A General Classifier of Whisker Data using Stationary Naive Bayes: Application to BIOTACT Robots

Nathan F. Lepora[1], Charles Fox[1], Mat Evans[1], Ben Mitchinson[1], Asma Motiwala[1], J. Charlie Sullivan[2], Martin J. Pearson[2], Jason Welsby[2], Tony Pipe[2], Kevin Gurney[1], and Tony J. Prescott[1]

[1] Adaptive Behaviour Research Group, Department of Psychology, University of Sheffield, UK.
{n.lepora,c.fox,mat.evans,b.mitchinson,a.motiwala, k.gurney,t.j.prescott}@sheffield.ac.uk
[2] Bristol Robotics Laboratory, Bristol Business Park, Bristol, UK.
{charlie.sullivan,martin.pearson,jason.welsby,tony.pipe}@brl.ac.uk

Abstract. A general problem in robotics is how to best utilize sensors to classify the robot's environment. The BIOTACT project (BIOmimetic Technology for vibrissal Active Touch) is a collaboration between biologists and engineers that has led to many distinctive robots with artificial whisker sensing capabilities. One problem is to construct classifiers that can recognize a wide range of whisker sensations rather than constructing different classifiers for specific features. In this article, we demonstrate that a stationary naive Bayes classifier can perform such a general classification by applying it to various robot experiments. This classifier could be a key component of a robot able to learn autonomously about novel environments, where classifier properties are not known in advance.

Keywords: BIOTACT, Active touch, Whiskers, Bayes' rule, Classifier

1 Introduction

Robotics has much to learn from the ways that animals are constructed, control their bodies and utilize their sensory capabilities [1]. All of the robots described in this article result from a long-term collaboration between biologists and engineers that aims to further understanding of biological vibrissal (whisker) systems and determine the potential applications to engineered systems such as autonomous robots [2]. Research into these systems was undertaken as part of a European Framework 7 project termed BIOTACT (BIOmimetic Technology for vibrissal Active Touch) [3]. The vibrissal sensing technology used in these robots was inspired by many aspects of biological whiskers [4], including their morphology, control and sensory information processing.

This article focuses on how to process the sensory information from artificial vibrissal sensors to categorize and recognize the robot's nearby environment.

This classification problem relies on characterizing stimuli from the environment that are similar every time they are encountered, for example the speed of a contacting object is typified by the peak whisker deflection it produces [5]. Classifiers of sensory data from artificial vibrissae have generally focussed on distinctive features of the whisker signal produced by the stimulus. Examples include peak deflection amplitude and duration for contact speed and radial distance [5], the profile of the whisker deflection power spectrum for texture [6–8], and the change in peak deflection across many whiskers for surface shape [9]. These classifiers perform well in the situations they are designed for, but are not accurate outside their operating range. For example, peak whisker deflection may not accurately characterize texture.

While it is possible to choose *specific* classifiers for robots sensing artificial environments, in more natural environments it would be useful to employ *generic* classifiers that apply across a broad range of stimuli. Traditionally this goal has not been a focus of classifier design for whiskered robots, because the first priority was to find a set of classifiers that allows the robot to interact with test environments. However, now that a fairly comprehensive library of classifiers is becoming available [5–14], such questions about the choice and flexibility of classifiers are becoming relevant.

In recent work, we found that a classifier based on generic probabilistic methods could recognize surface texture accurately by utilizing the probability distribution (histogram) of whisker deflections over a time window of data [12]. In principle, such a method could also characterize other stimuli by utilizing the overall statistical properties of whisker deflections to determine the salient aspects of each stimulus. To explore the hypothesis that a probabilistic classifier could apply across general types of stimuli, we demonstrate here that one method, stationary naive Bayes, applies well across a broad range of stimuli encountered by several robot platforms constructed for the BIOTACT project.

2 Classification with stationary naive Bayes

The probabilistic classifier used here is based on Bayes' rule, but makes two assumptions for simple application to time series of whisker data. First, the sensor measurements are assumed independent, or naive, and second the measurement distributions are assumed stationary in time. Naive Bayes is usually considered across different data dimensions rather than over time series with the stationary assumption, and so we refer to the present method as stationary naive Bayes.

