 CMC, vol.57, no.1, pp.25-48, 2018

# Multi-task Joint Sparse Representation Classification Based on Fisher Discrimination Dictionary Learning

**Rui Wang[1] , Miaomiao Shen[1, *], Yanping Li[1] and Samuel Gomes[2]**

**Abstract:** Recently, sparse representation classification (SRC) and fisher discrimination dictionary learning (FDDL) methods have emerged as important methods for vehicle classification. In this paper, inspired by recent breakthroughs of discrimination dictionary learning approach and multi-task joint covariate selection, we focus on the problem of vehicle classification in real-world applications by formulating it as a multi-task joint sparse representation model based on fisher discrimination dictionary learning to merge the strength of multiple features among multiple sensors. To improve the classification accuracy in complex scenes, we develop a new method, called multi-task joint sparse representation classification based on fisher discrimination dictionary learning, for vehicle classification. In our proposed method, the acoustic and seismic sensor data sets are captured to measure the same physical event simultaneously by multiple heterogeneous sensors and the multi-dimensional frequency spectrum features of sensors data are extracted using Mel frequency cepstral coefficients (MFCC). Moreover, we extend our model to handle sparse environmental noise. We experimentally demonstrate the benefits of joint information fusion based on fisher discrimination dictionary learning from different sensors in vehicle classification tasks.

**Keywords:** Multi-sensor fusion, fisher discrimination dictionary learning (FDDL), vehicle classification, sensor networks, sparse representation classification (SRC).

## 1 Introduction

The past several years have witnessed the rapid development of multi-sensor fusion technology, as well as its successful applications both in military and non-military tasks [Gravina, Alinia, Ghasemzadeh et al. (2016); Zhang, Lin, Chang et al. (2016)]. Recently, multi-sensor fusion techniques are also hopeful to achieve glorious results in classification tasks. The success of classification based on multi-sensor fusion owes firmly to the fact that it takes advantage of obtaining related information from different heterogeneous sensors while the same physical events are recorded [Yan, Wang, Xue et al. (2016); Nguyen, Nasrabadi and Tran (2011)]. Meanwhile, a variety of approaches have been proposed in several literatures to optimize this problem. One popular strategy

[1] School of Communication and Information Engineering, Shanghai University, No. 99 Shangda Road, Shanghai, 200444, China.

[2] School of Mechatronics, Methods, Models and Skills Laboratory (M3M), Universite de Technologie de Belfort-Montbeliard, 90010 Belfort cedex, Belfort, France.

[*] Corresponding Author: Miaomiao Shen. Email: mmloven@shu.edu.cn.

is decision fusion, Duarte et al. [Duarte and Hu (2004)] proposed to perform local classification for each sensor signals and utilized the local decisions which incorporated through Maximum A Posterior estimator to make the final classification decision. Another popular strategy is feature fusion, Klausner et al. [Klausner, Tengg and Rinner (2007)] proposed to extract temporal gait patterns from both visual and acoustic sensors, and utilized them as inputs for the classifier. Extensive experiments are conducted on real data sets [Zhang, Lin, Chang et al. (2016)], and the results show that the feature fusion approach can obtain higher classification accuracy compared to the decision fusion approach in pattern classification tasks.

In this paper, we consider the vehicle classification problem to be an important signal processing tasks [Oh, Chung and Myung (2017); Tian, Dong and Jia (2014); Eom (1999)] and the multiple heterogeneous sensors are applied for such a purpose. We employ sparse representation theory [Wright, Yang and Ganesh (2009); Mei and Ling (2011); Gao, Tsang and Chia (2010)] to adapt to different situations and improve the classification performance, while making the use of the feature fusion approach to study the vehicle classification tasks. For sparse representation, dictionary learning plays a quite important role, and a number of dictionary learning methods [Mairal, Bach and Ponce (2012); Mairal, Bach, Ponce et al. (2008a); Mairal, Bach, Ponce et al. (2008b); Gu, Zhang, Zuo et al. (2014)] have been proposed to promote the discrimination of the learned dictionary.

