form is described by the function $y = f(t)$ over the closed interval [`start`,`finish`].

$$\begin{aligned} m_{0q} &= \sum_y y^q \\ \mu_{0q} &= \sum_y (y - \overline{y})^q, \text{ where} \\ \overline{y} &= \frac{m_{01}}{m_{00}}, \text{ and} \\ \eta_{0q} &= \frac{\mu_{0q}}{\mu_{00}^{\gamma}}, \text{ where} \\ \gamma &= \frac{q}{2} + 1 \text{ for } q = 2, 3, \ldots \end{aligned}$$

The final group, `skew` and `kurtosis`, are effected by both scale and the shape. In all there were initially 22 features representing an echo, 11 for each channel.

5 GENERATION OF THE TRAINING DATA

The sensor comprises of a transmitter and two receivers, although it is intended that it should be eventually generalized to three dimensional array. The transmitter is a small piezo-electric element that produces high frequency pressure waves at 40 kHz. It is operated in pulsed rather than continuous mode and the duration of one pulse is 200 μs long. The full aperture angles of both the receiver and the transmitter is approximately 60°. The distances between the transmitter and receivers is 7 cm and the array is mounted on the robot at a height of 12 cm above the floor.

The received signal is rectified and filtered with a fifth order low pass Bessel function filter with its -3 dB point at approximately 7 kHz, the resulting waveform is digitised using an 8 bit ADC ($\pm\frac{1}{2}$ LSB) to produce a unsigned representation of the amplitude. The sampling rate is one sample every 14 μs with 250 samples in each signal, giving the sensor a range of approximately 60 cm at 20° C. Thus the echoes from objects at the extremes of the active volume were truncated and the sensors range limited to the immediate neighbourhood of the robot (i.e., 15–60 cm and $\pm 30°$ about the focal axis of the transmitter). Figure 3 shows typical mean echoes for a 'flat surface' at various positions within the active volume.

The echos can be partition into three separate regions. The first begins at $t = 0$ and lasts for approximately the first 0.25 ms, it is the result of reverberation within the sensor and since it is independent of the reflectors shape, it is not part of the echo. The second region contains the majority of the power and is the primary echo, whilst the latter variations are due to multiple reflections of the primary echo. Since the parasitic regions of the echo are difficult to isolate from the primary echo precisely for more complex reflectors, the signal is regarded as a time bounded signature rather than a direct measure of an individual property of the local environment.

The training data was obtained by placing objects, which were to act as representatives of particular generic classes, at different positions within the active volume of the sensor. Twenty-five echoes for each of the positions in Figure 1 were recorded for each object shown in Figure 4, where the array is facing into the diagram from the right. The dimensions of the edges and corners were chosen so that they filled the active volume at any position, the diameters of poles were 5.2, 10.5 and 37 cm, respectively. The

Table 1: Generated decision tree sizes and accuracies on classifying untaught scenes

Tree	Attributes	Tree size	Total error
1	22	139	0.1%
2	16	171	0.2%
3	14	189	1.5%

generic classes were chosen to represent components of surfaces that the mobile robot would be likely to encounter in a laboratory environment (i.e. corners, edges, and flats). Furthermore, a selection of poles that would be used to classify untaught examples encountered by the robot and not recognised as corners etc. were exposed to the sensor.

Two groups of test data were also taken: one in which the same objects used in the training data were present, but at different positions and orientations to the sensor, and a group of objects which did not belong to the taught class (including toppled waste paper bins and chairs etc.). In particular, this second group included a subgroup which had more than one object in the active volume and so provided a test for more complex scenes (e.g., legs of a chair).

6 RESULTS

C4.5 was used to create three decision trees to investigate the necessary features for the classification of the training data and the trees' relative performance when applied to the test data. The first tree was generated using all features (i.e. 22 attributes), the second was generated without the `start`, `finish`, and `maxpos` intervals (i.e. 16 attributes), while the final tree used only the second and third groups of features (i.e. 14 attributes). The tree sizes (total number of nodes) after pruning and their classification errors are shown in Table 1.

Since, the moment-based features are averaged values, the lower order statistics tend to be more immune to noise (such as quantisation noise etc.). Hence, the function of trees 2 and 3 was to allow the investigation of the effects of the scale sensitive features upon classification performance.

When the trees were translated into production rules, it was apparent that the small error in Tree 1 was related to the use of `maxvalue` at low signal-to-noise ratios, where it misclassified 'small' and 'big poles' at the extreme of the sensors' range. This was also the case for Tree 2, but it additionally misclassified 'flat surfaces' as 'enormous poles' at the maximum range. Removing the `maxvalue` from the feature set in Tree 3 resulted in additional misclassifications in which 'edges' and 'small poles' were confused at the larger angles of azimuth.

Even though removing the three scale-sensitive features from the feature set does not result in dramatic increase in classification error, when moving from Tree 1 to Tree 2, there is noticeable increase in the tree's size. In other words, deleting the three features can be compensated to an extent by the other features, but not without a cost. In many cases a combination of the remaining features is needed in place of a deleted feature. Removing the final feature of the first group results in further increase in tree size. This time the feature's absence cannot be fully compensated.

