Available online at www.sciencedirect.com
SCIENCE DIRECT
Sensors and Actuators B 116 (2006) 17–22
SENSORS AND ACTUATORS B CHEMICAL
www.elsevier.com/locate/snb

A dimensionality-reduction technique inspired by receptor convergence in the olfactory system

A. Perera *, T. Yamanaka, A. Gutiérrez-Gálvez, B. Raman, R. Gutiérrez-Osuna

Department of Computer Science, Texas A&M University, College Station, TX, USA

Received 11 July 2005; received in revised form 3 October 2005; accepted 23 November 2005
Available online 2 May 2006

Abstract

In this paper, we propose a new technique for feature extraction/selection based on the projection of sensor features in class space while taking into account the sensor variance. The proposed technique is inspired by the organization of the early stages in the biological olfactory system. The algorithm proves to be highly suitable for high-dimensional feature vectors. The performance shows robustness with problems where only a small number of samples are available as a training dataset. We demonstrate the method on experimental data from two metal oxide sensors driven by a sinusoidal temperature profile.

Keywords: Bioinspired; High dimensionality; Electronic nose; Feature selection; Olfactory model; Gas sensor

1. Introduction

In a large number of applications using smart chemical sensors (sometimes referred to as electronic noses), the goal is defined as a categorization into a limited number of discrete classes from the information given by a set of descriptors or features. In these applications, statistical pattern recognition and other signal processing algorithms are common tools applied in several stages right after signals are translated to the digital domain. Usually, stages are split into filtering, signal preprocessing, feature extraction or feature selection and a final stage that classifies unknown input patterns into known previously trained classes. For some applications, where the output should be quantitative, the last stage is known as a regressor. Other approaches are conceptually different than classifiers/regressors like those found in event detection or change point detection applications. However, all approaches share a strong dependence on the quality of the early stages, predominantly feature extraction and feature selection.

On the other hand, certain applications add additional difficulties for a proper design of the pattern recognition algorithms. A particular case is found at high-dimensional datasets, where the number of features generated by the instrument is sufficiently high to impede the use of classical algorithms. These characteristics typically translate into poor performance in validation, therefore showing a lack of generalization of the algorithm to unseen data.

In chemical sensing, several groups have proposed different alternatives that generate high-dimensional data from chemoresistive sensors. A first example is the dynamic information extracted from the evaluation of sensor transients [1,2] or sensor responses under certain temperature modulation profiles [3]. However, other authors expanded chemical information by means of a different sensor technology foundation. In 1996, White et al. presented a new sensor device built with an array of fiber-optic based chemosensors. Changes in dye fluorescence were recorded using a CCD device obtaining 256 channels [5]. High-density configurations reaching 20 *k*-fibers were achieved 2 years later by Michael et al. [6]. Other developments explore analytical based instruments that also provide with large dimensionalities. For instance, laser based ion mobility spectrometer instruments (IMS) can provide a set of time-of-flight spectrums for the same sample when varying the laser frequency [7]. These works hint at the future availability of sensor systems providing large dimensionalities.

It is known that high-dimensional spaces penalize the generalization performance of most classifiers. In 1968, Hughes showed that with a fixed design pattern sample, recognition

* Corresponding author. Tel.: +1 34 934016963; fax: +1 34 934017045.
E-mail address: aperera@cs.tamu.edu (A. Perera).

0925-4005/$ – see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.snb.2005.11.082

accuracy increased with the number of measurements made, but decreased when the measurement complexity was higher than some optimum value [4]. This points out to an optimum number of features for a given a training set size which, unfortunately, is not known a priori. Furthermore, it has been shown that data distributed in high-dimensional spaces has some interesting properties. Jimenez and Landgrebe proved that the volume of a hypercube concentrates in the corners as the dimensionality increases, suggesting than most of the multivariate dataset space is empty and that data distribution on high-dimensional spaces might be counterintuitive, making density estimation a more difficult problem [8]. Additionally, the required size of the training set increases as a function of the dimensionality, linearly for a linear classifier and to the square of the dimensionality for a quadratic classifier [9]. However, typical algorithms found in chemical sensor array literature for feature extraction (e.g., Fisher's linear discriminant analysis) or selection (e.g., sequential forward floating selection) are prone to over-fitting or computational ill-conditioning when the ratio of dimensionality to samples is large.

