

# Feature Selection for Gas Identification with a Mobile Robot

Marco Trincavelli and Amy Loutfi

**Abstract**—In this paper we analyze the problem of discrimination of gases with mobile robots. Previously, it has been shown that the conditions in which data is collected heavily influence the characteristics of the signal to be identified. As a result, the already difficult task of selecting features which characterize a gas is made more challenging by the absence of a steady state response. This is often due to the movement of the robot, and/or the physical properties of the environment, e.g., turbulent airflow creating patches and eddies in the plume. In this work we compare two approaches for feature selection which are able to consider explicitly the information on the experimental setup and optimize the subset of features used in the recognition process. The approaches are tested on a large data set collected with a mobile robot moving in different environments (outdoors and indoors). The results show that the classification performance is improved resulting in a higher average accuracy and lower variance in the accuracy across the different experimental setups.

## I. INTRODUCTION

The ability to discriminate and identify gases is important for mobile robot applications rely on gas detection and include search and rescue and environmental robotics. This is especially the case when sensor technologies such as tin dioxide semiconductors are used as they are intrinsically partially selective and display cross sensitivities to other substances within the same gas family [1]. With these types of sensor technologies are employed, discrimination becomes particularly relevant for adjacent mobile olfactory tasks, such as source declaration [2], odour based navigation [3], [4] and gas distribution mapping [5] particularly when deployment in real world settings must contend with the presence of multiple gaseous agents.

Discrimination of gases using an array of partially selective semiconductor gas sensors presents a number of challenges that specifically relate to mobile robotics. One challenge deals with the extraction of features which correspond to the dynamic properties of the signal response, as steady responses are seldom reached due to the movement of the robot as well as the physical properties of the environment (e.g. patches and eddies due to turbulent airflow) [6]. This particular fact makes the classification of gases with a mobile robot, in which the array of sensors is directly exposed to the environment, a different problem than classification of odours when the array is in a chamber with controlled humidity, temperature and sampling procedure. While the related works concerning the classification of

gases with a static electronic nose is dense [7] only few works have addressed the classification of odours based uniquely on the transient [6], [8], [9]. Previous work has shown that it is possible to use a set of standard feature extraction tools to obtain information on the dynamic part of the signal [6]. When the sensors are used on a mobile platform, however, it has also been shown that not only is the information about the gases present in the feature set, but also present is the intrinsic information about how the robot is interacting with the environment [10]. Indeed it is visible in Figures 2 and 3 how a change in the orientation in which the robot sweeps the environment impacts the shape of the collected signal significantly. Ultimately, for discrimination on a mobile robot to be generic, it is necessary to select a set of features which are independent of the experimental setup. In this way, we can enable the possibility to train a mobile robot on a specific gas in one environment and deploy the robot in another environment having different properties and/or having different interaction with the plume. Indeed this ability to select features which are independent of the experimental setup is not only important for mobile olfactory discriminating robot but also for other fields within mobile robotics characterized by multivariate sensor data, and differing environmental conditions between training and deployment.

The general approach we consider in this paper for dealing with the above problem is to extract features from the sensor signals and select features that show regularity across the experimental setups while providing enough discrimination between different analytes. This approach differs from other dimensionality reduction methods that are based on projecting the original feature space to a lower dimensional one in that in the proposed approach we do not only rely on the structure of the data (PCA, KPCA) or the label information (LDA, KDA), rather we inject into the system some auxiliary information that comes from the knowledge on how the data set has been constructed. We apply two different approaches for feature selection. One approach is a *filter approach* in which we rank all the features according to a score and we select the first features in the ranking. The other approach is a *wrapper approach*, in which the different features subsets are evaluated using the same classifier that then is used to perform the final identification. We validate our approach on a dataset composed of experiments carried out in five different setups. A five fold cross validation is performed where in every fold data from one experimental setup is left out for validation and only data from the four other setups are used for selecting features and training the classifier. In this way we evaluate the ability of our system to perform

This work was supported by the Swedish Research Council - Vetenskapsrådet.

The authors are with the AASS Research Center, School of Science and Technology, Örebro University, SE-701 82 Örebro, Sweden  
name.surname@aass.oru.se

well in a different configuration than the one on which it has been trained.

The remainder of the paper is organized as follows: Section II presents briefly the pattern recognition algorithm, Section III gives a description of the general feature selection problem as well as explains the proposed algorithms and the main contribution of this work, Section IV describes the different experimental setups, Section V presents the results and finally Section VI summarizes the paper with a discussion and future works.

