

# Odor Recognition in Multiple E-nose Systems with Cross-domain Discriminative Subspace Learning

Lei Zhang, *Member, IEEE*, Yan Liu, and Pingling Deng

**Abstract**—In this paper, we propose an odor recognition framework for multiple electronic noses (E-nose), machine olfaction odor perception systems. Straight to the point, the proposed transferring odor recognition model is called cross-domain discriminative subspace learning (CDSL). General odor recognition problems with E-nose are single domain oriented, that is, recognition algorithms are often modeled and tested on the same one domain dataset (*i.e.*, from only one E-nose system). Different from that, we focus on a more realistic scenario: the recognition model is trained on a prepared source domain dataset from a master E-nose system A, but tested on another target domain dataset from a slave system B or C with the same type of the master system A. The internal device parameter variance between master and slave systems often results in data distribution discrepancy between source domain and target domain, such that single domain based odor recognition model may not be adapted to another domain. Therefore, we propose domain adaptation based odor recognition for addressing the realistic recognition scenario across systems. Specifically, the proposed CDSL method consists of three merits: 1) an intra-class scatter minimization and inter-class scatter maximization based discriminative subspace learning is solved on source domain. 2) a data fidelity and preservation constraint of the subspace is imposed on target domain without distortion. 3) a mini-patch feature weighted domain distance is minimized for closely connecting the source and target domains. Experiments and comparisons on odor recognition tasks in multiple E-noses demonstrate the efficiency of the proposed method.

**Index Terms**—Odor recognition, electronic nose, domain adaptation, cross-domain learning, subspace learning

## I. INTRODUCTION

ODOR recognition by using an electronic nose (E-nose) is an interesting but challenging issue in machine olfaction. The challenging aspect lies in the technical gap of olfactory sensors, with respect to the vision sensors (e.g., imaging sensor). In machine olfaction community, E-nose plays an important role in odor perception and data analysis based on pattern recognition algorithms [1, 2]. E-nose, as a cross-sensitive odor perceptron device with intelligent signal processing and pattern

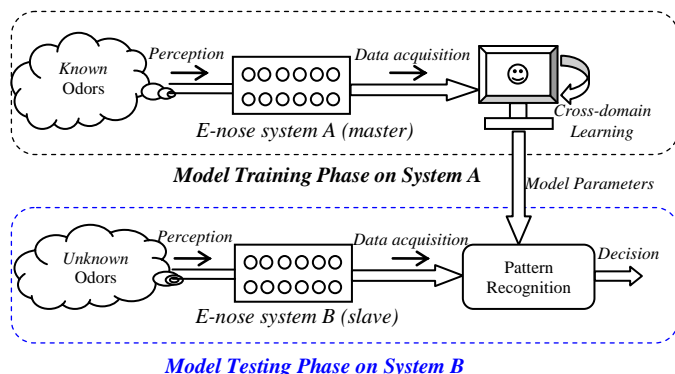


Fig. 1. Diagram of the proposed odor recognition framework. The model training (cross-domain learning) is implemented in PC based on the acquired odor data from master system A, and the well-trained model parameters will be used to recognize the odors from slave B.

recognition units, has witnessed a wide progress in systems, applications, and algorithms during the past two decades [3-6]. Specifically, Flammini *et al.* proposed a low-cost interface to high-value resistive sensors over a wide range, such that a wide detection range is possible [7]. Brudzewski *et al.* [8] proposed a differential electronic nose for recognition of coffees. Herrero-Carrón *et al.* [9] proposed an active and inverse temperature modulation E-nose for odorant classification. Gosangi and Gutierrez-Osuna [10] proposed an active temperature programming method for odor recognition. Yin *et al.* [11] also proposed a temperature modulation method in E-nose for gases recognition, and the recognition accuracy with fewer sensors can also be guaranteed by modulation.

In odor recognition, a number of pattern recognition algorithms for classification and regression have been presented for E-nose [12-15], such as support vector machines (SVM), neural networks (ANN), discriminant analysis (DA), learning vector quantization (LVQ), etc. Tudu *et al.* [16] proposed an incremental fuzzy approach for classification of black tea quality with an E-nose, such that the newly presented patterns can be automatically included in the training set because of the incremental learning ability. Zhang *et al.* [17] proposed a hybrid linear discriminant analysis based support vector machine method for classification of six kinds of air contaminants, and achieved the best accuracy. For those readers of interest, some excellent reviews in machine olfaction

This work was supported in part by National Natural Science Foundation of China (Grant 61401048) and in part by the research fund for Central Universities.

L. Zhang, Y. Liu and P. Deng are with the College of Communication Engineering, Chongqing University, Chongqing 400044, China. (e-mail: leizhang@cqu.edu.cn; yanliu@cqu.edu.cn; dengpl@cqu.edu.cn)

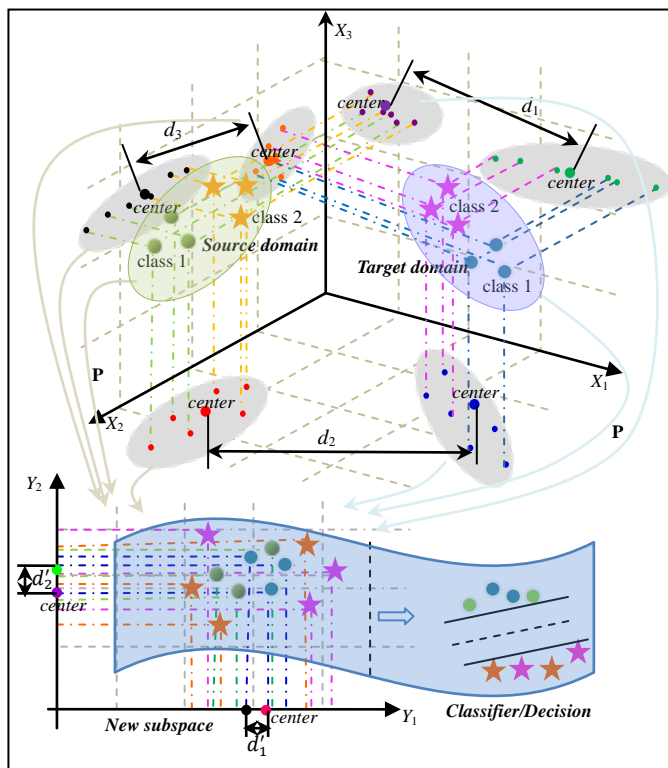


Fig. 2. Schematic diagram of the proposed CDSL method; after a subspace projection  $\mathbf{P}$ , the source domain and target domain of different space distribution lie in a latent subspace with good distribution consistency (the centers of both domains become very close and drift is removed); in this latent subspace, the classification of two classes is successfully achieved. Formally, the upper coordinate system denotes the raw data points of source domain and target domain in three dimensions. We use the word “center” to represent the mean of the features. From the upper figure, we can see that the difference between the mean of source domain and the mean of target domain is large in each dimension. After a subspace projection  $\mathbf{P}$  in the below figure, we can see that the values of  $d'_1$  and  $d'_2$  become smaller, which demonstrate that the distribution difference becomes small, and both domains of different space distribution lie in a latent common subspace with good distribution consistency.

and E-nose that have described pattern classification and signal processing methods can be referred to as [18,19].

By reviewing these research findings described above, we can observe a common property, that is, all these works tend to study highly efficient and effective odor recognition algorithms in a single E-nose system. This can be recognized to be method driven. It is useful to reveal some better algorithms for improving the odor recognition performance of an E-nose. However, in real application scenarios, multiple E-nose systems of the same type would be developed for odor detection and recognition. An essential issue that we should pay attention to is that although multiple systems are of the same type, the well-trained recognition algorithm based on one system cannot be easily adapted to another system. An explicit reason is the internal output differences (e.g., signal shift) between gas sensor arrays in multiple systems. Another reason is the slow *instrumental aging* of sensors (e.g., signal drift) when exposed to air for a long time. It is just the challenging

aspect of machine olfaction claimed at the beginning of this section. The undesired result caused by this issue is that the obtained machine learning algorithm based on E-nose system  $A$  cannot be transferred to another E-nose system  $B$  of the same type. Therefore, the diverse application of E-noses is seriously restricted. This is exactly what we are paying attention to and aiming to solve in this paper.

In terms of the non-transferrable restriction of E-nose in odor recognition, we propose a novel *transferring odor recognition* framework, which targets at odor recognition across systems. Specifically, the diagram of the proposed odor recognition framework is described in Fig. 1, which includes two parts: model training phase on system  $A$  (defined as master) and model testing phase on system  $B$  (defined as slave). To effectively address this issue, inspired by transfer learning [20] and domain adaptation [21] in machine learning community, each system is treated as a domain. Thus, the master E-nose system is viewed as source domain and other slave systems are viewed as target domains in this paper. From the viewpoint of domain distribution (i.e., data distribution), the data distribution between source domain (master system) and target domain (slave system) is different. That is, the data in source and target domain lie in different feature spaces. Therefore, we propose a cross-domain discriminative subspace learning (abbreviated as CDSL) method, by pursuit of a common (shared) subspace of both domain data, such that the data from different domains lie in the same (common) subspace. Then, with the proposed CDSL method, we are able to achieve transferring odor recognition across multiple systems (i.e., across domains).

Specifically, the merits of the proposed CDSL method are three-fold:

- ✧ The aim of the proposed CDSL is to learn a common subspace (projected by a transformation  $\mathbf{P}$ ) such that the source and target domains share similar feature distribution. Considering that our final task is classification, and the classification capability should be augmented in the learned subspace. Thus, in CDSL, a discriminative mechanism on the source domain data is integrated, by minimizing the intra-class scatter matrix and simultaneously maximizing the inter-class scatter matrix. That is, the class separability of the source domain data in the subspace can be further guaranteed with the discriminative learning mechanism.
- ✧ To preserve the structure of the target domain data in the learned subspace and avoid the target data distortion, a data fidelity constraint is imposed in CDSL. With this constraint, much available information in the target domain we are interested (e.g., drift knowledge) is effectively preserved.
- ✧ To learn a common subspace of source and target data, we propose a *source to target* domain distance minimization term in the learned subspace, such that the domain distribution consistency can be improved. Further, considering that the domain distribution is feature-specific, that is, different feature dimension may have different distribution discrepancy degree, we further propose a novel and effective mini-patch feature weighted domain distance representation. The weights measure the degree of feature mismatch between source and target domain. A larger

weight denotes a higher feature mismatch degree.

With above characteristics of the proposed CDSL, a common (shared), discriminative, and robust subspace projected by  $\mathbf{P}$  can be achieved. Further, the transferring odor recognition across domains (across multiple systems) can be improved with the proposed CDSL approach. Visually, the schematic diagram of the proposed CDSL is shown in Fig. 2, from which we can see that the ultimate goal is to achieve transferring odor recognition (classification) in the learned new subspace. The learning process of  $\mathbf{P}$  is completed by jointly modeling on the source domain and target domain, based on three key characteristics (merits): discriminative learning on source domain, data fidelity on target domain and mini-patch feature weighted domain distance.

The remainder of this paper is as follows. Section II presents the related works of shift calibration, drift correction and subspace learning. Section III describes the proposed cross-domain discriminative subspace learning (CDSL) approach including the model formulation, optimization and classification. The experiments and results have been discussed in Section IV. Finally, Section V concludes this paper.