On the other hand, very high-dimensional structures are also found on the biological side. In the mammalian olfactory bulb, a huge amount of olfactory sensory neurons (OSN) are found showing different selectivities with a degree of redundancy. It is know that mammals develop around 2 million olfactory sensory neurons where each neuron expresses only one type of odorant receptor gene out of a repertoire of up to 1000 genes [12,13].

It is also known that inside the olfactory signal pathway, signals are processed through a topographic mapping where OSN expressing the same receptor project onto a single or few glomeruli [11], as is schematically shown in Fig. 1.

Additionally, to the interest in the basic understanding of the internal structure of the biology, the study, characterization and modeling of this signal pathway, can lead to the development of neuromorphic signal-processing techniques. These techniques can be applied to gas sensor array signals with a dual purpose: to achieve a better understanding of the signal processing on the biological side, or to propose alternative coding and processing techniques to the classical pattern recognition applied to chemical sensors [14,15].

In this paper, we propose an alternative statistical formulation inspired by the biological convergence principle by means of a convergence map created from the training set. We demonstrate its suitability for high-dimensional problems with small training set sizes. The method is shown with the help of a synthetic based and a metal oxide based dataset. The synthetic dataset is used for the characterization of the algorithm at different training set sizes. The metal oxide dataset shows the characterization of the algorithm at different dimensionalities.

2. Method

The proposed method creates a set of descriptors for each feature in order to compute similarities between features, as opposite to compute similarities to patterns. This is inspired by the early stage of the olfactory system were OSN expressing the same receptor converge onto one or few glomeruli. It can be understood that biology converges (clusters) signals from different receptors (features) into different glomeruli (feature groups) depending on the receptor class. The convergence connectivity process itself controls receptor redundancy balancing noise reduction and discrimination. The proposed algorithm is split in a two-stage process. First two quality measures are computed so that these describe the behavior of a given feature against the different classes in terms of (1) discrimination and (2) confidence in the discrimination power of the feature. These two qualities are computed in the so-called class space as described later in this section. The second stage groups features according to these two quality measures. An alternative method is proposed substituting the second stage with a linear projection, based in a per-feature weight, computed from the relationship between feature discrimination and its confidence. The later alternative can be understood as feature extraction whereas the former can be stated as feature selection, as it also generates a number of feature-sets with different levels of discrimination power.