This method relies on calculating the probability distributions of measured time series from the empirical frequencies with which values occur in classes of training data. The probability of a single measurement x being from a class is

$$P(x|\mathrm{C}_l) = \frac{n_x(\mathrm{C}_l)}{\sum_x n_x(\mathrm{C}_l)}, \tag{1}$$

where $n_x(\mathrm{C}_l)$ is the total number of times that the value x occurs in the time series for class C_l. The conditional probability $P(x|\mathrm{C}_l)$ is commonly referred to as a likelihood of the sensor measurement x occurring.

Given a new set of test data, Bayes' rule states that the (posterior) probability $P(\mathrm{C}_l|x_1,\cdots,x_n)$ for a time-series of measurements $x_1,\cdots,x_n$ being drawn from the training data for class C_l is proportional to the likelihood of those measurements $P(x_1,\cdots,x_n|\mathrm{C}_l)$ estimated from that training data

$$P(\mathrm{C}_l|x_1,\cdots,x_n) = \frac{P(x_1,\cdots,x_n|\mathrm{C}_l)P(\mathrm{C}_l)}{P(x_1,\cdots,x_n)}, \tag{2}$$

where $P(\mathrm{C}_l)$ is the (prior) probability of the data being from class C_l and $P(x_1,\cdots,x_n)$ is the (marginal) probability of measuring $x_1,\cdots,x_n$ given no other information. A Bayesian classifier finds which class C has maximum *a posteriori* probability given the measurement time series

$$\mathrm{C} = \arg\max_{\mathrm{C}_l} P(\mathrm{C}_l|x_1,\cdots,x_n) = \arg\max_{\mathrm{C}_l} \frac{P(x_1,\cdots,x_n|\mathrm{C}_l)P(\mathrm{C}_l)}{P(x_1,\cdots,x_n)}, \tag{3}$$

where arg max refers to the argument (the class) that maximizes the posterior.

The following formalism uses that the marginals $P(x)$ are independent of the classes and can be ignored in the arg max operation. It will also be convenient to use the logarithm of the posterior probability to turn products of probabilities into sums. Then the classification in Eq. 3 is equivalent to finding

$$\mathrm{C} = \arg\max_{\mathrm{C}_l} \; \log P(x_1,\cdots,x_n|\mathrm{C}_l) + \log P(\mathrm{C}_l), \tag{4}$$

which is just the maximum over the log-likelihoods plus log priors of a series of measurements $x_1,\cdots,x_n$ being from each class of training data.

An important simplification occurs if the measurements are assumed statistically independent at each time across the data and the probability distributions from which these measurements are drawn are assumed stationary in time. Then the overall conditional probability of a series of measurements factorizes into a product of conditional probabilities for each individual measurement. Consequently, the classification can be rewritten as

$$\mathrm{C} = \arg\max_{\mathrm{C}_l} \; \sum_{i=1}^{n} \log P(x_i|\mathrm{C}_l) + \log P(\mathrm{C}_l), \tag{5}$$

Thus the most probable class is found from the maximum over the summed log-likelihoods for the time window, shifted by its log-prior. This equation defines the stationary naive Bayes classifier used in this paper, with 'naive' denoting the assumption of statistical independence over time and stationary denoting that the same likelihood distribution applies for all times.

The following examples are based around robot experiments in which similar amounts of data are collected for all classes. There is then no biasing towards any class from prior knowledge of its occurrence frequency. Therefore, we assume that the priors are equal and ignore them in the classification (Eq. 5). Although the classification is then equivalent to maximum likelihood, we retain the priors in the formalism to emphasize the Bayesian nature of the method in that in principle it applies also to situations with *a priori* class knowledge.