## II. RELATED WORKS

### A. Shift Calibration (Reproducibility)

The signal shift in E-nose is an inherent problem, which results from the reproducibility of gas sensors [21]. Briefly, when two identical E-nose systems are exposed to the same conditions, their outputs are not the same. One reason is that the system is related to the physical condition such as temperature, humidity and pressure. Researchers have proposed different methods to calibrate the signal shift, such as global affine transformation based on robust weighted least square (GAT-RWLS) [22], the windowed piecewise direct standardization (WPDS) [23], partial least square regression (PLS) [24], and univariate direct standardization [25]. These methods are successful in shift calibration. However, the calibration model construction depends on the data amount, and thus the calibration becomes task specific. That is, for different odor recognition, the re-calibration should be conducted and also the generality is too weak. Additionally, the calibration is directly shown in signal magnitude, and independent of the subsequent learning algorithm.

### B. Drift Correction (Aging)

Drift, that is supposed to be some slow, continuous and uncertain effect, is affecting the classification performance of an E-nose. Sensor drift effect is caused by many objective factors such as aging, poisoning, and the fluctuations of the ambient environmental variables (*e.g.*, humidity, temperature) [25]. As a result, the instrument responds differently to a constant concentration of some contaminant at different ambient conditions. Drift is once thought to be an ill-posed problem due to its very irregular characteristics. Although researchers have proposed several methods to correct drift [26-31], the results are still unsatisfactory when both drift and shift happen. More importantly, these methods were proposed

for drift compensation independent of shift. Our proposed work aims to solve more complex transferring odor recognition across multiple systems (*i.e.*, drift plus shift). Specifically, Yan and Zhang [26] proposed an autoencoder method for time-varying drift correction. Big data should be used for training a deep model. Martinelli *et al.* proposed an artificial immune system based adaptive classifier for drift mitigation [27], however the classifier has no knowledge transfer ability. Vergara *et al.* proposed a classifier ensemble model for drift compensation, in which multiple SVMs with weighted ensemble are used [28]. The ensemble method can improve the classifier robustness, but it lacks transfer ability across systems. Zhang *et al.* [29] proposed a domain adaptation extreme learning machine based transfer learning method for drift compensation, which aims at proposing a robust classifier but lacks feature representation ability. Padilla *et al.* [30] proposed an orthogonal signal correction method for recognizing and removing the drift components, which imposes a strong assumption about the data property (*e.g.*, orthogonality). Di Carlo *et al.* [31] proposed to correct drift evolutionarily by learning a transformation with respect to the best pattern recognition accuracy, in which overfitting is easily caused due to the lack of necessary constraint on the transformation.

### C. Subspace Learning

Subspace learning aims at learning a low-dimensional subspace. Several popular subspace methods include principal component analysis (PCA) [32], linear discriminant analysis (LDA) [33], manifold learning based locality preserving projections (LPP) [34], and marginal fisher analysis (MFA) [35]. Specifically, PCA, as an unsupervised method, aims at preserving the maximum variance (energy) of the raw data. LDA is a supervised dimension reduction method, which targets maximizing the between-class scatter matrix trace and minimizing the within-class scatter matrix trace, such the linear separability is improved in the discriminative subspace. LPP is an unsupervised dimension reduction technique with manifold assumption and graph embedding, which preserves the local affinity structure in low dimensional embedding. MFA is recognized to be a comprehensive version of LDA and LPP, which integrates the intra-class compactness and the graph embedding. These subspace learning methods are applicable to single domain scenarios, but cannot be adapted to the transferring odor recognition scenario across domains. Therefore, cross-domain subspace learning model is desired for the proposed transferring scenario.

## III. CROSS-DOMAIN DISCRIMINATIVE SUBSPACE LEARNING

### A. Notations

In this paper, the source and target domain are defined by subscript “S” and “T”, respectively. The training data of source and target domain is denoted as  $\mathbf{X}_S = [\mathbf{x}_S^1, \dots, \mathbf{x}_S^{N_S}] \in \mathcal{R}^{D \times N_S}$  and  $\mathbf{X}_T = [\mathbf{x}_T^1, \dots, \mathbf{x}_T^{N_T}] \in \mathcal{R}^{D \times N_T}$ , respectively, where  $D$  is the dimensionality,  $N_S$  and  $N_T$  are the number of training samples in source and target domains.  $\mathbf{t}_S = [t_S^1, \dots, t_S^{N_S}]^T \in \mathcal{R}^{N_S}$  denotes the labels with respect to the source data  $\mathbf{X}_S$ . Let  $\mathbf{P} \in \mathcal{R}^{D \times d}$

represents a basis transformation that would map the source and target data from the raw space  $\mathfrak{R}^D$  to a new lower-dimensional subspace  $\mathfrak{R}^d$ . The symbol  $\|\cdot\|_F$  and  $\|\cdot\|_2$  denotes the Frobenius norm and  $l_2$ -norm, respectively.  $Tr(\cdot)$  denotes the trace operator and  $(\cdot)^T$  denotes the transpose operator. Throughout this paper, matrix is written in capital bold face, vector is shown in lower bold face, and variable is written in italics.

### B. Problem Formulation

As illustrated in Fig. 2, we aim to learn a basis transformation  $\mathbf{P}$  that maps the original space of source data  $\mathbf{X}_S$  and target data  $\mathbf{X}_T$  to a new subspace, *i.e.*,  $\mathbf{Y}_S$  and  $\mathbf{Y}_T$ , such that the feature distribution between the mapped source and target data  $\mathbf{Y}_S = [\mathbf{y}_S^1, \dots, \mathbf{y}_S^{N_S}] \in \mathfrak{R}^{d \times N_S}$  and  $\mathbf{Y}_T = [\mathbf{y}_T^1, \dots, \mathbf{y}_T^{N_T}] \in \mathfrak{R}^{d \times N_T}$  becomes similar. Therefore, it is rational to have an idea that the mean distribution discrepancy between  $\mathbf{Y}_S$  and  $\mathbf{Y}_T$  can be minimized. Considering that the discrepancy between domains is feature-specific (sensor-specific), that is, the discrepancy of each feature dimension between domains may be different in real application, we propose a mini-patch feature specific domain distance (MPD) to exactly describe the distribution discrepancy by imposing different weights on each mini-patch feature combination. Then, after domain adaptation, the MPD is expected to be minimized. Specifically, the proposed MPD minimization is formulated as follows.

$$\min \sum_{i=1}^M w_i \left\| \boldsymbol{\mu}_S^i - \boldsymbol{\mu}_T^i \right\|_2^2 = \min \sum_{i=1}^M w_i \left\| \frac{1}{N_S} \sum_{j=1}^{N_S} \mathbf{y}_S^{ij} - \frac{1}{N_T} \sum_{k=1}^{N_T} \mathbf{y}_T^{ik} \right\|_2^2 \quad (1)$$

where  $M$  denotes the size of mini-patch features,  $w_i$  denotes the weight of the  $i$ -th mini-patch,  $\boldsymbol{\mu}_S^i$  and  $\boldsymbol{\mu}_T^i$  denote the source and target domain center of the  $i$ -th mini-patch after domain adaptation.  $\mathbf{y}_S^{ij}$  denotes the projected source data sample  $j$  under the  $i$ -th mini-patch and  $\mathbf{y}_T^{ik}$  represents the projected target data sample  $k$  under the  $i$ -th mini-patch.

As can be seen from the MPD minimization, the weight  $w_i$  is expected to be larger if the  $i$ -th mini-patch has smaller distribution discrepancy. Therefore,  $w_i$  is computed as follows.

$$w_i = \frac{\sum_{j=1}^M \left\| \mathbf{u}_S^j - \mathbf{u}_T^j \right\|_2^2 / \left\| \mathbf{u}_S^i - \mathbf{u}_T^i \right\|_2^2}{\sum_{k=1}^M \sum_{j=1}^M \left\| \mathbf{u}_S^j - \mathbf{u}_T^j \right\|_2^2 / \left\| \mathbf{u}_S^k - \mathbf{u}_T^k \right\|_2^2} \quad (2)$$

where  $w_i$  is calculated based on between-center distance which can reflect the drift or discrepancy degree, and  $0 < w_i < 1$ .  $\mathbf{u}_S^i$  and  $\mathbf{u}_T^i$  represent the centers of the  $i$ -th mini-patch of source and target data, respectively. They can be computed as

$$\mathbf{u}_S^i = \frac{1}{N_S} \sum_{j=1}^{N_S} \mathbf{x}_S^{ij} \quad (3)$$

$$\mathbf{u}_T^i = \frac{1}{N_T} \sum_{k=1}^{N_T} \mathbf{x}_T^{ik} \quad (4)$$

where  $\mathbf{x}_S^{ij}$  denotes the  $j$ -th source data sample under the  $i$ -th mini-patch and  $\mathbf{x}_T^{ik}$  represents the  $k$ -th target data sample under the  $i$ -th mini-patch.

According to the subspace projection  $\mathbf{P}$ , the new representation of source and target data of the  $i$ -th mini-patch in the lower-dimensional subspace can be formulated as

$$\mathbf{Y}_S^i = \mathbf{P}^T \mathbf{X}_S^i = \mathbf{P}^T \left[ \mathbf{x}_S^{i,1}, \dots, \mathbf{x}_S^{i,N_S} \right] = \left[ \mathbf{y}_S^{i,1}, \dots, \mathbf{y}_S^{i,N_S} \right] \quad (5)$$

$$\mathbf{Y}_T^i = \mathbf{P}^T \mathbf{X}_T^i = \mathbf{P}^T \left[ \mathbf{x}_T^{i,1}, \dots, \mathbf{x}_T^{i,N_T} \right] = \left[ \mathbf{y}_T^{i,1}, \dots, \mathbf{y}_T^{i,N_T} \right] \quad (6)$$

Therefore, we have  $\mathbf{y}_S^{ij} = \mathbf{P}^T \mathbf{x}_S^{ij}$  and  $\mathbf{y}_T^{ik} = \mathbf{P}^T \mathbf{x}_T^{ik}$ . By substituting Eqs.(5) and (6) into Eq.(1), the minimization problem Eq.(1) can be reformulated as

$$\min_{\mathbf{P}} \sum_{i=1}^M w_i \left\| \frac{1}{N_S} \sum_{j=1}^{N_S} \mathbf{P}^T \mathbf{x}_S^{ij} - \frac{1}{N_T} \sum_{k=1}^{N_T} \mathbf{P}^T \mathbf{x}_T^{ik} \right\|_2^2 \quad (7)$$

The narrowed mean distribution discrepancy between source and target domain after using the learned projection  $\mathbf{P}$  can be guaranteed by minimizing (7). However, the discriminative property among multiple classes cannot be explicitly shown. That is, the separability of multiple odors in the source domain is not effectively described. Therefore, in the proposed CDSL model, we would also like to design a discriminative term that tends to minimize the trace of the intra-class scatter matrix  $\mathbf{S}_W^S$  and simultaneously maximize the trace of the inter-class scatter matrix  $\mathbf{S}_B^S$  of the source data, such that the intra-class compactness and inter-class separability can be improved. As a result, the classification benefits from discrimination becomes easier in the learned linear subspace. Specifically, we aim to maximize the ratio between the inter-class and the intra-class scatter as follows,

$$\begin{aligned} \max_{\mathbf{P}} \frac{Tr(\mathbf{S}_B^S)}{Tr(\mathbf{S}_W^S)} &= \max_{\mathbf{P}} \frac{Tr \left( \sum_{c=1}^C (\boldsymbol{\mu}_{S,c} - \boldsymbol{\mu})(\boldsymbol{\mu}_{S,c} - \boldsymbol{\mu})^T \right)}{Tr \left( \sum_{c=1}^C \frac{1}{N_c} \sum_{\mathbf{y} \in \text{Class } c} (\mathbf{y} - \boldsymbol{\mu}_{S,c})(\mathbf{y} - \boldsymbol{\mu}_{S,c})^T \right)} \\ &= \max_{\mathbf{P}} \frac{Tr \left( \mathbf{P}^T \sum_{c=1}^C (\mathbf{u}_{S,c} - \mathbf{u})(\mathbf{u}_{S,c} - \mathbf{u})^T \mathbf{P} \right)}{Tr \left( \mathbf{P}^T \sum_{c=1}^C \frac{1}{N_c} \sum_{\mathbf{x} \in \text{Class } c} (\mathbf{x} - \mathbf{u}_{S,c})(\mathbf{x} - \mathbf{u}_{S,c})^T \mathbf{P} \right)} \end{aligned} \quad (8)$$

where  $\boldsymbol{\mu}_{S,c} = \mathbf{P}^T \mathbf{u}_{S,c}$  and  $\boldsymbol{\mu} = \mathbf{P}^T \mathbf{u}$ .