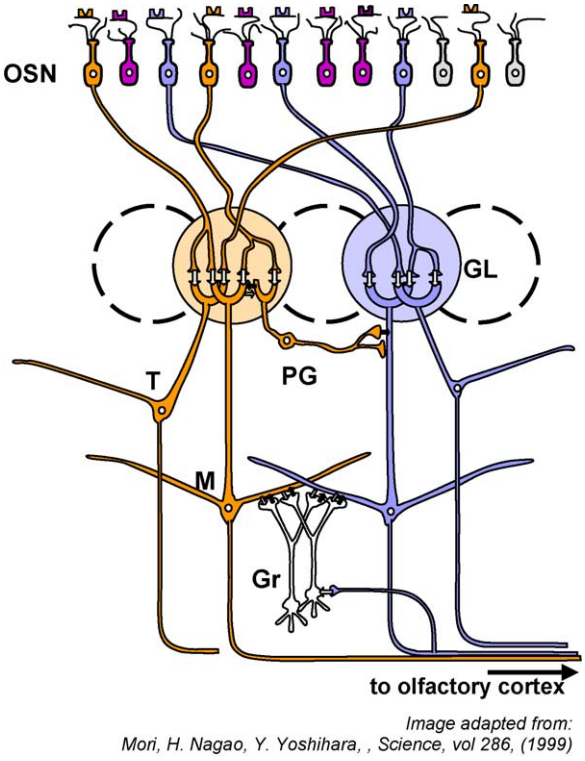


Fig. 1. Schematic of the signal pathway in the olfactory system. Olfactory sensory neurons (OSN) expressing the same genes converge onto the same glomeruli (GL) units. Similarly, the proposed algorithm cluster features showing similar behavior against dataset classes. Image adapted from ref. [10].

We can define the response of the sensor system as a vector $\boldsymbol{x}$ in a D-dimensional feature space $\Re^D$, where D corresponds to the complete number of features extracted from the sensor array. Assuming that all classes have Gaussian likelihoods, data distribution for the training set can be described by the mean response of a given sensor (or feature) k to class c, $\mu_{c,k}$ and its standard deviation $\sigma_{c,k}$. Assuming C classes, a vector μ_k in a C-dimensional *class space* $\Re^C$ is defined to represent the mean response of each feature across all classes, where $\mu_k = \{\mu_{1,k}, \ldots, \mu_{c,k}\}$. Note that this *class space* is the dual to the conven-

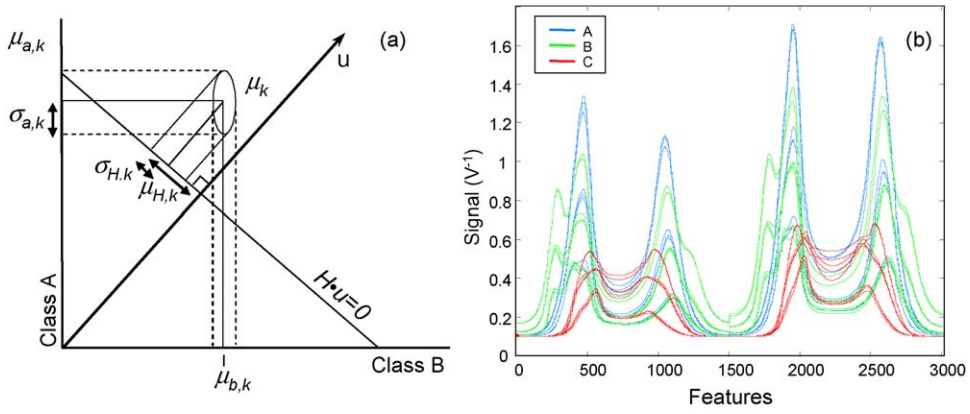


Fig. 2. (a) Illustration of the projection of features ion the subspace $\boldsymbol{H}$, orthogonal to $\boldsymbol{u}$. (b) Response of two sensors under a sinusoidal temperature profiles and three analytes.

tional *feature space*, in which an odor sample is represented by its response across all sensors. The standard deviations for each feature ($\sigma_{c,k}$) define confidence figures for μ_k values. This confidence values can be generally expressed by a covariance matrix Σ_k in $\Re^C$.

The variance of a given feature k in class space is symmetric with respect to the axis. This is due to the fact that a feature can be placed in *class space* using any pattern from class c_i training set against any pattern from class c_j training set. This enforces a diagonal covariance matrix $\Sigma_{k,c} = \sigma_c^2$.