## II. THE PATTERN RECOGNITION ALGORITHM

The pattern recognition algorithm consists of baseline subtraction, signal segmentation, feature extraction, feature selection and classification. The baseline is the value that a gas sensor gives as output when it is exposed to clean air. This value depends on temperature, humidity and long/short term drift [1]. The baseline is subtracted from the output value of the sensors using a differential baseline subtraction. The baseline value is measured for 60 s at the beginning of every experiment, when the gas source is closed and the robot is still. After performing baseline subtraction the signal is smoothed using an average filter of dimension 5 (samples, corresponding to 4 s at a sampling frequency of 1.25 Hz) in order to suppress the noise due to sampling and quantization. The smoothed signal is then segmented into three different phases, namely baseline, rise and decay according to the value of the first derivative. Note that the steady state response is never reached in our experimental sessions. The segmentation procedure can be easily explained using a finite state machine as shown in Figure 1.

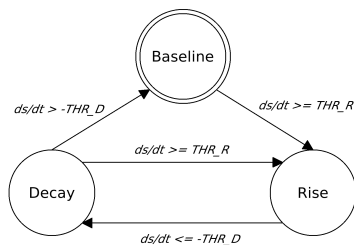


Fig. 1. Finite State Machine that illustrates the segmentation algorithm.

In this figure the first derivative is denoted as  $ds/dt$  and the threshold for the rise and decay are  $THR_R$  and  $THR_D$  respectively. Two different thresholds are needed since the rise and decay phase are described using a first-order model and the time constant for the rising phase is smaller [11]. A complete response to a patch is considered to be the ensemble of a consecutive rise and decay phase. This isolated response is passed to the feature extraction module that fits a second degree polynomial to the points in the response. The choice of the polynomial degree is due to the similarity of the shape of an isolated sensor response with the one of a parabola (second order polynomial). The feature vector is built by concatenating the 3 coefficients obtained by fitting each of the five sensors, obtaining a 15 dimensions vector. The dimensionality of the feature vector is then reduced

selecting one subset of the feature using one of the two proposed approaches described in Section III. Finally, the reduced feature vector is classified using a multi-class (3 classes were considered, one for each gas) Support Vector Machine (SVM) with Gaussian kernel, constructed using the *one-vs-one* approach [12].

## III. THE FEATURE SELECTION PROBLEM

In general with data from gas sensors arrays, the sensor responses are highly correlated and therefore the features contain redundant information, which needs to be removed with further processing. If redundant or noisy information is not removed before trying to learn a model, the problem of the *Curse of Dimensionality* [13] may arise. This refers to the fact that for high-dimensional spaces it is difficult to collect enough samples to attain a high enough density in order to obtain a valid estimate for a function or a discriminant. The most common way of dealing with this problem is to reduce the dimensionality of the feature space by either projecting the original  $N$  dimensional space into a  $M$  dimensional one where  $M < N$  (feature extraction), or selecting  $M$  out of the  $N$  original features (feature selection). In this work we consider a feature selection approach able to select the features that provide discriminative power while independent on the experimental session. Feature selection methods proposed in literature fall into two main categories, the filter approaches and the wrapper approaches [14]. The filter based methods produce a ranking of the features based on an optimality criterion and then select the first  $M$  features in the ranking, where  $M$  can be arbitrarily chosen. Various optimality criteria for filter methods have been proposed in literature where the most common are linear correlation criterion and information theoretic ranking criterion [14]. The latter rely on an estimation of the mutual information between each feature and the labels vector. The mutual information,  $I(X, Y)$ , in between two random variables  $X$  and  $Y$  can be calculated according to the following formula:

$$I(X, Y) = \int_Y \int_X P(X, Y) \log \frac{P(X, Y)}{P(X)P(Y)} dX dY \quad (1)$$

The mutual information is a quantity that measures the mutual dependence of the two variables. It is lower bounded by the value zero, obtained in case the two variables are independent and upper bounded by the entropy of one of the two random variables in case they are coincident.