Therefore, the discriminative term (8) can be further simplified as

$$\max_{\mathbf{P}} \frac{Tr(\mathbf{P}^T \mathbf{S}_B^S \mathbf{P})}{Tr(\mathbf{P}^T \mathbf{S}_W^S \mathbf{P})} \quad (9)$$

where  $\mathbf{S}_B^S$  and  $\mathbf{S}_W^S$  can be computed as follows

$$\mathbf{S}_B^S = \sum_{c=1}^C (\mathbf{u}_{S,c} - \mathbf{u})(\mathbf{u}_{S,c} - \mathbf{u})^T \quad (10)$$

**Algorithm 1.** The proposed CDSL

**Input:** Source data  $\mathbf{X}_S \in \mathbb{R}^{D \times N_S}$ , target data  $\mathbf{X}_T \in \mathbb{R}^{D \times N_T}$ ,  $\lambda_0, \lambda_1$ ,  $d$ , and the mini-patch size  $M$ ;

**Procedure:**

1. Compute the center  $\mathbf{u}_S^i$  of the  $i$ -th ( $i=1, \dots, M$ ) mini-patch via (3);
2. Compute the center  $\mathbf{u}_T^i$  of the  $i$ -th ( $i=1, \dots, M$ ) mini-patch via (4);
3. Compute the inter-class scatter matrix  $\mathbf{S}_B^S$  via (10);
4. Compute the intra-class scatter matrix  $\mathbf{S}_W^S$  via (11);
5. Compute the weight  $w_i$  of the  $i$ -th mini-patch ( $i=1, \dots, M$ ) via (2);
6. Compute the matrix  $\mathbf{R}$  and  $\mathbf{Q}$  via (15);
7. Solve the eigenvalue decomposition problem (21);
8. Compute the optimal subspace  $\mathbf{P}^* = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_d]$  via (22);

**Output:** The basis transformation  $\mathbf{P}$  (*i.e.*, subspace projection).

$$\mathbf{S}_W^S = \sum_{c=1}^C \sum_{\mathbf{x} \in \text{Class } c} (\mathbf{x} - \mathbf{u}_{S,c}) (\mathbf{x} - \mathbf{u}_{S,c})^T \quad (11)$$

where  $\mathbf{u}$  represents the center of source data and  $\mathbf{u}_{S,c}$  represents the center of class  $c$  of source data in the raw space.

Further, to guarantee that the projection  $\mathbf{P}$  does not distort the data of target domain, much available information should be preserved in structure under the cross-domain subspace representation. Therefore, for target data, it is rational to maximize the following term,

$$\max_{\mathbf{P}} \text{Tr} \left( \left( \mathbf{P}^T \mathbf{X}_T \right) \left( \mathbf{P}^T \mathbf{X}_T \right)^T \right) = \max_{\mathbf{P}} \text{Tr} \left( \mathbf{P}^T \mathbf{X}_T \mathbf{X}_T^T \mathbf{P} \right) \quad (12)$$

It means that by adding Eq. (12) as a constraint in the model, the variance (energy) of the target data in the new subspace can also be maximized. Therefore, much available information in target domain can be preserved without distorting the data.

For learning a robust cross-domain subspace, three units such as Eq.(7), Eq.(9) and Eq.(12) have been formulated in the proposed CDSL model. In summary, the model has the following three characteristics:

- ✧ As formulated in Eq. (7), the mini-patch feature specific domain distance is minimized, such that the proposed model can effectively handle the feature-specific distribution discrepancy. From the viewpoint of electronic nose, the drift is sensor-specific. Therefore, it can be treated individually or in mini-patch as shown in our model.
- ✧ As formulated in Eq. (9), the discriminative property (*i.e.*, separability) can be well described, such that in the learned subspace, different classes of the source domain can be easily classified. The objective of cross-domain subspace learning is to improve the probability distribution consistency (*i.e.*, domain adaptation), such that the final classifier can be adapted to both domains. Note that the ultimate goal of the proposed model is for classification.
- ✧ As formulated in the constraint of Eq. (12), which is motivated by principal component analysis (PCA), it is easy to observe that the structure of the target domain data can be well preserved, such that the target data would not be distorted after subspace projection.

After a detailed description of the three units in the proposed CDSL model, by incorporating the Eq. (7), Eq. (9) and Eq. (12) together, a complete CDSL model is formulated as follows

$$\max_{\mathbf{P}} \frac{\text{Tr}(\mathbf{P}^T \mathbf{S}_B^S \mathbf{P}) + \lambda_0 \cdot \text{Tr}(\mathbf{P}^T \mathbf{X}_T \mathbf{X}_T^T \mathbf{P})}{\text{Tr}(\mathbf{P}^T \mathbf{S}_W^S \mathbf{P}) + \lambda_1 \cdot \sum_{i=1}^M w_i \left\| \frac{1}{N_S} \sum_{j=1}^{N_S} \mathbf{P}^T \mathbf{x}_S^{ij} - \frac{1}{N_T} \sum_{k=1}^{N_T} \mathbf{P}^T \mathbf{x}_T^{ik} \right\|_2^2} \quad (13)$$

where  $\lambda_0$  and  $\lambda_1$  represent the regularization (trade-off) coefficients. From the CDSL model (13), we observe that it can be further simplified in formulation for easier solving.

By substituting (3) and (4) into (13), the proposed CDSL model (13) is reformulated as

$$\begin{aligned} & \max_{\mathbf{P}} \frac{\text{Tr}(\mathbf{P}^T \mathbf{S}_B^S \mathbf{P}) + \lambda_0 \cdot \text{Tr}(\mathbf{P}^T \mathbf{X}_T \mathbf{X}_T^T \mathbf{P})}{\text{Tr}(\mathbf{P}^T \mathbf{S}_W^S \mathbf{P}) + \lambda_1 \cdot \sum_{i=1}^M w_i \left\| \mathbf{P}^T \mathbf{u}_S^i - \mathbf{P}^T \mathbf{u}_T^i \right\|_2^2} \\ &= \max_{\mathbf{P}} \frac{\text{Tr}(\mathbf{P}^T (\mathbf{S}_B^S + \lambda_0 \cdot \mathbf{X}_T \mathbf{X}_T^T) \mathbf{P})}{\text{Tr}(\mathbf{P}^T \mathbf{S}_W^S \mathbf{P}) + \lambda_1 \cdot \sum_{i=1}^M w_i \text{Tr}(\mathbf{P}^T (\mathbf{u}_S^i - \mathbf{u}_T^i) (\mathbf{u}_S^i - \mathbf{u}_T^i)^T \mathbf{P})} \\ &= \max_{\mathbf{P}} \frac{\text{Tr}(\mathbf{P}^T (\mathbf{S}_B^S + \lambda_0 \cdot \mathbf{X}_T \mathbf{X}_T^T) \mathbf{P})}{\text{Tr}(\mathbf{P}^T \mathbf{S}_W^S \mathbf{P}) + \lambda_1 \cdot \sum_{i=1}^M w_i \text{Tr}(\mathbf{P}^T (\mathbf{u}_S^i - \mathbf{u}_T^i) (\mathbf{u}_S^i - \mathbf{u}_T^i)^T \mathbf{P})} \\ &= \max_{\mathbf{P}} \frac{\text{Tr}(\mathbf{P}^T (\mathbf{S}_B^S + \lambda_0 \cdot \mathbf{X}_T \mathbf{X}_T^T) \mathbf{P})}{\text{Tr}(\mathbf{P}^T (\mathbf{S}_W^S + \lambda_1 \cdot \sum_{i=1}^M w_i (\mathbf{u}_S^i - \mathbf{u}_T^i) (\mathbf{u}_S^i - \mathbf{u}_T^i)^T) \mathbf{P})} \end{aligned} \quad (14)$$

As can be seen from (14), the CDSL model is solving a ratio maximization problem, which is non-convex. Additionally, there will be a group of solutions  $\mathbf{P}$ . Therefore, in optimization, we have imposed an equality constraint, such that the solving is transformed into an efficient eigenvalue decomposition problem. The solver is described in the following section.

### C. Model Optimization

The optimal solution to (14) is equivalent to solving the Eigenvalue problem shown in the following Theorem 1.

*Theorem:* An optimal solution to (14) is given by chosen  $\mathbf{P}$  as the matrix whose columns are the eigenvectors  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_d$ , corresponding to the first  $d$  largest eigenvalues  $\rho_1, \rho_2, \dots, \rho_d$  of the following generalized eigenvalue decomposition problem:

$$\mathbf{Q}^{-1} \mathbf{R} \mathbf{v} = \rho \mathbf{v}$$

where  $\rho$  denotes the eigenvalues,  $\mathbf{Q}$  and  $\mathbf{R}$  are represented as follows

$$\begin{cases} \mathbf{R} = \mathbf{S}_B^S + \lambda_0 \cdot \mathbf{X}_T \mathbf{X}_T^T \\ \mathbf{Q} = \mathbf{S}_W^S + \lambda_1 \cdot \sum_{i=1}^M w_i (\mathbf{u}_S^i - \mathbf{u}_T^i) (\mathbf{u}_S^i - \mathbf{u}_T^i)^T \end{cases} \quad (15)$$

The proof of Theorem is shown in Appendix A. For easy following the proposed CDSL model in implementation, the algorithm is summarized in **Algorithm 1**.