When analysing the test data results it is only possible to declare a classification as incorrect if the decision is dramatically different from the expected shape (e.g., a convex rather than concave shape is chosen), since it the objective of system to bound the scene rather than describe it in every detail. Thus, when considering this fact, each of the trees performed similarly and not one of them failed to bound the scene within the arc of the active volume of the sensor. They only differed in their choice of shape at the extreme of ranges and azimuths and in how they the perceived radius of curvature of the surface (i.e., whether they chose a 'big pole' or a 'small pole' to bound a edge protruding into the active volume).

7 CONCLUSION

We have introduced a new approach to the use of ultrasonic sensors for the classification of surfaces. The light requirements posed to both software and hardware make the approach suitable to be used for local environment mapping in mobile robots. In addition, our initial experimental work with classifier construction, which is based upon classifying the echoes directly rather than using a model-based approach, highlights some of the advantages of symbolic learning methods in ill-posed problems.

References

Atlas, L. *et al.* (1990). Performance comparisons between backpropagation networks and classification trees on three real-world applications. In D. Touretzky (ed.), *Advances in Neural Information Processing Systems 2* (pp. 622–629). San Mateo, CA: Morgan Kaufmann.

Barshan, B. and Kuc, R. (1992). A bat-like sonar system for obstacle localization. *IEEE Transactions on Systems, Man and Cybernetics* **22**: 636–646.

Craven, M. and Shavlik, J. (1993). Learning to represent codons: a challenge problem for constructive induction. In *Proc. Thirteenth International Joint Conference on Artificial Intelligence* (pp. 1319–1324). San Mateo, CA: Morgan Kaufmann.

Dietterich, T., Hild, H., and Bakiri, G. (1990). A comparative study of ID3 and backpropagation for English text-to-speech mapping. In B. Porter and R. Mooney (eds.), *Machine Learning: Proceedings of the Seventh International Conference* (pp. 24–31). San Mateo, CA: Morgan Kaufmann.

Fisher, D. and McKusick, K. (1989). An empirical comparison of ID3 and back-propagation. In *Proc. Eleventh International Joint Conference on Artificial*

Intelligence (pp. 788–793). San Mateo, CA: Morgan Kaufmann.

Fisher, D., McKusick, K., Mooney, R., Shavlik, J., and Towell, G. (1989). Processing issues in comparisons of symbolic and connectionist learning systems. In *Proc. Sixth International Workshop on Machine Learning* (pp. 169–173). San Mateo, CA: Morgan Kaufmann.

Hickling, R. and Marin, S.P. (1986). The use of ultrasonics for gauging and proximity sensing in air. *The Journal of the Acoustical Society of America* **79**: 1151–1160.

Kira, K. and Rendell, L. (1992). A practical approach to feature selection. In D. Sleeman and P. Edwards (eds.), *Machine Learning: Proceedings of the Ninth International Workshop* (pp. 249–256). San Mateo, CA: Morgan Kaufmann.

Kuc, R. and Bozma, Ö. (1991). Building a sonar map in a specular environment using a single mobile sensor. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **13**: 1260–1269.

Mooney, R., Shavlik, J., Towell, G., and Gove, A. (1989). An experimental comparison of symbolic and connectionist learning algorithms. In *Proc. Eleventh International Joint Conference on Artificial Intelligence* (pp. 775–780). San Mateo, CA: Morgan Kaufmann.

Núñez, M. (1988). Economic induction: a case study. In D. Sleeman (ed.), *Proc. Third European Working Session on Learning* (pp. 139–145). London: Pitman.

Pagallo, G. and Haussler, D. (1990). Boolean feature discovery in empirical learning. *Machine Learning* **5**: 71–99.

Quinlan, R. (1993). *C4.5: Programs for Machine Learning*. San Mateo, CA: Morgan Kaufmann.

Shavlik, J., Mooney, R., and Towell, G. (1991). Symbolic and neural learning algorithms: An experimental comparison. *Machine Learning* **6**: 111–143.

Tan, M. and Schlimmer, J. (1990). Two case studies in cost-sensitive concept acquisition. In *Proc. Eighth National Conference on Artificial Intelligence* (pp. 854–860). Cambridge, MA: MIT Press.

Thrun, S. *et al*. (1991). The MONK's problems – a performance comparison of different learning algorithms. Report CMU-CS-91-197. School of Computer Science, Carnegie Mellon University.

Watanabe, S. and Yoneyama, M. (1992). An ultrasonic visual sensor for three-dimensional object recognition using neural networks. *IEEE Transactions on Robotics and Automation* **8**: 240–249.

Weiss, S. and Kapouleas, I. (1989). An empirical comparison of pattern recognition, neural nets, and machine learning classification methods. In *Proc. Eleventh International Joint Conference on Artificial Intelligence* (pp. 781–787). San Mateo, CA: Morgan Kaufmann.