In the case of the sparse representation with a synthesis dictionary [Mairal, Bach and Ponce (2012); Jiang, Lin and Davis (2013)], its representation coefficients of a signal are usually obtained through an $\ell_p$-norm$(p \leq 1)$ sparse coding process. Compared to the sparse representation with an analytical dictionary [Jiang, Zhe and Davis (2011)], it allows us to learn a desired dictionary from training samples more easily and can better model local structures of the complex image. Take the LC-KSVD [Jiang, Lin and Davis (2013)] for example, a single over-complete dictionary and an optimal linear classifier are learned jointly, which is suitable for the situation of limited memory resources, and it shows superior performance in face, action, scene, and object category classification. However, the shared dictionary loses the correspondences among the dictionary atoms and the class labels, so vehicle classification based on the class-specific representation residuals is not allowed.

To make the class labels of training samples available, sparse representation with a structured dictionary [Ramirez, Sprechmann and Sapiro (2010); Yang, Zhang and Feng (2011); Wang, Guo, Li et al. (2017)] exploits the class discrimination information and encodes the query samples over the learned dictionary, in which both the coding coefficients and the coding residuals can be used for classification. Ramirez et al. [Ramirez, Sprechmann and Sapiro (2010)] introduced a non-continuous promotion to ensure the independence of the subclasses of the different categories. Yang et al. [Yang, Zhang and Feng (2011)] proposed a fisher discrimination dictionary learning (FDDL) method based on the fisher discrimination criterion, in which the dictionary atoms have correspondences to the subject class labels, with which not only the representation residuals can be used to distinguish different classes, but also the representation coefficients have small within-class scatter and big between-class scatter. To incorporate FDDL with SRC [Mei and Ling (2011)], Wang et al. [Wang, Ramirez and Zhang (2017)] proposed an efficient algorithm for vehicle classification tasks and the vehicle classification scheme is handled through fisher discriminative dictionary learning method. However, the information among different sensors has not been considered. Therefore, Nam et al. [Nam, Nasser and Trac (2011)] proposed a novel multi-task multivariate (MTMV) sparse representation method for multi-task classification, which took advantage of different sensors having related information while recording the same physical event, and achieved excellent classification performance.

Motivated by discrimination dictionary learning approach and multi-task joint covariate selection [Zhang, Nasrabadi, Huang et al. (2011); Yuan and Yan (2010)], we consider the multi-task classification problem among multiple heterogeneous sensors [Cui (2015); Shrivastava, Patel and Chellappa (2014); Chen, Nasrabadi and Tran (2011)], and formulate the vehicle classification problem as a multi-task joint sparse representation classification model based on fisher discrimination dictionary learning. In the scheme of this proposed method, the acoustic and seismic sensor data sets are captured to measure the same physical event simultaneously by the multiple heterogeneous sensors during the real word wireless sensor networks (WDSN) experiment [Yan, Wang, Xue et al. (2016); Jiang, Zhe and Davis (2011)] , and the multi-dimensional frequency spectrum features of the sensor data are extracted using Mel frequency cepstral coefficients (MFCC) [Sahidullah and Saha (2013); Kinnunen, Saeidi and Sedlak (2012)], whose high efficiency has been proven in signal classification.

The contribution of our work is shown as follows. We consider a multi-task classification problem, and the correlations as well as complementary information among different sensors simultaneously. The experimental results show its great superiority when considering the importance of collaborative of heterogeneous sensors. Then, we propose a vehicle classification scheme associated with the multi-task joint sparse representation classification base on fisher discrimination dictionary learning, which uses the discriminative information in the reconstruction error together with sparse coding coefficients in classification tasks.

The remainder of this paper is organized as follows. Section II briefly introduces sparse representation classification and fisher discrimination dictionary learning methods. We present in Section III a framework of multi-task joint sparse representation model based on fisher discrimination dictionary learning in sensor networks for vehicle classification. Section IV mainly describes the algorithm of our proposed method in detail. Extensive experiments are shown in Section V and conclusions are drawn in Section VI.