### D. Classification

The proposed CDSL is used to learn a cross-domain discriminative subspace  $\mathbf{P}$  for domain adaptation. After projecting the data from source and target domains, the classifier is learned on the projected source data  $\mathbf{X}_S' = \mathbf{P}^T \mathbf{X}_S$  (master system), and the final task is for accurate classification

**Algorithm 2.** Classification

**Input:** Source data  $\mathbf{X}_S \in \mathbb{R}^{D \times N_S}$ , target data  $\mathbf{X}_T \in \mathbb{R}^{D \times N_T}$ , source label  $\mathbf{t}_S$ , and the subspace projection  $\mathbf{P}$ .

**Procedure:**

1. Compute the projected source data  $\mathbf{X}'_S = \mathbf{P}^T \mathbf{X}_S$ ;
2. Compute the projected target data  $\mathbf{X}'_T = \mathbf{P}^T \mathbf{X}_T$ ;
3. SVM classifier training on  $\{\mathbf{X}'_S, \mathbf{t}_S\}$  by solving (16);
4. Classification of  $\mathbf{X}'_T$  by using (17);

**Output:** the classification results of target data.

of multiple kinds of odors in target data  $\mathbf{X}'_T = \mathbf{P}^T \mathbf{X}_T$  (slave system). For classifier training, support vector machine (SVM) is used in this paper. For easier following, the briefs of SVM are simply provided as follows.

Given a training set of  $N$  data points  $\{\mathbf{x}_i, y_i\}_{i=1}^N$ , where the label  $y_i \in \{-1, 1\}$ ,  $i = 1, \dots, N$ . According to the structural risk minimization principle, SVM aims at solving the following risk bound minimization problem with inequality constraint.

$$\begin{aligned} \min_{\mathbf{w}, \xi} \quad & \frac{1}{2} \|\mathbf{w}\|^2 + C \cdot \sum_{i=1}^N \xi_i, \\ \text{s.t.} \quad & \xi_i \geq 0, y_i [\mathbf{w}^T \varphi(\mathbf{x}_i) + b] \geq 1 - \xi_i \end{aligned} \quad (16)$$

where  $\varphi(\cdot)$  is a linear/nonlinear mapping function,  $\mathbf{w}$  and  $b$  are the parameters of classifier hyper-plane,  $C$  denotes the penalty coefficient, and  $\xi$  denotes prediction error. Generally, for optimization, the raw problem (16) of SVM can be transformed into its dual formulation with equality constraint by using Lagrange multiplier method.

After solving Eq.(16), in classification, the goal of SVM is to construct the following decision function

$$f(\mathbf{x}) = \text{sgn} \left( \sum_{i=1}^M \alpha_i y_i \kappa(\mathbf{x}_i, \mathbf{x}) + b \right) \quad (17)$$

where  $\kappa(\cdot)$  is a kernel function.  $\kappa(\mathbf{x}_i, \mathbf{x}) = \varphi(\mathbf{x}_i)^T \varphi(\mathbf{x}) = \mathbf{x}_i^T \mathbf{x}$  for linear SVM and  $\kappa(\mathbf{x}_i, \mathbf{x}) = \exp(-\|\mathbf{x}_i - \mathbf{x}\|^2 / \sigma^2)$  for nonlinear Gaussian SVM. In this paper, Gaussian kernel function is used due to its generalization and better performance in experiments. The odor classification algorithm based on SVM is summarized in **Algorithm 2**.

**E. Remark**

The proposed CDSL model is a cross-domain subspace learning method, and the learned subspace includes three important aspects: discriminative property of source data, data fidelity of target data, and mini-patch feature domain distance minimization. The task of CDSL is to learn a common subspace for source and target domains and achieve domain adaptation. In modeling, the labels of source domain are supposed to be available, while the labels of target domain are unavailable or very few labels are available. Also, the mini-patch feature is designed by considering that the domain difference may be different with respect to different features. When the mini-patch size  $M=1$ , all features contribute the same to the distribution inconsistency between source and target domains. In experiments, different  $M$  values are discussed, and different number of labeled target data is also evaluated and explored.

## IV. EXPERIMENTS

**A. Experimental data**

In this section, three odor datasets, including source domain dataset (S), target domain dataset 1 ( $T_1$ ), and target domain dataset 2 ( $T_2$ ), are experimented. The three datasets are with large knowledge shift (instrumental related) and drift (time related). These datasets are collected by using three electronic nose systems with completely the same type of metal oxide semi-conductor gas sensors, including TGS2620, TGS2602, TGS2201A and TGS2201B. An extra module for temperature and humidity sensing is also used. For each observation, the steady state response point is extracted, and as a result, a 6-dimensional feature vector is formulated. Also, the source dataset was collected 5 years earlier than target dataset 1 and target dataset 2. That is, shift and drift are implied between source dataset and target dataset, while only shift is implied between target dataset 1 and target dataset 2. For each dataset, six kinds of odors (air contaminants) are included, such as toluene ( $C_7H_8$ ), benzene ( $C_6H_6$ ), ammonia ( $NH_3$ ), carbon monoxide (CO), nitrogen dioxide ( $NO_2$ ), and formaldehyde ( $CH_2O$ ). The detailed description of the three datasets is shown in Table I. The raw features of all samples acquired by using master system, slave 1 system and slave 2 system are described in Fig. 3(a), with respect to each class. As can be seen from Fig. 3(a), six kinds of features (feature dimensionality) are included in each E-nose system and the feature value (sensing value) is normalized into (0, 1). The significant feature distribution difference between source and target domains for the same class can be easily observed. Further, the mean feature for master, slave 1 and slave 2 computed based on Fig. 3(a) is shown in Fig. 3(b) with respect to each class. It is clear that the mean source feature is different from slave features for the same class. To visualize the scatter points of the source dataset (master), target dataset 1 (slave 1) and target dataset 2 (slave 2) in raw feature space, PCA is analyzed on the three datasets, respectively. The scatter points of the first two principal components are shown in Fig. 4, from which we observe that the data points from different classes are almost clustered.

**B. Mini-patch Feature Combination**

As described in the experimental data, 6 kinds of features (*i.e.*, six sensors) are formed, which is nominated as  $f_1, f_2, \dots, f_6$ , respectively. In this paper, different mini-patch size  $M$  is discussed. Specifically, the description of the mini-patch (feature combination) is shown in Table II. In experiments, different  $M$  values have been discussed separately. Note that when  $M=5$ , two cases of feature combination are considered. From Table II, it is clear that when  $M=1$ , it is a general problem. When  $M=6$ , each kind of feature is viewed to be one patch, that is, the feature dimensionality for each patch is 1.

**C. Experimental Setting**

In experiments, according to the availability of the target data labels, two cross-domain settings are experimented respectively. The performance on target data is reported.

**Cross-domain setting 1:** In CDSL training, the labels of the target domain data are unavailable. That is, the target labels are not used for model training and classifier learning and only the source data and source labels are used.

TABLE I  
DESCRIPTION OF EXPERIMENTAL DATASETS

<i>E-nose systems</i>	<i>Dimensionality</i>	<i>Toluene</i>	<i>Benzene</i>	<i>Ammonia</i>	<i>Carbon monoxide</i>	<i>Nitrogen dioxide</i>	<i>Formaldehyde</i>	<i>Total</i>
Source domain	6	66	72	60	58	38	126	420
Target domain 1	6	106	108	81	98	107	108	608
Target domain 2	6	94	87	84	95	108	108	576

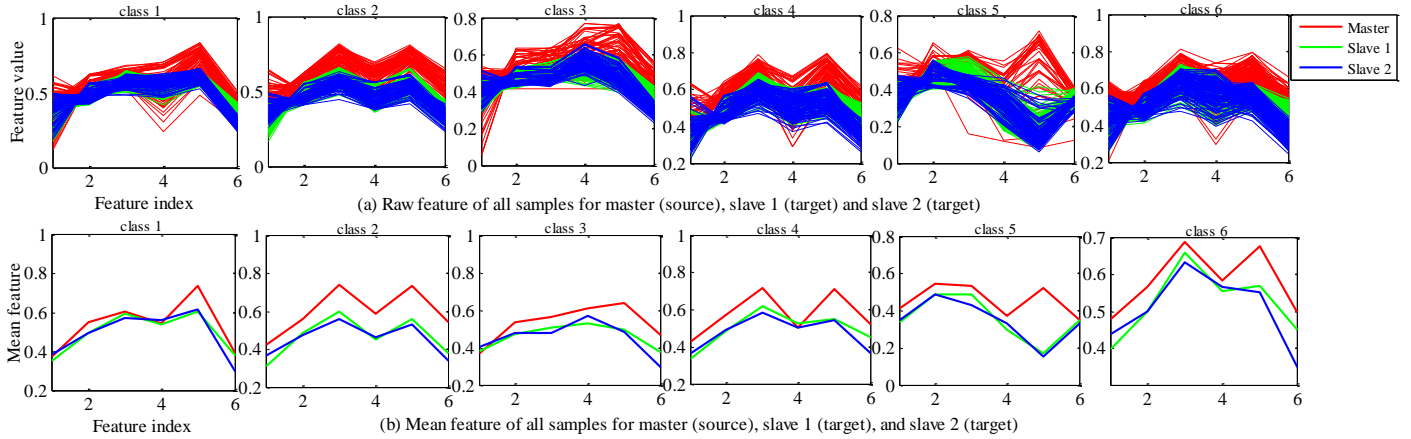


Fig. 3. Description of features of the source domain and target domain with respect to different classes

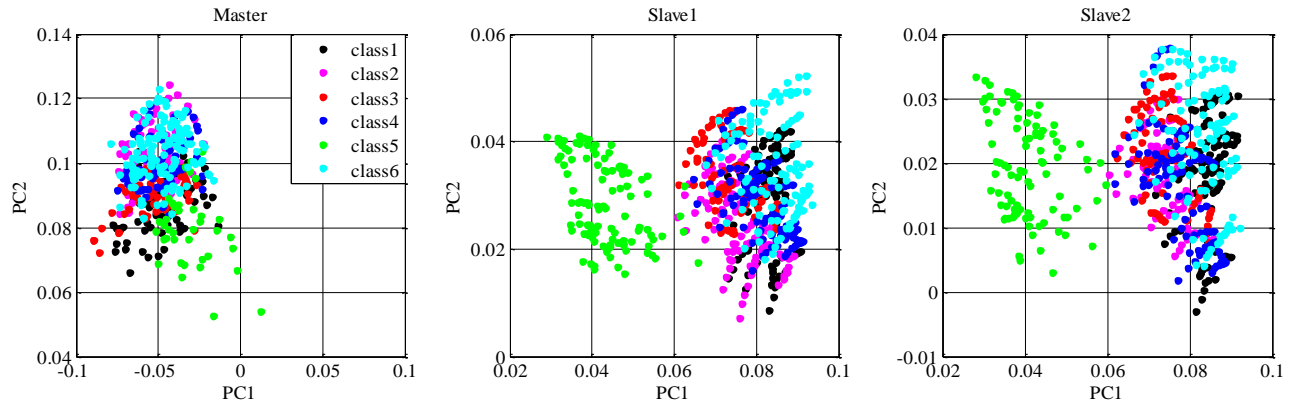


Fig. 4. The principle component analysis (PCA) results of the three datasets: source domain (*i.e.*, master), target domain 1 (*i.e.*, slave 1), and target domain 2 (*i.e.*, slave 2). Class 1-class 6 correspond the odors  $C_7H_8$ ,  $C_6H_6$ ,  $NH_3$ ,  $CO$ ,  $NO_2$ , and  $CH_2O$ .