3 Experiment 1: Surface texture classification

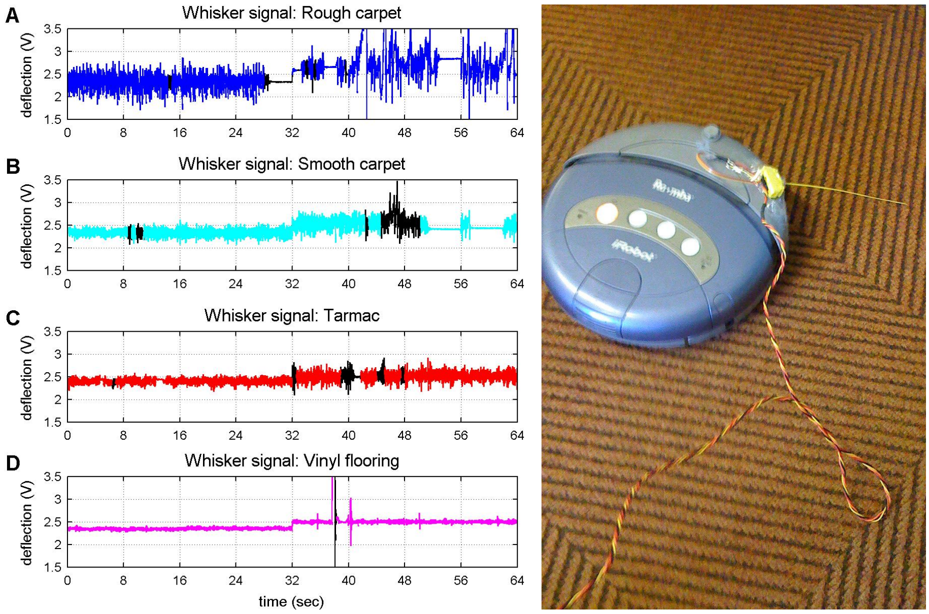

Fig. 1. Surface texture classification with the iRobot Roomba.
The Y-deflection validation data is shown for each of the four textures. On each trace, the first four half-trials of 8 seconds are for anticlockwise motion and the latter four half-trials are for clockwise motion. Hits are shown in color and misses shaded in black.

Methods: Surface texture data was collected from whisker attached to an iRobot Roomba mobile robotic platform [7].

The whisker was mounted on the robot at a 45 degree angle to the direction of forwards motion with a downwards elevation sufficient to contact the floor (Fig. 1). The whisker sensor consisted of a flexible plastic whisker shaft mounted into a short, polyurethane rubber filled tube called a follicle case. Overall dimensions were approximately five-times a rat whisker (200mm long, 2mm diameter), with material properties matched to scaled-up biological whiskers [15].

Four surfaces were chosen for classification: two carpets of different roughnesses, a tarmac surface and a vinyl surface. Deflections of the whisker shaft were measured as 2-dimensional (X,Y) displacements at the base using a Hall effect sensor sampled at 2KHz. In these experiments, the robot moved in a stereotyped rotating manner, spinning either anticlockwise or clockwise (4 trials each of 16 seconds). Here the classification applies over both types of motion.

Results: The data resemble noisy time series interspersed with dead zones, jumps and systematic changes in the mean or variance (Figs 1A-D). The initial 8 seconds of each trial was used for training data, leaving the final 8 seconds for validation testing. The texture likelihoods were then found from the histograms of the empirical frequencies for measurement values of the four textures (Methods, Eq. 1). In previous work, we used likelihood distributions that assumed statistical independence between horizontal (X) and vertical (Y) deflections of the whisker [12]. Here, we build on that study by utilizing a two-dimensional joint-likelihood for classification. This study also aims to present texture classification in the wider context of identifying other aspects of the environment, such as contact speed and distance or shape classification, as discussed later.

For validation, the data was separated into discrete segments of fixed temporal window size 500ms over which the texture is determined. The stationary naive Bayes classifier then considers the log-likelihood values for the four candidate textures from the conditional probability distributions at the same measurement value in the training data. As discussed above, two-dimensional likelihoods over pairs of X and Y deflection values were considered, binning measurements over a grid of resolution 10mV by 10mV. The log-likelihood values over the segments of validation data were then summed for each of the four textures (Methods, Eq. 5), with the maximal value giving the classified texture.