Wrapper methods consist in using the prediction performance of a given classifier to assess the relative usefulness of subsets of variables. Since the number of possible feature subsets of  $N$  features is  $2^N$ , an exhaustive search is unfeasible even for small  $N$ . Therefore wrapper algorithms use a search heuristic to perform a partial exploration of the feature subsets space. The two simplest search strategies are *Forward Selection* and *Backward Elimination*. In *Forward Selection* the algorithm starts from the empty set of features and at every iteration adds the feature that gives the greatest improvement in classification performance. In *Backward*

*Elimination*, the search starts from the full feature set and at every iteration the feature that causes the smaller degradation in performance is removed. Notice that, especially in the first iterations of the backward elimination, when the number of features included in the subset is large, removing a feature is likely to either increase (or not degrade) the performance. The stopping criteria for both algorithms is either when the desired number of features is achieved or when a drop in the classification performance is obtained.

In this work we propose two different feature selection methods, one that belongs to the filter family and one that belongs to the wrapper family. Both approaches are aimed at selecting features that are at the same time discriminative and not heavily dependent on the experimental setup. For the filter approach, we calculate a score for each feature  $f$  according to the following formula:

$$\gamma(f) = \frac{I(f, S)^\alpha}{I(f, C)} \quad (2)$$

where  $S$  is the experimental setups vector,  $C$  is the analyte labels vector,  $I$  is the mutual information between two random variables and  $\alpha$  is a parameter that modulates the relative importance of the two factors. The best features in the set have the smallest values for  $\gamma$ . It is important to notice that for  $\alpha = 0$ , the expression degenerates to  $\gamma(f) = I(f, C)^{-1}$  and therefore we would select the features with the highest mutual information with respect to the class vector, that is equivalent to the traditional information theoretic ranking criterion. For increasing values of  $\alpha$  we tend to prefer features that do not carry any information about the experimental setup and therefore are more robust to changes in the environment or in the moving strategy of the robot. The joint and marginal distributions ( $P(f, C), P(f, S), P(f), P(C)$  and  $P(S)$ ) used in the calculation of the mutual information are estimated using histogram techniques.

In the wrapper approach we propose a modification to the *Backward Elimination* algorithm. Indeed one of the weak points of the *Backward Elimination* algorithm is that many features would be good candidates for elimination since the performance of the subsets of the remaining features does not drastically change. Rather than perform an uninformed choice on which feature to eliminate (since they are equivalent with respect to our criterion), we isolate the features which obtain a comparable classification accuracy. These features are then ranked according to the mutual information with respect to the experimental setup and the highest ranked feature is permanently eliminated.

The algorithm can be described by the following four steps:

- 1) Perform an 8-fold cross validation using the whole feature set in order to estimate the hyperparameters  $C$  and  $\sigma$  for the SVM classifier with Gaussian kernel (the classifier used in the pattern recognition algorithm).
- 2) Given that the current feature subset contains  $N$  features (in the beginning  $N$  is the total number of features), consider all the possible subsets containing

$N - 1$  features and calculate the classification accuracy for each of them.

- 3) Select all the features that obtained the best performance with a certain tolerance  $\epsilon$  and order them according to the mutual information with respect to the experimental setup.
- 4) Remove permanently the feature with the highest mutual information with respect to the experimental setup. If the desired number of features has been reached then stop, otherwise go back to point 2<sup>1</sup>.

It should be noted that the classifier chosen to select features with a wrapper approach would bias the choice of the features. Therefore it is preferable to use the same classifier that will be used for the final identification. Since the feature selection procedures requires  $O(N^2)$  training sessions (where  $N$  is the number of features), the classifier should be chosen carefully as to not incur a computationally intractable procedure. The SVM in this case has been chosen since its training can be expressed as a convex optimization problem and therefore solved efficiently.

#### IV. EXPERIMENTAL SETUPS

The robot used in the experiments is an ATRV-JR all terrain robot equipped with the Player Robot Device Interface [15]. Player provides both the interface to the sensors and the actuators, and high level algorithms to address robotic tasks such as localization (*amcl* driver) and navigation (*vfh* and *wavefront* drivers). The robot is equipped with an electronic nose, an actively ventilated aluminum tube containing an array of five metal oxide gas sensors, mounted in front of the robot at a height of 0.1 m on the ground. The sensors present in the array are listed in Table I together with their target gases.

Model	Gases Detected	Quantity
Figaro TGS 2600	Hydrogen, Carbon Monoxide	2
Figaro TGS 2602	Ammonia, Hydrogen Sulfide, VOC (volatile organic compound)	1
Figaro TGS 2611	Methane	1
Figaro TGS 2620	Organic Solvents	1

TABLE I  
GAS SENSORS USED IN THE ELECTRONIC NOSE.