TABLE II  
DESCRIPTION OF FEATURE COMBINATIONS WITH DIFFERENT MINI-PATCH SIZE  $M$

$M$	<i>Feature Patches</i>					
1	<i>Patch 1</i>					
	<b>Feature</b> =[ $f_1, f_2, f_3, f_4, f_5, f_6$ ]					
2	<i>Patch 1</i>			<i>Patch 2</i>		
	<b>Feature 1</b> =[ $f_1, f_2$ ]			<b>Feature 2</b> =[ $f_3, f_4, f_5, f_6$ ]		
3	<i>Patch 1</i>		<i>Patch 2</i>		<i>Patch 3</i>	
	<b>Feature 1</b> =[ $f_1, f_2$ ]		<b>Feature 2</b> =[ $f_3, f_4$ ]		<b>Feature 3</b> =[ $f_5, f_6$ ]	
	<i>Patch 1</i>		<i>Patch 2</i>		<i>Patch 3</i>	
4	<i>Patch 1</i>		<i>Patch 2</i>		<i>Patch 3</i>	
	<b>Feature 1</b> =[ $f_1, f_2$ ]		<b>Feature 2</b> =[ $f_3$ ]		<b>Feature 3</b> =[ $f_4$ ]	
	<i>Patch 1</i>		<i>Patch 2</i>		<i>Patch 3</i>	
5	<i>Patch 1</i>		<i>Patch 2</i>		<i>Patch 3</i>	
	Case 1	<b>Feature 1</b> =[ $f_1, f_2$ ]	<b>Feature 2</b> =[ $f_3$ ]	<b>Feature 3</b> =[ $f_4$ ]	<b>Feature 4</b> =[ $f_5$ ]	<b>Feature 5</b> =[ $f_6$ ]
	Case 2	<b>Feature 1</b> =[ $f_1$ ]	<b>Feature 2</b> =[ $f_2$ ]	<b>Feature 3</b> =[ $f_3$ ]	<b>Feature 4</b> =[ $f_4$ ]	<b>Feature 5</b> =[ $f_5, f_6$ ]
6	<i>Patch 1</i>		<i>Patch 2</i>		<i>Patch 3</i>	
	<b>Feature 1</b> =[ $f_1$ ]		<b>Feature 2</b> =[ $f_2$ ]		<b>Feature 3</b> =[ $f_3$ ]	
	<i>Patch 4</i>		<i>Patch 5</i>		<i>Patch 6</i>	
	<b>Feature 4</b> =[ $f_4$ ]		<b>Feature 5</b> =[ $f_5$ ]		<b>Feature 6</b> =[ $f_6$ ]	

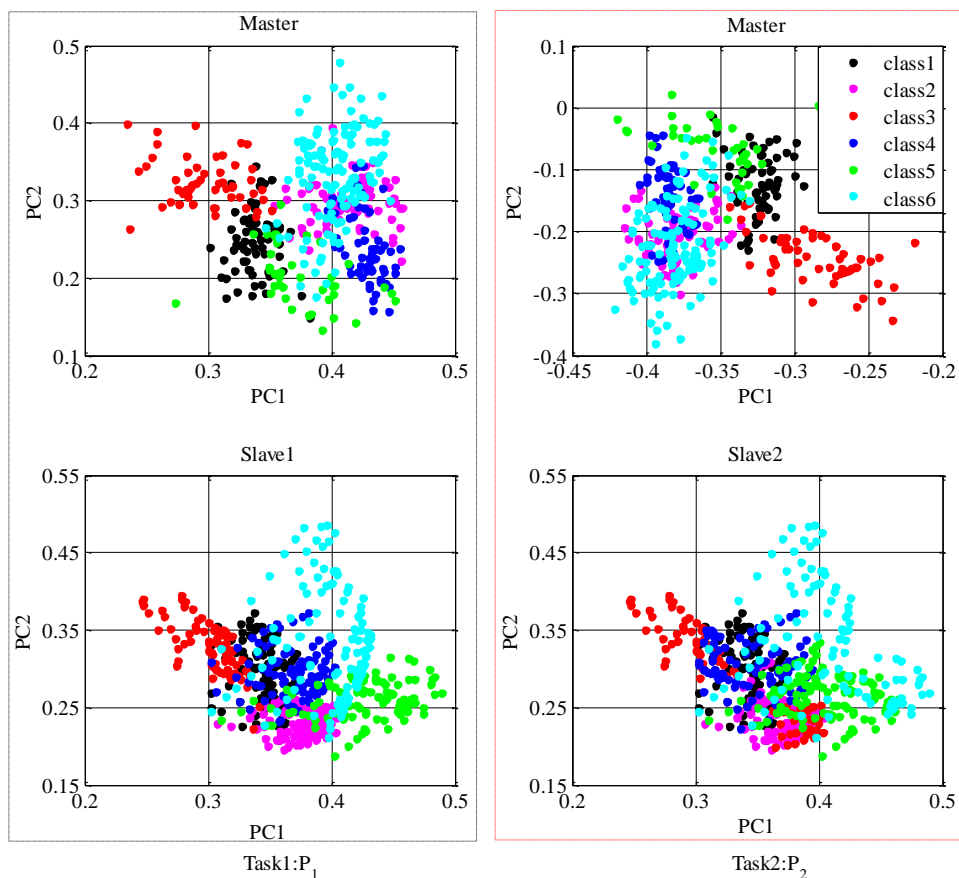


Fig. 5. The scatter points of the first two components by using the proposed CDSL method under **Setting 1**. The left denotes the task 1 (i.e.,  $P_1$  of source domain  $\rightarrow$  target domain 1) and the right denotes the task 2 (i.e.,  $P_2$  of source domain  $\rightarrow$  target domain 2). Class 1~class 6 correspond the odors  $C_7H_8$ ,  $C_6H_6$ ,  $NH_3$ ,  $CO$ ,  $NO_2$ , and  $CH_2O$ , respectively.

TABLE III  
COMPARISONS OF RECOGNITION ACCURACY ON TWO TASKS (SETTING 1)

Cross-domain task	SVM	PCA	LDA	LPP	NPE	NCA	MDS	LFDA	GFK	SGF	SA	OSC	DS	GLSW	CDSL
Source domain $\rightarrow$ target domain 1	51.97	51.97	51.97	53.95	53.62	41.28	51.15	61.84	33.88	55.10	41.10	34.38	45.00	40.46	<b>71.88</b>
Source domain $\rightarrow$ target domain 2	60.59	60.59	56.77	57.81	54.69	33.85	58.51	61.63	32.81	57.49	31.12	36.46	42.62	53.65	<b>71.88</b>

TABLE IV  
RECOGNITION ACCURACY WITH DIFFERENT M-VALUE IN THE PROPOSED CDSL

Cross-domain recognition task	CDSL ( $M=1$ )	CDSL ( $M=2$ )	CDSL ( $M=3$ )	CDSL ( $M=4$ )	CDSL ( $M=5$ , case 1)	CDSL ( $M=5$ , case 2)	CDSL ( $M=6$ )
Source domain $\rightarrow$ target domain 1	<b>71.88</b>	68.59	60.03	62.01	60.53	62.83	59.87
Source domain $\rightarrow$ target domain 2	<b>71.88</b>	69.44	62.50	65.45	62.50	63.19	62.50

**Cross-domain setting 2:** In classifier training, partial labels of target domain can be used. Specifically, for each class in the target domain,  $k$  labeled samples are used for classifier learning, and  $k=1, 3, 5, 7, 9$  is discussed separately. The compared methods follow the same setup.

**Training and Testing Protocol:** In experiments, the source domain data is fixed as training set for model training (CDSL) and classifier learning (SVM). A few target domain data is used as a validation set for the best regularization parameters tuning. The remaining target domain data is used for testing. Note that there is no overlap among training, validation and testing sets.

#### D. Compared Methods

To show the effectiveness of the proposed method, we have chosen 14 machine learning based methods of four categories. First, 3 baseline methods such as support vector machine (SVM), principal component analysis (PCA) and linear discriminant analysis (LDA) are compared. Second, 3 representative calibration transfer methods in E-nose, such as orthogonal signal correction (OSC) [36], generalized least squares weighting (GLSW) [37] and direct standardization (DS) [38] are compared. Third, 5 semi-supervised learning methods



TABLE V

COMPARISONS OF RECOGNITION ACCURACY ON SETTING 2 (TASK 1: SOURCE DOMAIN  $\rightarrow$  TARGET DOMAIN 1) WITH DIFFERENT NO. OF LABELED TARGET DATA PER CLASS

No. of labeled target data per class ( $k$ )	1	3	5	7	9	Average
SVM	59.14	63.22	62.80	70.49	70.76	65.28
PCA	59.14	63.39	65.40	70.49	70.76	65.84
LDA	67.77	71.36	74.22	75.09	76.17	72.92
LPP	65.46	69.83	71.45	72.08	71.48	70.06
NPE	64.78	64.07	63.49	71.55	71.84	67.15
NCA	52.49	50.85	53.81	50.00	63.00	54.03
MDS	61.13	64.75	65.57	70.32	72.92	66.94
LFDA	62.13	67.12	71.63	76.86	74.91	70.53
GFK	33.39	34.07	35.81	37.63	38.45	35.87
SGF	66.61	66.95	67.13	70.14	72.74	68.71
SA	49.67	58.64	59.69	58.83	61.91	57.75
OSC	41.03	45.25	45.85	43.29	45.13	44.11
DS	54.69	62.56	59.78	59.09	59.49	59.12
GLSW	52.99	51.69	69.90	60.42	59.75	58.95
CDSL ( $M=1$ )	<b>72.76</b>	<b>77.97</b>	<b>79.07</b>	<b>81.80</b>	<b>83.21</b>	<b>78.96</b>
CDSL ( $M=2$ )	71.10	76.61	78.20	79.51	81.05	77.29
CDSL ( $M=3$ )	69.27	75.08	76.99	77.39	79.06	75.56
CDSL ( $M=4$ )	69.10	75.25	77.34	77.92	78.88	75.70
CDSL ( $M=5$ , case 1)	70.27	75.25	77.16	78.80	79.60	76.22
CDSL ( $M=5$ , case 2)	69.10	75.25	77.16	78.09	78.88	75.70
CDSL ( $M=6$ )	68.94	75.25	77.16	77.74	79.24	75.67

TABLE VI

COMPARISONS OF RECOGNITION ACCURACY ON SETTING 2 (TASK 2: SOURCE DOMAIN  $\rightarrow$  TARGET DOMAIN 2) WITH DIFFERENT NO. OF LABELED TARGET DATA PER CLASS

No. of labeled target data per class ( $k$ )	1	3	5	7	9	Average
SVM	69.65	72.76	73.63	74.16	74.90	73.02
PCA	69.65	72.76	73.63	74.16	74.90	73.02
LDA	71.40	75.45	75.82	79.03	78.74	76.09
LPP	69.82	74.19	73.63	72.85	76.82	73.46
NPE	71.05	71.15	71.43	72.28	74.14	72.01
NCA	56.32	47.49	52.01	55.62	58.03	53.90
MDS	72.11	73.48	73.81	75.28	75.29	74.00
LFDA	65.26	70.61	73.08	75.47	77.01	72.29
GFK	35.26	39.78	43.22	44.57	46.17	41.80
SGF	68.07	71.51	73.08	73.03	73.56	71.85
SA	68.95	71.86	72.34	73.22	73.37	71.95
OSC	33.68	41.04	33.70	41.57	31.42	36.28
DS	48.59	52.97	52	51.30	48.52	50.68
GLSW	66.49	60.04	63.92	75.09	76.82	68.47
CDSL ( $M=1$ )	<b>78.25</b>	78.67	<b>80.04</b>	81.46	81.99	80.08
CDSL ( $M=2$ )	77.37	78.85	79.67	<b>81.84</b>	<b>83.14</b>	<b>80.17</b>
CDSL ( $M=3$ )	75.96	78.49	78.57	80.90	81.23	79.03
CDSL ( $M=4$ )	75.96	78.85	78.57	80.90	81.23	79.10
CDSL ( $M=5$ , case 1)	75.96	<b>79.21</b>	79.12	80.90	81.23	79.28
CDSL ( $M=5$ , case 2)	76.32	78.85	78.75	80.90	81.23	79.21
CDSL ( $M=6$ )	75.96	78.49	78.75	80.90	81.23	79.07

based on manifold learning, including locality preservation projection (LPP) [34], multidimensional scaling (MDS) [39], neighborhood component analysis (NCA) [40], neighborhood preserving embedding (NPE) [41], and local fisher discriminant analysis (LFDA) [42] are explored and compared. Finally, 3 popular subspace transfer learning methods such as geodesic flow kernel (GFK) [43], sampling geodesic flow (SGF) [44] and subspace alignment (SA) [45] are also explored and compared. These methods are closely related with CDSL.