The correct classifications, or hits, for each of the four textures is shown on Figs 1A-D by the colored segments of data and the misses by black segments. The classification for all four textures was highly accurate (mean hit rates 89%, 88%, 91% and 99%). These results are apparently little different from those assuming independence between the X and Y directions (*c.f.* [12]). Thus in this situation, a saving in computational complexity can be made by moving to a classifier that naively treats the X and Y values as independent; however, we expect that in general the dependence between separate streams of information is important, and the determination of when it is remains an open research question.

In general, the stationary naive Bayes classifier was robust to changes in robot motion, since an identical classifier was used for both clockwise and anticlockwise rotations. This is perhaps surprising considering that such changes cause large systematic effects in the whisker deflection data (Fig. 1 at 32 seconds). It is also curious that the misses occurred mainly near jumps in the whisker deflection, corresponding to the whisker spuriously catching the floor surface, rather than directly confusing textures on normal data. Thus the accuracy on data without such artifacts would improve upon the 92% mean hit rate found here.

4 Experiment 2: Speed and radial distance classification

Methods: An XY positioning robot moved a vertical bar onto a fixed, horizontal whisker to probe the whisker deflection dependency on contact speed and radial distance [5]. The data set consisted of four repeats of 2626 distinct contact speeds and distances along the whisker shaft.

The whisker sensor was similar to that in the preceding texture study, with slightly different dimensions (whisker length 185mm, tapering from 3mm to 0.5mm) and material properties again scaled up from biological rat whiskers [15]. Objects were moved onto the whisker and when contact was detected the robot

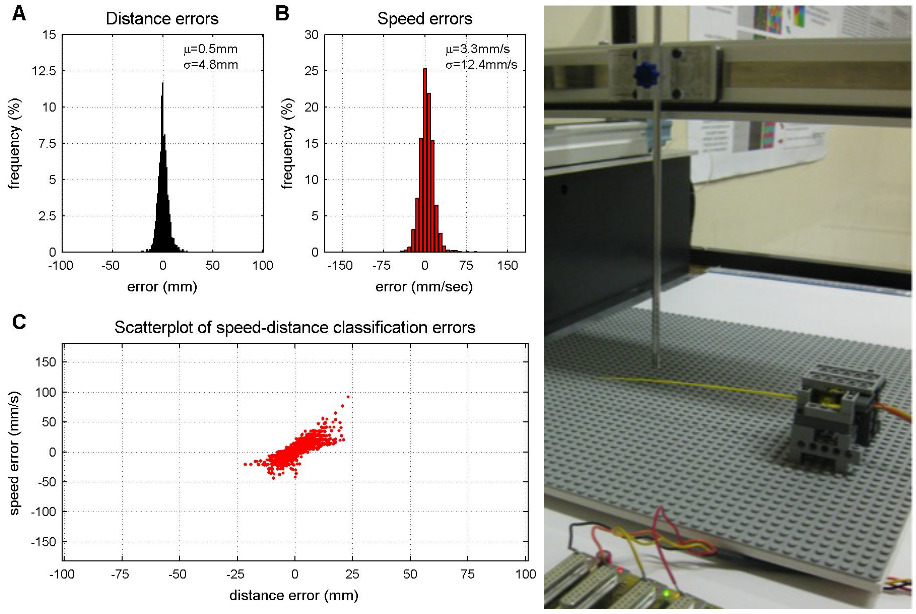

Fig. 2. Speed and radial distance classification with the XY table.
Panels A and B show the contact speed errors and radial distance errors between the target value for the validation trial and the classified value from the stationary naive Bayes classifier. Panel C shows a scatter plot of these errors.

retracted the bar at its maximum speed after a short 300ms delay. Thus the contacts were actively feedback-controlled, as are whisker contacts observed in animal behaviour studies. Radial contact distance along the whisker shaft away from the base was sampled in 1mm steps from 80mm to 180mm and contact speed in 7.2mm/sec steps from 36mm/sec to 216mm/sec, giving $101 \times 26 = 2626$ distinct contacts. For more details we refer to [5].