The projection of the features into class space provides some interesting options due to the space properties. Features alongside the hyper-diagonal $\boldsymbol{u} = [1\,1\,\ldots\,1]^{\mathrm{T}}$ (see Fig. 2(a)) in class space contain no discriminatory information since they provide the same response to all the classes. Therefore, a measure of the discriminatory information can be obtained by the projection of the mean vector μ_k onto the subspace $\boldsymbol{H}$, namely $\mu_{\boldsymbol{H}k}$, where $\boldsymbol{H}$ is a $C-1$ dimensional subspace orthogonal to $\boldsymbol{u}$. Identically, to measure the confidence in the information content of feature k, we project the covariance matrix Σ_k onto the subspace $\boldsymbol{H}$ orthogonal to $\boldsymbol{u}$ (Fig. 2(a)), obtaining the projected mean $\mu_{\boldsymbol{H}k}$ and projected $\Sigma_{\boldsymbol{H}k}$, both in $\Re^{C-1}$. Information about the discriminatory information, represented by the mean $\mu_{\boldsymbol{H}k}$, and its uncertainty, represented by its variance $\Sigma_{\boldsymbol{H}k}$, can be combined to find a set of values $\boldsymbol{w}_k = f(\mu_{\boldsymbol{H}k}, \Sigma_{\boldsymbol{H}k})$.

The selection of $f(\cdot)$ determines the method used to weight the discriminatory information once the uncertainty of the affinities is known from the training set. In the following, we propose a heuristic method for $f(\cdot)$, although other options are possible. This heuristic is based on the intuitive idea that the relative relationship between $\mu_{\boldsymbol{H}k}$, and the projection of $\Sigma_{\boldsymbol{H}k}$ on $\boldsymbol{H}$-axis, $\sigma_{\boldsymbol{H}k}$, determine the confidence on the value of $\mu_{\boldsymbol{H}k}$. This is computed element-wise using the function $f(\mu_{\boldsymbol{H}k}, \sigma_{\boldsymbol{H}k}) = \mu_{\boldsymbol{H}k} \exp(|\mu_{\boldsymbol{H}k}| - \sigma_{\boldsymbol{H}k})$, which rewards features that are distant from the hyper-diagonal $\boldsymbol{u}$ and have low standard deviation.

Next, we propose two different alternatives in the use of the resulting factors. For feature selection, similar discriminant vectors $\boldsymbol{w}_k$ are grouped using a partition of the subspace $\boldsymbol{H}$ (referred as ConvI). As a result, features with similar behavior (similar affinities to the set of C classes) are clustered together, in the same manner that olfactory receptor neurons expressing the same receptor converge onto the same glomeruli. Note that this is done via a *partition* of $\boldsymbol{H}$, as opposite to a *clustering* (e.g., with a self organizing map, SOM) in order to avoid a density estimation process in class-space.

Estimation methods like a SOM would split high-density clusters creating multiple nodes although they would provide very similar discriminative power. However, SOM based or k-means based clustering of the features can be beneficial in certain situations.

A second algorithm is proposed by using the set of factors $\boldsymbol{w}_k$ directly as a projection matrix, W. This projection maps $\Re^D$ into $\Re^{C-1}$ in a feature extraction sense (we refer to this method as ConvII). The first method needs the assignment of a feature to a group or partition in the subspace $\boldsymbol{H}$. This step can be computationally expensive when the number of classes defined in the problem is high (class space will show high dimensionalities). The second method avoids this partition and therefore will be computationally cost-effective.

3. Robustness to over-fitting

Both PCA and LDA use the data covariance for building the linear projection. In the case of PCA, the complete dataset covariance is computed whereas LDA computes per-class scatter matrices. These covariance matrices result in strong computational issues under high-dimensional datasets. The proposed algorithm evades covariance matrices computation and builds a linear projection by using only variance–mean relationship information. This is important in the case where covariance based methods are prone to over-fit the training dataset, degrading their generalization properties. In order to show this property, a synthetic experiment was designed with three normal distributions generated in a three-dimensional space. The dimensionality was extended by padding feature vectors with normal noisy channels $N(0,\alpha)$ (with $\alpha = 0.1$), to achieve a 200 dimensional space. Each class was generated with a normal distribution $N(\mu, \Sigma)$ with a selection of parameters that provides with a dataset with discriminatory information in both mean and covari-