### E. Results

In this section, the experimental results on cross-domain

settings are reported to evaluate the proposed CDSL method.

#### ✧ Setting 1:

Under the cross-domain setting 1, we first observe the qualitative result shown in Fig. 5 by plotting the scatter points of the first two components after subspace projection. We can see that after cross-domain subspace projection (*i.e.*,  $\mathbf{P}_1$ ) between source domain and target domain 1, and  $\mathbf{P}_2$  between source domain and target domain 2, the separability among data points from different classes (represented as different symbols) is much improved in the learned common subspace compared

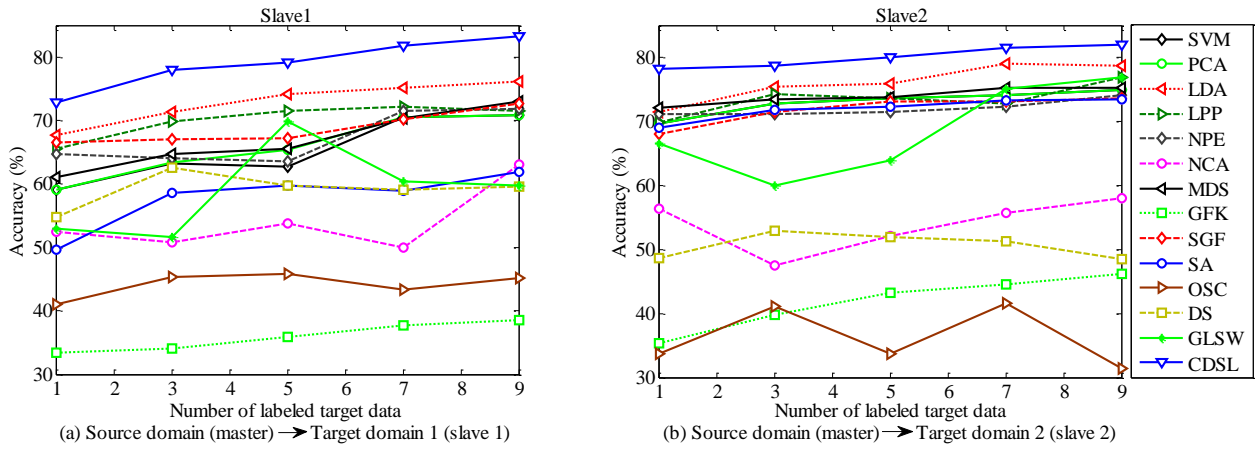


Fig. 6. Odor recognition accuracy with respect to different number of labeled target data per class for two cross-domain tasks: (a) Source domain  $\rightarrow$  target domain 1 and (b) source domain  $\rightarrow$  target domain 2.

to Fig. 4 that is without cross-domain learning.

Further, the odor classification accuracy of the target domain data has been reported in Table III. We can clearly observe that the proposed CDSL method significantly outperforms other methods. The recognition accuracy for two cross-domain tasks achieves 71.88%, which is almost 10% improvement in comparisons. Note that the result of CDSL is obtained when the mini-patch size  $M=1$ . The results with different  $M$ -values are shown in Table IV, from which we observe that  $M=1$  shows better performance. The comparison results clearly demonstrate the effectiveness of the proposed CDSL method in cross-domain classification tasks. Note that, for cross-domain classification task, the classifier is trained on one domain, but tested on another domain. Traditional classification implies that the distribution between training data and testing data is the same or similar. However, due to the domain distribution discrepancy, this assumption is violated, such that cross-domain learning and classification (*i.e.* the proposed CDSL) is desired. In this paper, the odor datasets acquired by using different electronic noses (master and slaves) at different time interval are recognized as different domains.

#### ◇ Setting 2:

Under the cross-domain setting 2, two tasks, *i.e.*, source domain (master)  $\rightarrow$  target domain 1 (slave 1) and the source domain (master)  $\rightarrow$  target domain 2 (slave 2), are completed by using the proposed method respectively. Different from Setting 1 that only the source data are used for classifier training, in Setting 2,  $k$  labeled target data are also used as auxiliary data of source data for classifier training. Considering that the number of labeled target data is very limited,  $k=1, 3, 5, 7, 9$  samples per class (odor) are randomly selected from target domain for classifier learning. The recognition accuracy of the first task (*i.e.*, source domain  $\rightarrow$  target domain 1) is reported in Table V, from which we can clearly observe that the proposed CDSL performs the best cross-domain recognition with different  $M$ -value. Particularly, when  $M=1$ , the best accuracy for each  $k$  is achieved, which is about 5% improvement comparing to LDA based recognition result.

Similarly, the recognition accuracy of the second cross-domain task (*i.e.*, source domain  $\rightarrow$  target domain 2) is

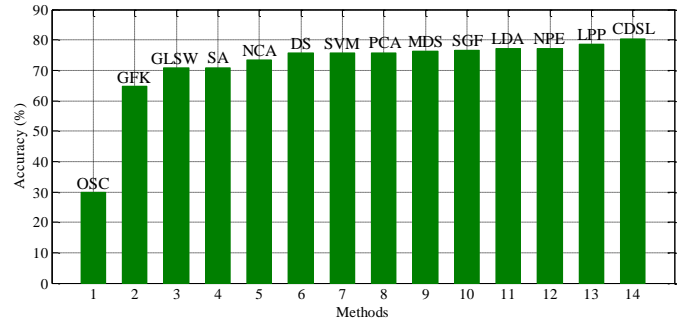


Fig. 7. Recognition accuracies in single E-nose system by using 14 methods.

reported in Table VI. From the results, we also observe the better performance of the proposed CDSL method by comparing with other methods. These results demonstrate that the proposed method is very effective in cross-domain classification (*i.e.*, odor recognition in multiple E-noses).

The performance curves with respect to different number of labeled target data for two cross-domain classification (transferring odor recognition) tasks are shown in Fig. 6. From the accuracy curves, we can observe that the proposed CDSL method shows the best recognition results. Additionally, the accuracy becomes higher with the increasing of labeled target samples containing drift information, that are used as auxiliary data for training a more robust CDSL model. Therefore, the recognition performance and generalization ability become better by adding more target data into the source training data. However, in real applications it is difficult to obtain much labeled target data points, we have therefore discussed very few labeled target data as shown in Fig. 6.

The performance of the proposed method has been validated in odor recognition across multiple E-nose systems. To show the performance of CDSL in odor recognition based on a single E-nose system, we have conducted further exploration. We select 50% samples from the master system to train a model using CDSL, and the remaining 50% samples are used for testing the model. The recognition accuracies with comparisons are described in Fig. 7, from which we can observe that the proposed CDSL still outperforms others. This experiment shows the double reliability of CDSL method in not only multiple E-noses, but single E-nose system.

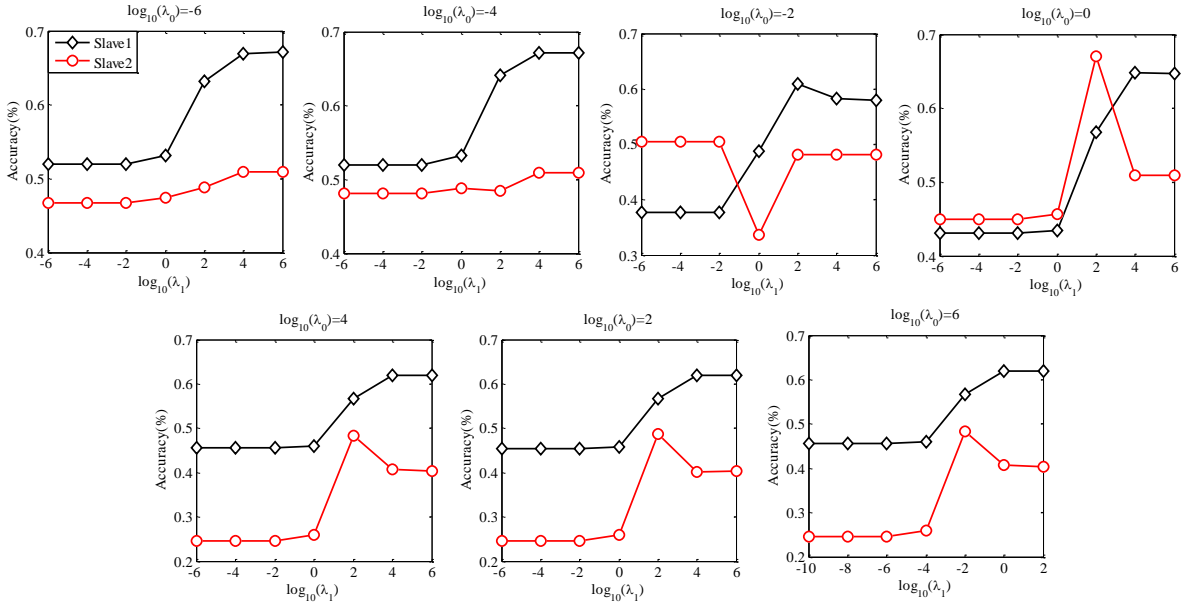


Fig. 8. Performance variation curves with respect to the logarithm of  $\lambda_1$  by frozen  $\log_{10}(\lambda_0)=-6, -4, -2, 0, 2, 4,$  and  $6,$  respectively.