Results: Whisker signals had a characteristic profile in which their X-deflection increased smoothly from rest until reaching their peak after a few hundred milliseconds, and then quickly returned towards baseline followed by damped oscillatory ringing of period $\sim$50ms [5, Fig. 3]. As observed in the original analysis, deflections of a similar peak amplitude can be produced for many contact speed and distance combinations (by compensating the increase in speed with a decrease in radial distance).

Traces from 100ms before the initial deflection to 650ms after were used for analysis. Empirically, it was found that whisker velocity traces gave a better localization of contact speed and distance than whisker deflection traces. Thus, the positional deflections were converted to velocities for the analysis (by taking their numerical derivative and Gaussian smoothing to reduce noise). For each contact combination, three trials were used for training and the remaining trial saved for validation. These three trials were concatenated and the likelihoods

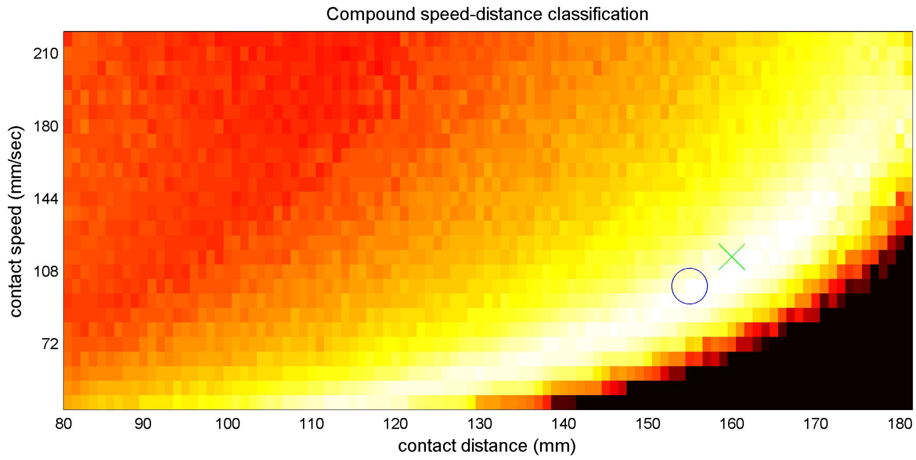

Fig. 3. Speed and radial distance classification with the XY table.
The brightness of each pixel represents the magnitude of the log-posterior for each combination of speed and distance marked on the axes. The green cross represents the target speed and distance of the test data and the blue circle the classified values according to the maximal log-posterior (at the brightest pixel).

found from the histograms of the empirical frequencies for the time series values binned into 10mV/sec intervals (Methods, Eq. 1). This procedure resulted in 2626 distinct likelihood distributions, one for each speed-distance combination.

For validation, each likelihood distribution was used to construct a log-posterior from the test trial data (Methods, Eq. 5). Generally, the distribution of log-posteriors for each trial had an extended curving segment of high values close to the target speed and distance of the test trial, surrounded by a drop-off of values further away from these targets (example shown in Fig. 3). The maximum log-posterior gave the classified speed and radial distance for that test trial.

Histograms of the errors between the classified and target speeds and distances were approximately Gaussian with little systematic bias of their centers from zero error (Figs 2A,B). A two-dimensional scatter-plot of these errors was consistent with these observations, and also revealed correlations between the speed and distance errors (Fig. 2C) consistent with the prevailing direction of the high value region in the log-posterior distribution (*c.f.* Fig. 3, bright region). This direction in contact speed and radial distance coincided with the region of constant peak amplitude examined in the original study of feature-based classifiers [5]. Hence, the stationary naive Bayes classifier appears highly dependent on peak deflection amplitude, but must also use other aspects of the whisker signal to localize the contact speed and distance close to the target values. Overall, the spread of distance errors was around 5mm compared with an overall range of 100mm, and the spread of speed errors was around 12mm/sec compared with an overall range of 180mm/sec. Thus the overall accuracy was around 93-95%.