The experiments have been performed in three different locations using four moving strategies which attempt to vary the interaction of the robot with a possible plume. In all experiments the robot was moving with a speed of 0.05 m/s. The gas source was a cup full of the analyte placed on the ground. The first location that has been considered is a large closed room in which the robot followed a sweeping trajectory with two orthogonal orientations that we name N-S and E-W. Figures 2 and 3 provide a graphical representation of the two paths followed by the robot together with the signal collected during two experimental runs. The second

<sup>1</sup>The choice of performing the hyperparameters selection only with the full feature set is an heuristic to reduce the computational burden.

Experimental Setup	Location	Moving Strategy	Number of Runs
1	Large Room	Sweep N-S	15
2	Large Room	Sweep E-W	15
3	Classroom	Spiral	18
4	Classroom	Spiral with Stops	72
5	Courtyard	Spiral with Stops	16

TABLE II

SUMMARY OF THE EXPERIMENTAL CONDITIONS IN WHICH THE DATA HAVE BEEN COLLECTED.

set of experiments has been carried out in a small classroom whose door has been left open. In this environment the robot performed two different types of spiral path: a spiral without any stops from the beginning to the end of the experiment and a spiral with stops when an odour is detected, at which point the robot stands static until enough information is obtained to perform a classification. The rooms were ventilated after each experimental run. The last experimental location was a courtyard with an uneven grass surface. In this case the robot performed a spiral movement stopping when a gas is detected similar to the one performed in the classroom. Figure 4(c) is a snapshot taken during an experiment in the courtyard. Table II summarizes the five different experimental configurations. The experiments have been repeated multiple times (more than 100) with three different substances, ethanol, acetone and isopropyl, that are the target substances for our classification problem. During one experimental run multiple responses were collected, for a total of 592 responses evenly distributed among the three analytes.

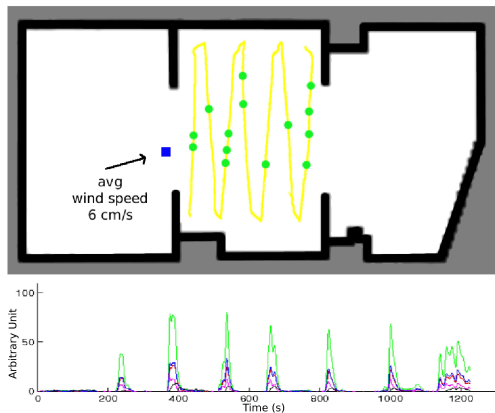


Fig. 2. Upper: Example of experiment in which the analyte used was isopropyl. The robot follows a N-S sweeping trajectory frequently entering and exiting the plume. The arrow shows the average direction and magnitude of the wind flow. The square indicates the position of the source. The solid line is the trajectory of the robot. The circles are locations in which we obtained a sensor response. Lower: The actual sensor readings collected during the run.

## V. RESULTS

The algorithms proposed have been evaluated performing a 5-fold cross validation where every fold has been formed by taking all the samples collected in a specific experimental setup. This evaluation scheme has been chosen in order to

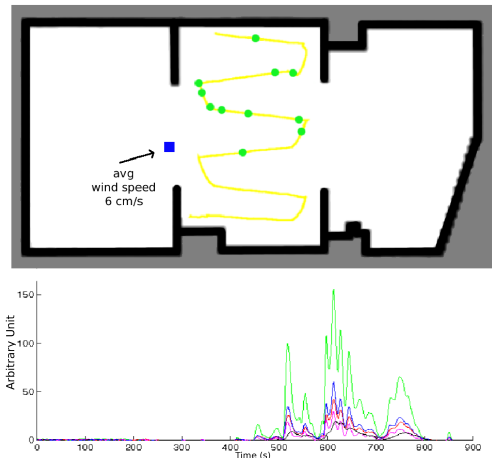


Fig. 3. Upper: Example of experiment in which the analyte used was ethanol. The robot follows a E-W sweeping trajectory and remains continuously in the plume. The arrow shows the average direction and magnitude of the wind flow. The square indicates the position of the source. The solid line is the trajectory of the robot. The circles are locations in which we obtained a sensor response. Lower: The actual sensor readings collected during the run.

analyze the ability of the system to generalize and perform well in an unknown experimental setup. Figure 5 gives the classification performance obtained selecting features with the proposed filter approach with  $\alpha = 10$  and with  $\alpha = 0$  (based only on the mutual information between the features and the labels). The optimal value of  $\alpha$  has been iteratively evaluated.