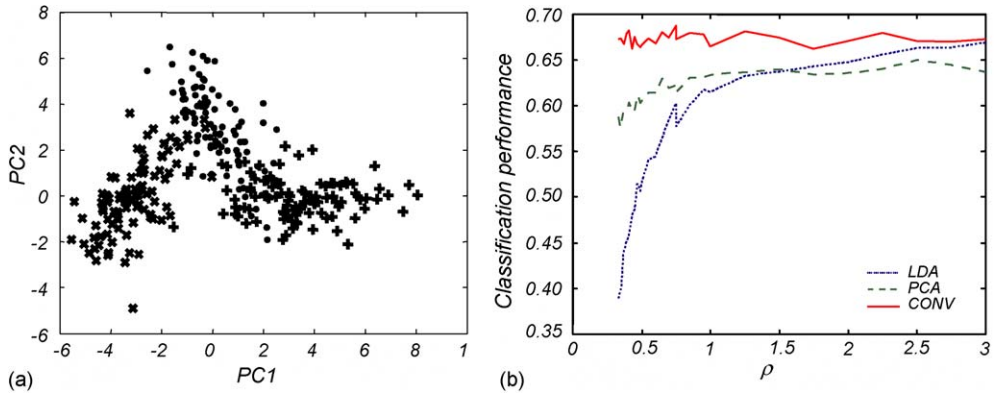


Fig. 3. (a) PCA score plot of the synthetic dataset. The synthetic problem consists of three Gaussian distributions with different mean and covariance. (b) Comparative response of PCA, LDA and convergence (method II) under over-fitting conditions.

ance, and defined by,

$$\sum_1 = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 0.6 & 0 \\ 0 & 0 & 0.1 \end{bmatrix} \sum_2 = \begin{bmatrix} 1.45 & -1.47 & 0 \\ -1.47 & 3.15 & 0 \\ 0 & 0 & 0.1 \end{bmatrix} \sum_3 = \begin{bmatrix} 1.45 & 1.47 & 0 \\ 1.47 & 3.15 & 0 \\ 0 & 0 & 0.1 \end{bmatrix}$$
$$\mu_1 = \begin{bmatrix} -3 \\ 0 \\ 2 \end{bmatrix} \mu_2 = \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix} \mu_3 = \begin{bmatrix} 0 \\ 3 \\ 0 \end{bmatrix} \quad (1)$$

A PCA plot illustrating the distribution of the dataset is shown in Fig. 3(a). Notice that both, distribution mean and covariance is different for each class.

The performance of a k-NN classifier in validation was computed after a reduction to two principal components using PCA, LDA and convergence (ConvII). The number of samples per class used in the training set was varied from 66 to 600, corresponding to a $\rho = 0.33$–3, where ρ is defined as the number of samples per class over the dimensionality.

Classification ratio was computed by means of a validation set with 1000 samples generated with the aforementioned distributions. Each experiment was repeated a total of 100 times. Results are shown in Fig. 3b. Note that PCA holds a stable performance for all computed values of ρ. On the other side, although convergence and LDA provide similar classification rates at $\rho = 3$, there is a noticeable performance drop of LDA at low ρ due to ill-conditioned with-in scatter matrix (S_w^{-1}) in the Fisher criterion. Results for this simple setup suggest that convergence achieves similar performance to LDA, and even improves its performance avoiding data over-fitting effects.

4. Results

Proof-of-concept for the proposed method is illustrated with experimental data from two metal oxide sensors modulated in temperature. The sensors were exposed to dilutions of three different analytes (A–C), and their responses under a sinusoidal temperature profile (0–7 V; 2.5 min period; 10 Hz sampling frequency) were recorded (Fig. 2(b)). As a result, the input space dimensionality was $D = 3000$. The proposed feature selection algorithm was performed with a uniform grid of four units covering the subspace $\boldsymbol{H}$. As shown in Fig. 4, the algorithm splits the temperature profile into different regions, which depend on their discriminatory information. Note that peaks characteristics of class B are grouped together under the G1 group, and features that provide low response to C but high response to A are grouped as G3. Overall the method is providing a rich combinatorial selection of which features are helpful for obtaining discriminant projections.