5 Experiment 3: Shape and position classification

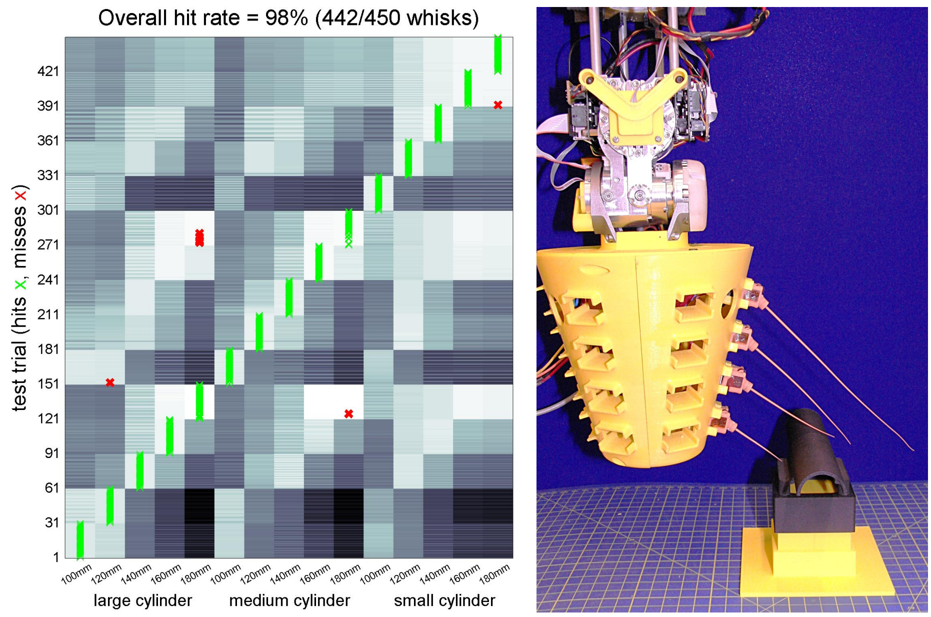

Fig. 4. Shape and position classification with the BIOTACT sensor.
The x-axis represents the cylinder configuration that is being tested for (5 positions and 3 curvatures). The y-axis denotes the test whisk, which were taken in 30 whisks trials in order of the configurations shown on the x-axis. Green crosses represent test trials that were correctly classified and red crosses represent incorrect classifications. The shade of pixels represents the value of the log-posterior according to the test cylinder configurations, such that brighter pixels denote larger log-posteriors (with the classified configuration being the brightest for each test trial).

Methods: The BIOTACT sensor (Fig. 4) was used to sense cylinders of various diameters placed horizontally at different positions along a platform and approached from a vertical direction.

The G1 (Generation 1) BIOTACT sensor consists of a truncated conical head that holds up to 24 whisker modules. The whisker modules are arranged in 6 radially symmetric rows of 4, oriented normally to the cone surface [16]. The head is mounted as the end-effector of a 7-degree-of-freedom robot arm (Fig. 4). For the present experiments, the head was fitted with a total of 4 whiskers in one row appropriate for sensing axially symmetric shapes such as cylinders aligned perpendicular to the whisker row. These whiskers were moved back-and-forth to repeatedly contact surfaces of interest akin to animal whisking behaviour [4], with individual whisker modules feedback modulated to make light, minimal impingement contacts on the surface [16]. Whiskers towards the front of the head were shorter than at the back (lengths 50mm, 80mm, 115mm and 160mm), and were designed with a taper from a 1.5mm base to a 0.25mm tip.

Three rigid plastic hemi-cylinders with radii of curvature 35mm, 25mm and 15mm were used as test objects. They were mounted with their curved surfaces lying upwards on bases of height 15mm, 25mm and 35mm to ensure their vertical dimensions were equal. Five different positions for each cylinder were used, with central axes offset 100mm, 120mm, 140mm, 160mm and 180mm from the central axis of the head cone of the BIOTACT sensor. The orientation of the BIOTACT head was such that the four whiskers contacted the cylinders along the horizontal and perpendicular to the cylinder axis (*i.e.* across the cylinder). When contacting a horizontal surface, the whisker tips are approximately 30mm apart, spanning a distance from 100mm to 190mm perpendicular to the head axis. The depth of the contacts was arranged to be equal for all trials and only contacting the curved sections of the test objects. Each of the 15 distinct trials was repeated twice with 30 whisks at 1Hz to give training and validation data sets.