The error bars display the average performance across the 5 folds together with the standard deviation. We can notice that the proposed filter clearly outperforms the filter based solely on the mutual information with the classes. Moreover the proposed approach obtains in average a smaller standard deviation for the performance across the fold. This is important because it shows how the feature subset obtained with the proposed filter are more robust with respect to variations in the data.

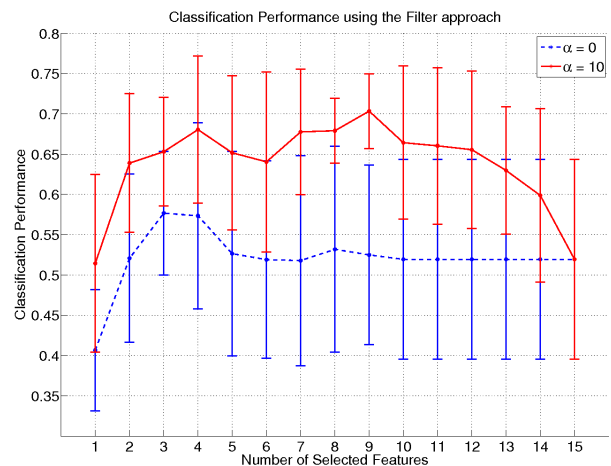
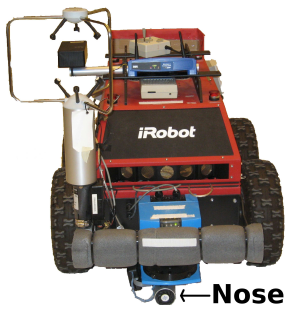


Fig. 5. Error bars displaying average and standard deviation of the performance of the classifier obtained selecting the features using the filter approach. The lines represent the performance of the proposed approach with  $\alpha = 10$  and  $\alpha = 0$ . Notice that  $\alpha = 0$  is a filter based only on the mutual information between each feature and the labels vector.



(a) The robot with the electronic nose and the anemometer



(b) The robot in the large room



(c) The robot in the courtyard

Fig. 4. The robot and snapshots from two experimental runs in different locations.

Figure 6 shows the classification performance obtained by the proposed wrapper approach compared with a wrapper that in case of ties selects to eliminate the first feature in the list. In the proposed approach features are considered equally ranked if the classification performance differs by less than  $\epsilon = 0.2\%$  (given that we have 592 samples, each sample contributes for 0.16%). Also in this case we can see how the proposed approach outperforms the traditional one both with a higher average performance and with a lower standard deviation for the performance across the folds.

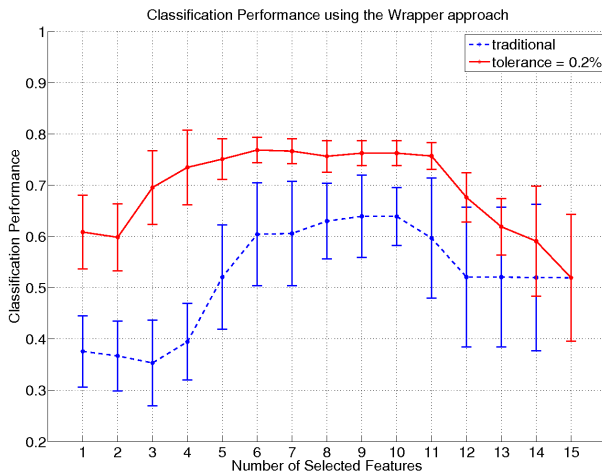


Fig. 6. Error bars displaying average and standard deviation of the performance of the classifier obtained selecting the features using the wrapper approach. The solid line represents the performance of the proposed approach while the dashed line represents the performance of a wrapper that eliminates the feature that when removed obtains the highest classification performance. In case of a tie the wrapper eliminates the first feature of the list.