After the convergence mapping is constructed, training and test data can be projected using grouping data using the convergence information, $g_i = (1/N)\sum_{k=1}^{D} c_{ik} x_k$, where g_i is the output of the group i, N_i the number of features grouped in the set i, x_k the feature k and c_{ik} takes 1 if the feature k converges to set i and 0 otherwise.

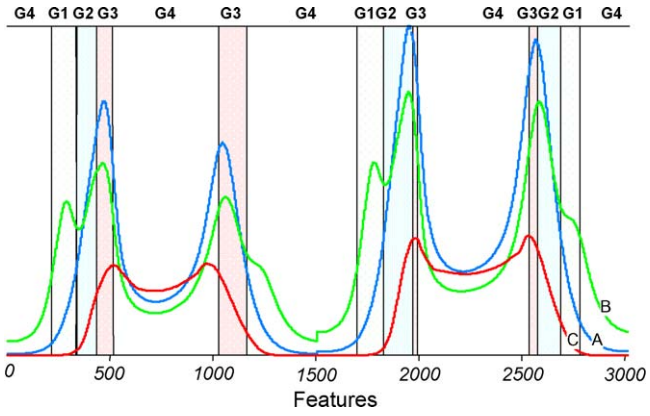


Fig. 4. Result of the feature grouping on class-space. For visualization purposes, only the response to the highest concentration to each analyte is shown. Zones corresponding to the different grouped features by the convergence algorithm are marked as G1–G4 (y-axis omitted for clarity).