Results: The four whisker contacts across the cylinder surfaces had a distinctive pattern depending upon the curvature and position of the cylinder. For the largest cylinder in a central position, four whiskers could contact simultaneously, while for the smallest cylinder only one or two whiskers could make contact.

Single whisk traces from all 4 whiskers were used for analysis, with each trace starting and finishing when the whisker was maximally retracted from the object. The 30 whisks from each training trial were concatenated and the likelihoods found from the histograms of the empirical frequencies for the time series values binned into 10mV/sec intervals (Methods, Eq. 1). This procedure resulted in 15 likelihood distributions, one for each of the 3 cylinder curvatures and 5 positions.

For validation, the classification accuracy over single test whisks were considered. Over the 15 test trials, this gave a total of 450 distinct test whisks. The stationary naive Bayes classifier considered the log-posterior values for the 15 candidate configurations of 3 curvatures and 5 positions, calculated by summing the associated log-likelihood values for the whisker deflections over each single test whisk (Methods, Eq. 5). Correct classifications, or hits, correspond to when the maximum log-posterior coincides with the test cylinder configuration.

Overall, the hit rate for this experiment was above 98% using single test whisks contacting a surface (Fig. 4; hits shown in green and misses in red). The most common mistake was confusing the largest and medium cylinders at their most distant position from the robot head, where only one or two whiskers could contact the object; even then, though, there were just 6 mistakes over 60 test whisks. Therefore, using four whiskers dabbing onto a surface, the BIOTACT sensor could reliably determine both the curvature and horizontal positioning of test cylinders. This was achievable with single whisks lasting less than a second. If instead the full trial of 30 whisks were considered, then all test configurations could be determined with a perfect classification rate of 100%.

6 Discussion

The main point of this article is that a classifier based on probabilistic methods can be a general classifier for artificial whisker sensors because it determines the salient aspects of the data on which to base the discrimination. Such a classifier contrasts with the more traditionally considered specific classifiers that are based on pre-chosen features of the signals, such as peak deflection or profile of the power spectrum. We see these two types of classifiers as complementary in their applications and use. Specific classifiers are appropriate for controlled environments in which many details are known in advance, whereas general classifiers should just ‘work out of the box’ and apply across a range of situations.

For this initial study, we used the simplest Bayesian classifier available for time series analysis, by assuming stationarity and naive independence over sensor measurements. Further work could investigate how relaxing these assumptions affects classification accuracy. That being said, all applications of stationary naive Bayes considered in this article gave highly accurate hit rates of 90% or greater over diverse stimuli and robots. Furthermore, the efficiency savings from computationally simple classifiers, compared with more sophisticated methods for Bayesian inference, are important for sensing in autonomous robots. Hence we expect that the classifier used here will be appropriate for many situations of practical interest for whiskered robots. That being said, future study should clarify when the classification is limited by the simplifying assumptions, for example when correlations in the time series degrade the performance.

One problem requiring classifiers with the general performance found here is to autonomously explore and learn about novel environments. Because novel stimuli are not known in advance, only general classifiers would be sufficiently versatile for each new situation that arises. Fortunately, the probabilistic classifier in this article, stationary naive Bayes, can also detect novelty [17]. In another study with a BIOTACT robot, SCRATCHbot, novelty detection was examined by training the classifier to recognize the whisker signals from a blank floor and then examining the classifier output when it passed a novel texture. The classifier output, corresponding to the log-probability of the data being from the floor, displayed an anomalous change that can be used to identify novel events. One future project would be to combine the novelty detecting properties of the stationary naive Bayes classifier with its classification abilities to investigate the autonomous learning of stimuli in unfamiliar environments.

A related question for future study is how to extend the standard training/testing protocol of presenting many stereotyped examples of a contact onto a stimulus and then testing with the same examples. In natural environments, the robot might be expected to be more varied in its movements while contacting an object. Or perhaps the motion of the robot would need to become stereotyped to achieve the most reliable classification? These questions relate to a central theme of BIOTACT on active sensing, in which the appropriate sensation for the task is actively selected by adapting the sensor in response to information received about the environment, and will be a focus of future work.