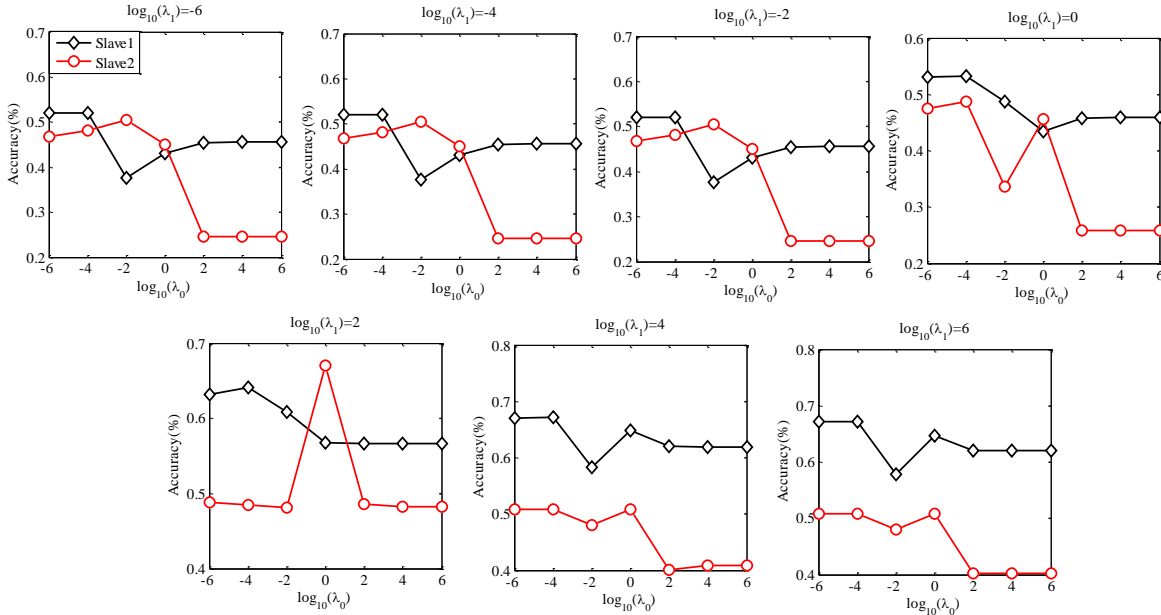


Fig. 9. Performance variation curves with respect to the logarithm of  $\lambda_0$  by frozen  $\log_{10}(\lambda_1)=-6, -4, -2, 0, 2, 4,$  and  $6,$  respectively.

### F. Discussion

In this paper, a mini-patch concept is used in modeling. From the results shown in Table IV, V and VI, we observe that a smaller value of  $M$  performs better results. We notice that when  $M=1$ , a single mini-patch with all sensors in multi-dimensional data takes place. This can better show the feature distribution discrepancy between systems. Unlike that  $M=6$ , each mini-patch only contains a single sensor, such that the distribution cannot be well learned and aligned based on a subspace projection. The cross-sensitivity among sensors may be lost and harmful to odor recognition.

Additionally, the proposed CDSL model is a cross-domain learning framework, which can be used to handle heterogeneous data classification problems. In E-nose, the well-trained classifier based on master device may not be adapted to the slave devices due to some inherent differences.

Also, in image classification, when the image data is from different domains (*e.g.*, different sensors of low and high resolution), the classifier trained with high resolution images may not be adapted to the low resolution images. Therefore, the proposed cross-domain learning method can be expected to solve such kind of generalized heterogeneous problem.

### G. Parameter Sensitivity Analysis

In the proposed CDSL model, there are two parameters: the regularization coefficients  $\lambda_0$  and  $\lambda_1$ . To observe the performance variations in tuning the two parameters,  $\lambda_0$  and  $\lambda_1$  are fine-tuned increasingly according to  $10^t$ , where  $t = \{-6, -4, -2, 0, 2, 4, 6\}$ . To show the performance with respect to each coefficient, one is tuned by frozen the other one. Specifically, the performance variation with respect to  $\lambda_1$  by frozen  $\lambda_0$  is shown in Fig. 8, from which we can see that a larger  $\lambda_1$  contributes positive effect on recognition of target domains

(i.e., slave 1 and slave 2). Similarly, the performance variation with respect to  $\lambda_0$  by frozen  $\lambda_1$  is shown in Fig. 9. We can see that a smaller  $\lambda_0$  contributes positive effect.

Further, from the proposed CDSL formulation (14), we can see that  $\lambda_1$  is the coefficient of the mini-patch feature based domain distance term. Since a larger  $\lambda_1$  shows better recognition performance, it demonstrates that the proposed mini-patch feature based domain distance term is very helpful for cross-domain learning. Additionally,  $\lambda_0$  is the coefficient of the target data fidelity term, which is used to reduce data distortion and preserve more useful information in the learned subspace. Since a smaller  $\lambda_0$  shows better performance, it demonstrates that the proposed cross-domain subspace is useful and effective without distorting the target data too much. Even that without the data fidelity term the proposed CDSL may still be effective.

## V. CONCLUSION

In this paper, we propose a cross-domain discriminative subspace learning (CDSL) model for handling transferring odor recognition tasks in multiple E-nose systems, and approximate actual application scenarios. The proposed transferring odor recognition concept denotes that the recognition model is learned on one odor dataset from master system  $A$  and tested on another odor dataset from slave system  $B$  of the same type as system  $A$  or another odor dataset still from system  $A$  but acquired at different time.

Three novel aspects are included in the proposed method. First, a model is formulated by minimizing the intra-class compactness and simultaneously maximizing the inter-class separability based on source domain. Second, the data fidelity term is imposed as constraint in CDSL based on target domain for avoiding distortion. Third, a mini-patch feature-specific domain distance is proposed, such that we can give different penalty coefficients in terms of drift degree (e.g., the further the distance is, the more serious the drift is). In this way, those mini-patches with more serious drift will be paid more attention to. Essentially, the proposed method in this paper is to address the issue of sensor drift in E-nose. Formally, we regard the drift issue as a cross-domain recognition problem that is rational and novel, such that cross-domain learning techniques can be developed to solve the transferring odor recognition problems. Different from existing subspace learning methods such as PCA and LDA that can only work on single domain, the proposed cross-domain method aims at learning a common subspace for connecting source and target domains.

Experiments on olfaction datasets of multiple kinds of odors by using three E-nose systems (i.e., one master and two slaves) demonstrate the effectiveness of the proposed method and the superiority in comparisons.

## APPENDIX A

### PROOF OF THEOREM

*Proof:* We can re-write (14) with an equality constraint as follows

$$\begin{aligned} & \max_{\mathbf{P}} \text{Tr}(\mathbf{P}^T (\mathbf{S}_B^S + \lambda_0 \cdot \mathbf{X}_T \mathbf{X}_T^T) \mathbf{P}) \\ & \text{s.t. } \text{Tr}(\mathbf{P}^T (\mathbf{S}_W^S + \lambda_1 \cdot \sum_{i=1}^M w_i (\mathbf{u}_S^i - \mathbf{u}_T^i)(\mathbf{u}_S^i - \mathbf{u}_T^i)^T) \mathbf{P}) = \varpi \\ & \quad \downarrow \\ & \max_{\mathbf{P}} \text{Tr}(\mathbf{P}^T \mathbf{R} \mathbf{P}) \\ & \text{s.t. } \text{Tr}(\mathbf{P}^T \mathbf{Q} \mathbf{P}) = \varpi \end{aligned} \quad (18)$$

where  $\mathbf{R}$  and  $\mathbf{Q}$  are represented in Eq. (14), and  $\varpi$  denotes a positive constant for equality constraint such that the solution can be normalized for uniqueness.

By introducing Lagrange multiplier  $\rho$ , the objective function of model (15) can be formulated as

$$J(\mathbf{P}, \rho) = \text{Tr}(\mathbf{P}^T \mathbf{R} \mathbf{P}) - \rho (\text{Tr}(\mathbf{P}^T \mathbf{Q} \mathbf{P}) - \varpi) \quad (19)$$

By computing the partial derivative of  $J(\mathbf{P}, \rho)$  with respect to  $\mathbf{P}$ , and let it be zero, there is

$$\begin{aligned} \frac{\partial J(\mathbf{P}, \rho)}{\partial \mathbf{P}} &= \mathbf{R} \mathbf{P} - \rho \cdot \mathbf{Q} \mathbf{P} = 0 \\ & \quad \downarrow \\ \mathbf{Q}^{-1} \mathbf{R} \mathbf{P} &= \rho \cdot \mathbf{P} \end{aligned} \quad (20)$$

where the optimal  $\mathbf{P}$  is the matrix whose columns are the eigenvectors  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_d$ , corresponding to the first  $d$  largest eigenvalues  $\rho_1, \rho_2, \dots, \rho_d$  of the following generalized eigenvalue decomposition problem,

$$\mathbf{Q}^{-1} \mathbf{R} \mathbf{v} = \rho \mathbf{v} \quad (21)$$

where the optimal subspace projection  $\mathbf{P}$  can be represented as

$$\mathbf{P}^* = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_d] \quad (22)$$

where  $d$  denotes the expected dimensionality of the new subspace. *Proof of the Theorem* is completed.

## APPENDIX B

### DATA PREPROCESSING

Smooth filtering and the vector standardization are used for preprocessing and normalization, respectively. In this paper, the filtered signal vector  $\mathbf{x}$  of each sensor can be calculated by

$$x(i) = \frac{\sum_{l=i}^{i+L-1} q_l - \max(q_i, \dots, q_{i+L-1}) - \min(q_i, \dots, q_{i+L-1})}{L-2} \quad (21)$$

where  $i=1, \dots, n-L+1$ ,  $\mathbf{q} = [q_1, q_2, \dots, q_n]^T$  is a response sequence for one sensor, and the length of raw signal is  $n$ . The filtered signal sequence of each sensor is indicated as  $\mathbf{x} = [x(i), \dots, x(n-L+1)]^T$ . The width  $L$  of the smoothing filter window is set as 20 in this work. By subtracting the maximum and minimum values, the noise may be removed, which is similar to averaging filters functioned essentially as a low-pass filter. For normalization, the steady-state point in each sensor sequence after filtering is selected as the feature of each observation, and divided by the *maximum* value of all observations.

## ACKNOWLEDGMENT

The authors would like to thank the Editor in Chief, Associate Editor and anonymous reviewers for their insightful comments.