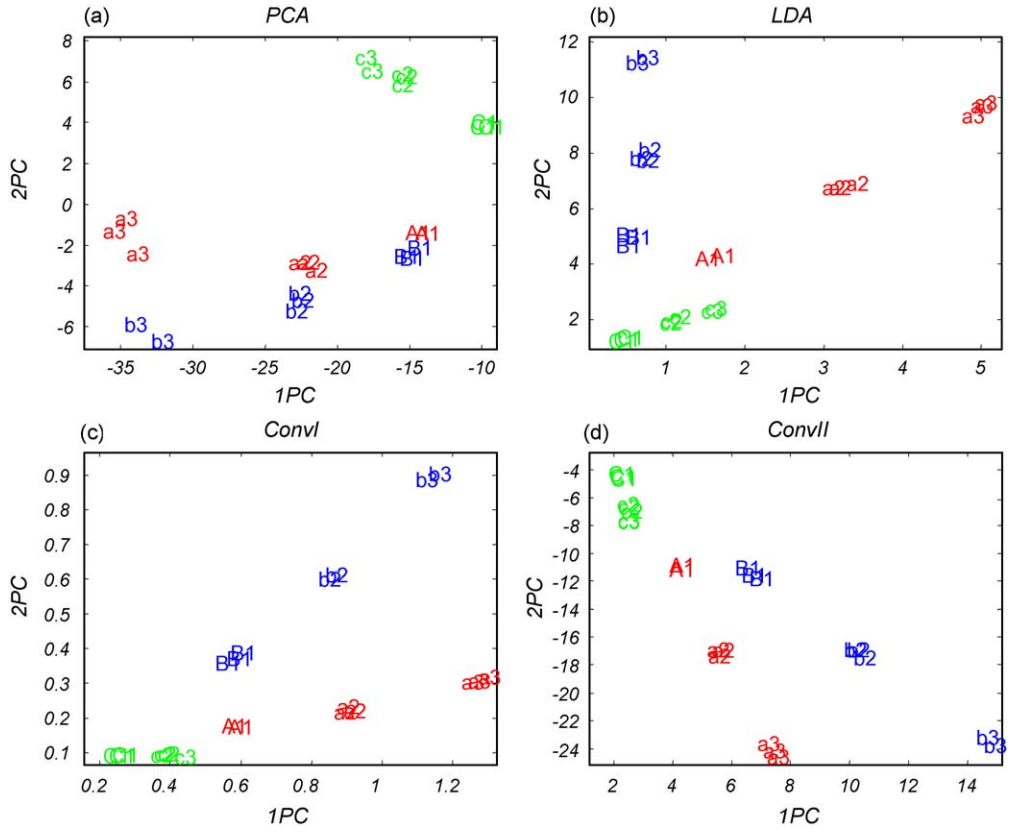


Fig. 5. Scatter-plots for the four methods. (a) Principal component analysis, (b) linear component analysis, (c) convergence type-I and (d) convergence type-II.

In Fig. 5, we show the score plots from the output of PCA (a) and LDA (b) compared with those generated by the two algorithms proposed (ConvI in (c) and ConvII in (d)). LDA projection was computed using a decimation of the waveform of the order 1:6. In the four algorithms, the models were trained using the lower dilutions (numbered as A1, B1 and C1) and the rest of dilutions were projected onto the respective models (2 and 3) as validation samples. In the PCA plot, it is noticeable that analytes A and B are confused at low concentration (dilutions 1 and 2) whereas in LDA those are correctly separated. The convergence based algorithms also separate the three analytes and show a similar behavior as LDA, with the difference that Convergence was computed using the complete sensor waveforms $D = 3000$.

5. Conclusions

The proposed algorithms shown in this paper are a result of two efforts: the study of the convergence seen from the population of olfactory sensory neurons to the glomerular layer form a signal processing view, and the construction of this convergence under the constrains given by the existence of a training set available to build the convergence map. This idea has led to the proposal of an algorithm based in grouping of features in class-space constructed with information that takes into account the relationship between mean and variance for each feature. An alternative method is derived from the first that avoids the feature clustering and builds a direct linear projection from the class space to a $C-1$ dimensional space. The algorithms are computationally efficient under high dimensionalities and well suited for small-sample-set input spaces since they do not involve the computation of covariance matrices in feature space.

Further work evaluate the method's generalization characteristics to other fields like image processing or genomics.

Acknowledgment

This work was supported by the National Science Foundation under CAREER Grant No. 9984426/0229598.

References

[1] Y. Hiranaka, T. Abe, H. Murata, Gas dependant response in the temperature transient of SnO_2 gas sensors, Sens. Actuators B, Chem. 9 (3) (1992) 177.

[2] R. Gutierrez-Osuna, H.T. Nagle, S.S. Schiffman, Transient response analysis of an electronic nose using multi-exponential models, Sens. Actuators B, Chem. 61 (1–3) (1999) 170–182.

[3] A.P. Lee, B.J. Reedy, Temperature modulation in semiconductor gas sensing, Sens. Actuators B, Chem. 60 (November (1)) (1999) 35–42.

[4] G.F. Hughes, On the mean accuracy of statistical pattern recognizers, IEEE Trans. Inf. Theory 14 (1) (1968) 55–63.

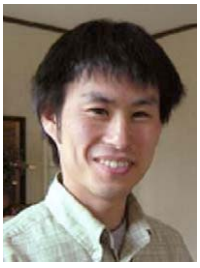
[5] J. White, J.S. Kauer, T.A. Dickinson, D.R. Walt, Rapid analyte recognition in a device based on optical sensors and the olfactory system, Anal. Chem. 68 (1996) 2191–2202.

[6] K.L. Michael, L.C. Taylor, S.L. Shultz, D.R. Walt, Randomly ordered addressable high-density optical sensor arrays, Anal. Chem. 70 (1998) 1242–1248.