Acknowledgments. The authors thank members of ATLAS (Active Touch Laboratory at Sheffield), the Bristol Robotics Laboratory, and the BIOTACT (BIOmimetic Technology for vibrissal ACtive Touch) consortium. This work was supported by EU Framework projects BIOTACT (ICT-215910), ICEA (IST-027819) and CSN (ICT-248986).

References

1. Y. Bar-Cohen. Biomimetics - using nature to inspire human innovation. *Bioinspiration & Biomimetics*, 1:P1, 2006.
2. T.J. Prescott, M.J. Pearson, B. Mitchinson, J.C.W. Sullivan, and A.G. Pipe. Whisking with robots from rat vibrissae to biomimetic technology for active touch. *IEEE Robotics and Automation Magazine*, 16(3):42–50, 2009.
3. BIOTACT consortium. `http://www.biotact.org`.
4. T.J. Prescott. Vibrissal behavior and function. *Scholarpedia* (In press).
5. M. Evans, C.W. Fox, and T.J. Prescott. Tactile discrimination using template classifiers: Towards a model of feature extraction in mammalian vibrissal systems. *From Animals to Animats SAB 2010.*
6. C.W. Fox, B. Mitchinson, M.J. Pearson, A.G. Pipe, and T.J. Prescott. Contact type dependency of texture classification in a whiskered mobile robot. *Autonomous Robots*, 26(4):223–239, 2009.
7. M. Evans, C.W. Fox, M.J. Pearson, and T.J. Prescott. Spectral Template Based Classification of Robotic Whisker Sensor Signals in a Floor Texture Discrimination Task. *Proceedings TAROS 2009*, pages 19–24.
8. J. Hipp, E. Arabzadeh, E. Zorzin, J. Conradt, C. Kayser, M.E. Diamond, and P. Konig. Texture signals in whisker vibrations. *J. neurophys.*, 95(3):1792, 2006.
9. D.E. Kim and R. Moller. Biomimetic whiskers for shape recognition. *Robotics and Autonomous Systems*, 55(3):229–243, 2007.
10. M. Fend, S. Bovet, H. Yokoi, and R. Pfeifer. An active artificial whisker array for texture discrimination. In *Proc. IEEE/RSJ Int. Conf. Intel. Robots and Systems IROS 2003*, volume 2.
11. M. Fend. Whisker-based texture discrimination on a mobile robot. *Advances in Artificial Life*, pages 302–311, 2005.
12. N.F. Lepora, M. Evans, C.W. Fox, M.E. Diamond, K. Gurney, and T.J. Prescott. Naive Bayes texture classification applied to whisker data from a moving robot. *Proc. IEEE World Congress on Comp. Int. WCCI 2010.*
13. J.H. Solomon and M.J. Hartmann. Biomechanics: Robotic whiskers used to sense features. *Nature*, 443(7111):525, 2006.
14. J.H. Solomon and M.J.Z. Hartmann. Artificial whiskers suitable for array implementation: Accounting for lateral slip and surface friction. *IEEE Transactions on Robotics*, 24(5):1157–1167, 2008.
15. M.J. Pearson, I. Gilhespy, C. Melhuish, B. Mitchinson, M. Nibouche, A.G. Pipe, and T.J. Prescott. A biomimetic haptic sensor. *International Journal of Advanced Robotic Systems*, 2(4):335–343, 2005.
16. J.C. Sullivan, B. Mitchinson, M.J. Pearson, M. Evans, N.F. Lepora, C.W. Fox, C. Melhuish, and T.J. Prescott. Tactile Discrimination using Active Whisker Sensors. *IEEE Sensors*, 99:1, 2011.
17. N.F. Lepora, M.J. Pearson, B. Mitchinson, M. Evans, C. Fox, A. Pipe, K. Gurney, and T.J. Prescott. Naive Bayes novelty detection for a moving robot with whiskers.