## REFERENCES

- [1] F. Röck, N. Barsan, and U. Weimar, "Electronic Nose: Current Status and Future Trends," *Chem. Rev.*, vol. 108, pp. 705-725, 2008.
- [2] L. Zhang and F. Tian, "Performance Study of Multilayer Perceptrons in a Low-Cost Electronic Nose," *IEEE Trans. Instrumentation and Measurement*, vol. 63, no. 7, pp. 1670-1679, 2014.
- [3] N. Bhattacharyya, R. Bandyopadhyay, M. Bhuyan, B. Tudu, D. Ghosh, and A. Jana, "Electronic Nose for Black Tea Classification and Correlation of Measurements With "Tea Taster" Marks," *IEEE Trans. Instrumentation and Measurement*, vol. 57, no. 7, pp. 1313-1321, 2008.
- [4] A. Fort, N. Machetti, S. Rocchi, M.B.S. Santos, L. Tondi, N. Olivieri, V. Vignoli, and G. Sberveglieri, "Tin Oxide Gas Sensing: Comparison Among Different Measurement Techniques for Gas Mixture Classification," *IEEE Trans. Instrumentation and Measurement*, vol. 52, no. 3, pp. 921-926, 2003.
- [5] F. Hossein-Babaei and A. Amini, "Recognition of complex odors with a single generic tin oxide gas sensor," *Sens. Actua. B*, vol. 194, pp. 156-163, 2014.
- [6] L. Zhang and F.C. Tian, "A new kernel discriminant analysis framework for electronic nose recognition," *Analytica Chimica Acta*, vol. 816, pp. 8-17, 2014.
- [7] A. Flammini, D. Marioli, and A. Taroni, "A Low-Cost Interface to High-Value Resistive Sensors Varying Over a Wide Range," *IEEE Trans. Instrumentation and Measurement*, vol. 53, no. 4, pp. 1052-1056, 2004.
- [8] K. Brudzewski, S. Osowski, and A. Dwulit, "Recognition of Coffee Using Differential Electronic Nose," *IEEE Trans. Instrumentation and Measurement*, vol. 61, no. 6, pp. 1803-1810, 2012.
- [9] F. Herrero-Carrón, D.J. Yáñez, F. de Borja Rodríguez, and P. Varona, "An active, inverse temperature modulation strategy for single sensor odorant classification," *Sens. Actua. B*, vol. 206, pp. 555-563, 2015.
- [10] R. Gosangi and R. Gutierrez-Osuna, "Active Temperature Programming for Metal-Oxide Chemoresistors," *IEEE Sensors Journal*, vol. 10, no. 6, pp. 1075-1082, 2010.
- [11] X. Yin, L. Zhang, F. Tian, and D. Zhang, "Temperature Modulated Gas Sensing E-nose System for Low-cost and Fast Detection," *IEEE Sensors Journal*, vol. 16, no. 2, pp. 464-474, 2016.
- [12] I. Rodriguez-Lujan, J. Fonollosa, A. Vergara, M. Homer, R. Huerta, "On the calibration of sensor arrays for pattern recognition using the minimal number of experiments," *Chemometrics and Intelligent Laboratory Systems*, vol. 130, pp. 123-134, 2014.
- [13] S.K. Jha, K. Hayashi, and R.D.S. Yadava, "Neural, fuzzy and neuro-fuzzy approach for concentration estimation of volatile organic compounds by surface acoustic wave sensor array," *Measurement*, vol. 55, pp. 186-195, 2014.
- [14] L. Zhang, D. Zhang, X. Yin, and Y. Liu, "A Novel Semi-supervised Learning Approach in Artificial Olfaction for E-Nose Application," *IEEE Sensors Journal*, vol. 16, no. 12, pp. 4919-4931, 2016.
- [15] S.J. Dixon and R.G. Brereton, "Comparison of performance of five common classifiers represented as boundary methods: Euclidean Distance to Centroids, Linear Discriminant Analysis, Quadratic Discriminant Analysis, Learning Vector Quantization and Support Vector Machines, as dependent on data structure," *Chemometrics and Intelligent Laboratory Systems*, vol. 95, pp. 1-17, 2009.
- [16] B. Tudu, A. Metla, B. Das, N. Bhattacharyya, A. Jana, D. Ghosh, and R. Bandyopadhyay, "Towards Versatile Electronic Nose Pattern Classifier for Black Tea Quality Evaluation: An Incremental Fuzzy Approach," *IEEE Trans. Instrumentation and Measurement*, vol. 58, no. 9, pp. 3069-3078, 2009.
- [17] L. Zhang, F. Tian, H. Nie, L. Dang, G. Li, Q. Ye, and C. Kadri, "Classification of multiple indoor air contaminants by an electronic nose and a hybrid support vector machine," *Sensors and Actuators B: Chemical*, vol. 174, pp. 114-125, 2012.
- [18] S. Marco and A. Gutiérrez-Gámez, "Signal and Data Processing for Machine Olfaction and Chemical Sensing: A Review," *IEEE Sensors Journal*, vol. 12, no. 11, pp. 3189-3214, 2012.
- [19] H. Alam, S.H. Saeed. "Modern applications of electronic nose: a review." *International Journal of Electrical and Computer Engineering*. vol. 3, no. 1, pp.52, 2013.
- [20] L. Zhang, W.M. Zuo, and D. Zhang, "LSDT: Latent Sparse Domain Transfer Learning for Visual Adaptation," *IEEE Trans. Image Processing*, vol. 25, no. 3, pp. 1177-1191, 2016.
- [21] Q. Liu, M. Ye, S.S. Ge, and X. Du, "Drift Compensation for Electronic Nose by Semi-Supervised Domain Adaption," *IEEE Sensors Journal*, vol. 14, no. 3, pp. 657-665, 2014.
- [22] L. Zhang, F. Tian, C. Kadri, B. Xiao, H. Li, L. Pan, and H. Zhou, "On-line sensor calibration transfer among electronic nose instruments for monitoring volatile organic chemicals in indoor air quality," *Sens. Actua. B*, vol. 160, no. 1, pp. 899-909, 2011.
- [23] K. Yan and D. Zhang, "Improving the transfer ability of prediction models for electronic noses," *Sens. Actua. B*, vol.220, pp. 115-124, 2015.
- [24] O. Tomic, H. Ulmer, and J.E. Haugen, "Standardization methods for handling instrument related signal shift in gas-sensor array measurement data," *Analytica Chimica Acta*, vol. 472, pp. 99-111, 2002.
- [25] L. Zhang, F. Tian, S. Liu, L. Dang, X. Peng, and X. Yin, "Chaotic time series prediction of e-nose sensor drift in embedded phase space," *Sens. Actua. B*, vol. 182, pp. 71-79, 2013.
- [26] K. Yan and D. Zhang, "Correcting Instrumental Variation and Time-varying Drift: A Transfer Learning Approach with Autoencoders," *IEEE Trans. Instrumentation and Measurement*, 2016.
- [27] E. Martinelli, G. Magna, S. De Vito, R. Di Fuccio, G. Di Francia, A. Vergara, and C. Di Natale, "An adaptive classification model based on the Artificial Immune System for chemical sensor drift mitigation," *Sens. Actua. B*, vol. 177, pp. 1017-1026, 2013.
- [28] A. Vergara, S. Vembu, T. Ayhan, M.A. Ryan, M.L. Homer, and R. Huerta, "Chemical gas sensor drift compensation using classifier ensembles," *Sens. Actua. B*, vol. 167, pp. 320-329, 2012.
- [29] L. Zhang and D. Zhang, "Domain Adaptation Extreme Learning Machines for Drift Compensation in E-nose Systems," *IEEE Trans. Instrumentation and Measurement*, vol. 64, no. 7, pp. 1790-1801, 2015.
- [30] M. Padilla, A. Perera, I. Montoliu, A. Chaudry, K. Persaud, and S. Marco, "Drift compensation of gas sensor array data by Orthogonal Signal Correction," *Chemometrics and Intelligent Laboratory Systems*, vol. 100, pp. 28-35, 2010.
- [31] S. Di Carlo, M. Falasconi, E. Sanchez, A. Scionti, G. Squillero, and A. Tonda, "Increasing pattern recognition accuracy for chemical sensing by evolutionary based drift compensation," *Pattern Recognition Letters*, vol. 32, pp. 1594-1603, 2011.
- [32] I. Jolliffe, "Principal Component Analysis," *Wiley Online Lib.*, 2002.
- [33] J. Ye, R. Janardan, and Q. Li, "Two-dimensional linear discriminant analysis," *NIPS*, 2004.
- [34] X. He and P. Niyogi, "Locality preserving projections," *NIPS*, 2004.
- [35] S. Yan, D. Xu, B. Zhang, H. Zhang, Q. Yang, S. Lin, "Graph embedding and extensions: a general framework for dimensionality reduction," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.29, no.1, pp.40-51, 2007.
- [36] T. Feam, "On orthogonal signal correction," *Chemometr. Intell. Lab. Syst.*, vol. 50, pp. 47-52, 2000.
- [37] L. Fernandez, S. Guney, A. Gutiérrez-Gámez and S. Marco. "Calibration transfer in temperature modulated gas sensor arrays," *Sensors and Actuators B: Chemical*, vol. 231, pp. 276-284, 2016.
- [38] J. Fonollosa, L. Fernández, A. Gutiérrez, R. Huerta and S. Marco. "Calibration transfer and drift counteraction in chemical sensor arrays using Direct Standardization," *Sensors and Actuators B: Chemical*, vol. 236, pp. 1044-1053, 2016.
- [39] W. S. Torgerson, "Multidimensional scaling: I. Theory and method," *Psychometrika*, vol. 17, no. 4, pp. 401-419, 1952.
- [40] J. Goldberger, S. Roweis, G. Hinton, R. Salakhutdinov, "Neighborhood component analysis," *NIPS*, 2004.
- [41] X. He, D. Cai, S. Yan, and H.J. Zhang, "Neighborhood Preserving Embedding," *ICCV*, 2005.
- [42] M. Sugiyama, "Local Fisher Discriminant Analysis for Supervised Dimensionality Reduction," *ICML*, pp. 905-912, 2006.
- [43] B. Gong, Y. Shi, F. Sha, and K. Grauman, "Geodesic flow kernel for unsupervised domain adaptation," *CVPR*, pp. 2066-2073, 2012.
- [44] R. Gopalan, R. Li, and R. Chellappa, "Domain adaptation for object recognition: An unsupervised approach," *ICCV*, 2011.
- [45] B. Fernando, A. Habrard, M. Sebban, and T. Tuytelarrs, "Unsupervised visual domain adaptation using subspace alignment," *ICCV*, pp. 2960-2967, 2014.



**Lei Zhang (M'14)** received his Ph.D degree in Circuits and Systems from the College of Communication Engineering, Chongqing University, Chongqing, China, in 2013. He was selected as a Hong Kong Scholar in China in 2013, and worked as a Post-Doctoral Fellow with The Hong Kong Polytechnic University, Hong Kong, from 2013 to 2015. He is currently a Professor/Distinguished Research Fellow with Chongqing University. He has authored more than 50 scientific papers in top journals, including the IEEE

TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, the IEEE TRANSACTIONS ON IMAGE PROCESSING, the IEEE TRANSACTIONS ON MULTIMEDIA, the IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT, the IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS: SYSTEMS, the IEEE SENSORS JOURNAL, INFORMATION FUSION, SENSORS & ACTUATORS B, and ANALYTICA CHIMICA ACTA. His current research interests include electronic olfaction, machine learning, pattern recognition, computer vision and intelligent systems. Dr. Zhang was a recipient of Outstanding Reviewer Award of Sensor Review Journal in 2016, Outstanding Doctoral Dissertation Award of Chongqing, China, in 2015, Hong Kong Scholar Award in 2014, Academy Award for Youth Innovation of Chongqing University in 2013 and the New Academic Researcher Award for Doctoral Candidates from the Ministry of Education, China, in 2012.



**Liu Yan** received her Bachelor degree in Information Engineering in 2014 from Chengdu Polytechnic University, China. Since September 2014, she is currently pursuing a MS degree in Chongqing University. Her research interests include electronic nose and intelligent algorithm.



**Pingling Deng** received her Bachelor degree in Information Science and Engineering in 2015 from Lanzhou University, China. Since September 2015, she is currently pursuing a MS degree in Chongqing University. Her research interests include machine learning and electronic nose.