[7] F. Mühlberger, K. Hafner, S. Kaesdorf, T. Ferge, R. Zimmermann, Comprehensive on-line characterization of complex gas mixtures by quasi-simultaneous resonance-enhanced multiphoton ionization, vacuum-UV single-photon ionization, and electron impact ionization in a time-of-flight mass spectrometer: setup and instrument characterization, Anal. Chem. 76 (2004) 6753–6764.
[8] L. Jimenez, D. Landgrebe, Supervised classification in high-dimensional space: geometrical, statistical and asymptotical properties of multivariate data, IEEE Trans. Syst. Man Cybertetics C 28 (January) (1998) 39–54.
[9] K. Fukunaga, Effects of sample size in classifier design, IEEE Trans. Pattern Anal. Machine Intell. 11 (8) (1989) 873–885.
[10] K. Mori, H. Nagao, Y. Yoshihara, The olfactory bulb: coding and processing of odor molecule information, Science 286 (October (5440)) (1999) 711–715.
[11] R. Vassar, S.K. Chao, R. Sitcheran, J.M. Nuñez, L.B. Vosshall, R. Axel, Topographic organization of sensory projection to the olfactory bulb, Cell 79 (December) (1994) 981–991.
[12] R. Axel, The molecular logic of smell, Sci. Am. 273 (4) (1995) 154–159.
[13] R. Vassar, S.K. Chao, R. Sitcheran, J.M. Nuñez, L.B. Vosshall, R. Axel, Topographic organization of sensory projection to the olfactory bulb, Cell 79 (December) (1994) 981–991.
[14] R. Gutierrez-Osuna, A self-organizing model of chemotopic convergence for olfactory coding, in: Proceedings of the second Joint EMBS-BMES Conference, Houston, TX, 23–26 October, 2002.
[15] B. Raman, R. Gutierrez-Osuna, A. Gutierrez-Galvez, A. Perera, Sensor-based machine olfaction with a neurodynamics model of the olfactory bulb, in: Proceedings 2004 IEEE/RSJ International Conference on International Robots and Systems, Sendai, Japan, September 28–October 2, 2004.

Biographies

Alex Perera received his BS (1996) and his PhD (2003) in physics from University of Barcelona and his BS in electrical engineering (2001) from the same university. He was with Texas A&M University in the period from 2003 to 2005 as PostDoctoral Research Associate. His research interests include applied pattern recognition focusing chemical sensors, machine olfaction and bioinformatics.



Takao Yamanaka received the Bachelor of Engineering, the Master of Engineering, and the PhD in engineering from Tokyo Institute of Technology in 1996, 1998, and 2004,respectively. He was with Canon, Inc. from 1998 to 2000 in the area of image and signal processing. He is currently a postdoctoral fellow at Texas A&M University, supported by JSPS postdoctoral fellowship for research abroad. His research interests include neuromorphic engineering, intelligent sensors, machine learning, computational neuroscience, and machine olfaction.

Agustin Gutierrez-Galvez received his BS in physics and BS in electrical engineering in 1995 and 2000, respectively, both from the Universitat de Barcelona. He received his PhD degree in computer engineering at Texas A&M University in 2005. His research interests are neuromorphic models of the olfactory system and pattern recognition applied to odor detection with sensor arrays.

Baranidharan Raman received the Bachelor of Engineering in computer science with distinction from the University of Madras in 2000, and the MS and PhD degrees in computer science from Texas A&M University in 2003 and 2005, respectively. His research interests include sensor based machine olfaction, intelligent systems, machine learning, neural computation, and robotics.

Ricardo Gutierrez-Osuna received the BS degree in electrical engineering from the Polytechnic University of Madrid in 1992, and the MS and PhD degrees in computer engineering from North Carolina State University in 1995 and 1998, respectively. From 1998 to 2002 he served on the faculty at Wright State University. He is currently an assistant professor of computer engineering at Texas A&M University. His research interests include pattern recognition, neuromorphic computation, machine olfaction and speech-driven facial animation.