Comparing the performances of the two proposed approaches we can notice that the number of features yielding the best classification performance is 9-10. Also, the wrapper outperforms the filter both for average classification performance and small variance in between the folds. This has to be expected since the wrapper approach scores the features according to the performance of the target classifier (in contrast with the filter approach that uses a score that is independent from the classifier). The main drawback of the wrapper approach is that it is computationally more expensive since it requires a training of the classifier for

Sensor	Filter	Wrapper
TGS 2600 (1)	7.65 - 9.81	6.52 - 8.55
TGS 2620	7.92 - 10.07	10.38 - 12.42
TGS 2611	5.59 - 7.74	6.25 - 8.28
TGS 2602	4.65 - 6.81	4.12 - 6.15
TGS 2600 (2)	8.79 - 10.94	7.65 - 9.68

TABLE III

95% CONFIDENCE INTERVALS FOR THE AVERAGE RANKING OBTAINED BY THE FEATURES GROUPED BY SENSOR.

Coefficient	Filter	Wrapper
a	2.66 - 3.57	6.46 - 7.62
b	8.34 - 9.25	4.02 - 5.18
c	11.62 - 12.53	11.78 - 12.94

TABLE IV

95% CONFIDENCE INTERVALS FOR THE AVERAGE RANKING OBTAINED BY THE FEATURES GROUPED BY PARABOLA COEFFICIENT.

every feature subset to be evaluated.

To further analyze the results it is also possible to examine the regularity between the ranking of the features across the different sensors and the coefficients of the parabola that we fit to the sensor response. In order to do this we group the features with respect to either the sensor (5 groups) or the coefficient to which they belong (3 groups). By analyzing the rankings obtained by the groups as random variables it is possible to get some insight on how a specific sensor/coefficient contributes to the classification task.

In Table III the average rankings for the features grouped by sensor with a 95% confidence interval are reported. We can see that the wrapper approach ranks sensor TGS2602 as best and TGS2620 as worse, while the filter approach ranks sensor TGS2602 significantly better than all the other sensors except the TGS2611. Instead in Table IV the average rankings for the features grouped by coefficients show that the constant coefficient  $c$  in the parabola equation is consistently ranked worse both by the filter and the wrapper approaches. A geometrical interpretation of this result can be seen in Figure 7 and suggests that the characteristics of the sensor response which are of interest are independent of the vertical position of the parabola. Noting that the characteristics of the sensor response which are of interest are those which contain discriminative power and at the same time are not dependent

on the experimental setup. The implication of this is that instead of estimating the 3 coefficients for the parabola we can translate the parabola so that it passes through the origin (parameter  $c = 0$ , see Figure 7) and then estimate only the two remaining parameters. The result of which makes the fitting procedure more robust since it requires the estimation of fewer parameters. Said differently, the parameter  $c$  can be interpreted as the state of the sensor when a gas patch hits the array and triggers a response. As the previous state information is not necessarily relevant for classification of the current signal and the parameter  $c$  can be discarded. In fact this can be seen by comparing Figure 2 and Figure 3, where in the former the value of  $c$  is low and in the latter high. Since the experimental conditions are similar this is most likely due to the different sweeping patterns of the robot with respect to the odour plume. Future work will address the possibility to use a parabola passing through the origin (2 parameters only) as feature extraction method.

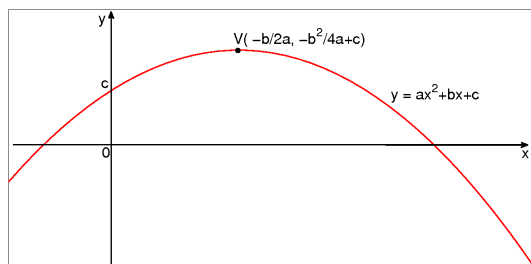


Fig. 7. Illustration of the shape of a parabola indicating the coordinates of the vertex  $V$ . Notice how coefficient  $c$  determines only the vertical position of the parabola but does not influence the concavity nor the horizontal position.

## VI. DISCUSSION AND FUTURE WORK

The purpose of this work has been to move a step forward towards an olfactory robot that can be deployed in a variety of settings and is therefore more suitable for real world applications. Indeed the intended use of mobile olfactory robots within search and rescue often involves environments that are difficult to predict in advance and for this reason it is important to derive from a limited amount of data, collected in a limited amount of situations, a model that is as general as possible. Clearly, the optimal solution would be to be able to extract features from the sensors response that are dependent only on the target analyte. However this task is extremely difficult given the multivariate nature of the sensor data and the difficulty to capture the properties of a response that are due to the target analyte versus the ones that depend on contingencies (airflow, temperature, humidity, interaction with the plume). For example, temperature and humidity which are assumed to have varied from the indoor experiments to the outdoor experiments, and between experimental runs in the outdoor experiments depending on weather conditions were not explicitly measured nor taken in account. Future investigations could measure explicitly these variables and take them into account in the algorithmic process to further improve the feature selection process. Nevertheless, in this work the starting point was to extract relevant features that

were not intrinsically coupled to deployment of the sensor on a mobile robot nor on the motion of the robot per se. Although in the presented works the robot was following a pre-defined path, the complexity of the problem is the same if the robot were to use another movement strategy.

To sum, this paper proposes two feature selection algorithms that improve the performance of an olfactory robotic system. Indeed one of the most crucial abilities for an olfactory robot is to be able to perform well in an unknown situation as is the case with other mobile robotic applications such as long term localization of an outdoor robot that has to cope with variations in the data due to different seasons [16]. In this work the unknown parameters include the properties of the environment as well as the task and behavior of the robot (e.g, different exploration strategies). The algorithms have been validated on a large data set of experiments performed in different uncontrolled conditions. Future works will include the merging of the discrimination ability with the other artificial olfaction tasks, namely source localization and gas distribution mapping.

## REFERENCES

- [1] J. W. Gardner and P. N. Bartlett, *Electronic Noses: Principles and Applications*, O. U. Press, Ed. Oxford Science Publications, 1999.
- [2] F. Li, Q.-H. Meng, J.-W. Sun, S. Bai, and M. Zeng, "Single odor source declaration by using multiple robots," in *Proc. AIP, The American Institute of Physics*, vol. 1137, 2009, pp. 73–76.
- [3] H. Ishida, "Robotic systems for gas/odor source localization: Gap between experiments and real-life situations," in *Proceedings of the IEEE International Conference on Intelligent Robots and Systems, 2005. (IROS 2005)*, 2007, pp. 3–8.
- [4] M. Vergassola, E. Villermaux, and B. I. Shraiman, "'Infotaxis' as a strategy for searching without gradients," *Nature*, vol. 445, pp. 406–409, January 2007.
- [5] M. Reggente and A. J. Lilienthal, "Using local wind information for gas distribution mapping in outdoor environments with a mobile robot," in *Proceedings of IEEE Sensors*, 2009.
- [6] M. Trincavelli, S. Coradeschi, and A. Loutfi, "Odour classification system for continuous monitoring applications," *Sens. Actuators B: Chem.*, vol. 139, pp. 265–273, 2009.
- [7] A. Hierlemann and R. Gutierrez-Osuna, "High-order chemical sensing," *ASC Chemical Reviews*, in press, 2008.
- [8] D. Martinez, O. Rochel, and E. Hughes, "A biomimetic robot for tracking specific odors in turbulent plumes," *Autonomous Robots*, vol. 20, pp. 185–195(11), June 2006.
- [9] N. Nimsuk and T. Nakamoto, "Improvement of capability for classifying odors in dynamically changing concentration using qcm sensor array and short-time fourier transform," *Sensors and Actuators B: Chemicals*, vol. 127, pp. 491–496, May 2007.
- [10] M. Trincavelli, S. Coradeschi, and A. Loutfi, "Classification of odours for mobile robots using an ensemble of linear classifiers," in *Proc. AIP, The American Institute of Physics*, vol. 1137, 2009, pp. 475–478.
- [11] A. J. Lilienthal and T. Duckett, "A stereo electronic nose for a mobile inspection robot," in *Proceedings of the IEEE International Workshop on Robotic Sensing (ROSE)*, Örebro, Sweden, 2003.
- [12] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer, August 2006.
- [13] V. Cherkassky and F. Mulier, *Learning from Data: Concepts, Theory and Methods*. Wiley Inter-Science, 1998.
- [14] I. Guyon, "An introduction to variable and feature selection," *Journal of Machine Learning Research*, vol. 3, pp. 1157–1182, 2003.
- [15] B. Gerkey, R. T. Vaughan, and A. Howard, "The Player/Stage Project: Tools for Multi-Robot and Distributed Sensor Systems," in *Proceedings of the IEEE International Conference on Advanced Robotics (ICAR)*, 2003, pp. 317–323.
- [16] C. Valgren and A. J. Lilienthal, "Sift, surf and seasons: Long-term outdoor localization using local features," *Robotics and Autonomous Systems*, to appear, 2009.