## 2 Related works

### *2.1 Sparse Representation Classification (SRC)*

Sparse representation is a typical signal processing method to represent the main information of a signal using non-zero coefficients as few as possible. For object recognition, our goal is to classify the testing sample using labeled training data. Here, our central approach is to represent the testing sample as a sparse linear combination of training samples.

We arrange the $k_i$ training samples from the $i$-th class as a matrix $A_i = [v_{i,1}, v_{i,2}, \cdots, v_{i,n_i}] \in R^{m \times k_i}$ and define a new matrix $A$ for the entire training sets as $A = [A_1, A_2, \cdots, A_k] \in R^{m \times n}$, $k$ is the number of classes in training set. Given sufficient training samples of the $i$-th class, any test sample $y \in R^m$ from the same class will approximately lie in the hyperplane spanned by the training samples of class $i$ as Eq. (1):

$$y = x_{i,1}v_{i,1} + x_{i,2}v_{i,2} + \cdots + x_{i,k_i}v_{i,k_i} \tag{1}$$

Then, the above representation of $y$ can be rewritten in matrix form as: $y = A_i x_i$, where $x_i = [x_{i,1}, x_{i,2}, \cdots, x_{i,k_i}] \in R^{k_i}$. This motivates us to seek the sparsest solution by solving the following optimization problem in Eq. (2):

$$(\ell_0): \quad \hat{x}_0 = \arg\min \|x\|_0 \quad s.t. \quad y = Ax \tag{2}$$

Generally, if the solution sought is sparse enough, the solution of the $\ell_0$-minimization problem is equal to the solution of the following $\ell_1$ -minimization problem as Eq. (3):

$$(\ell_1): \quad \hat{x} = \arg\min \|x\|_1 \quad s.t. \quad y = Ax \tag{3}$$

Let $\hat{x}$ be the solution of Eq. (3), for each class $i$, let $\delta_i$ be the characteristic function that selects the coefficients associated with the $i$-th class. Using the coefficients, one can approximate the given test sample $y$ as $\hat{y}_i = A\delta_i(\hat{x})$, where $\delta_i(\hat{x}) = [0, \cdots, x_{i,1}, x_{i,2}, \cdots, x_{i,k_i}, \cdots, 0]$. We then compute the residual $r_i(y)$ between $y$ and $\hat{y}$ as Eq. (4):

$$r_i(y) = \|y - A\delta_i(\hat{x})\|_2 \tag{4}$$

The test sample $y$ is identified by minimizing $r_i(y)$ as Eq. (5):

$$\text{identity}(y) = \arg\min_i r_i(y) \tag{5}$$

### *2.2 Fisher Discrimination Dictionary Learning (FDDL)*

The vast majority of the dictionary learning methods use a shared dictionary method to achieve the sparse representation of different types of signals, which makes signal classification can only be performed by reconstruction errors. To improve the performance of discrimination dictionary learning methods, Yang et al. [Yang, Zhang and Feng (2011)] construct a structured dictionary $D = [D_1, D_2, \cdots, D_j, \cdots, D_K]$, where $D_j$ is the class-specified sub-dictionary associated with class $j$ and $K$ is the classes of objects. Suppose the given training samples $A = [A_1, A_2, \cdots, A_j, \cdots, A_K]$ , where $A_j = [A_{j,1}, A_{j,2} \cdots A_{j,h_j}] \in R^{m \times h_j}$ is the subset of the training samples from class $j$, $h_j$ is the number of samples of class $j$ and $m$ is the features dimension of each sample. Denoted by the coding coefficient matrix $X = [X_1, X_2, \cdots, X_j, \cdots, X_K]$ of $A$ over $D$, $A \approx DX$, where $X_j$ is the sub-matrix containing coding coefficients of $A_j$ over $D$. Since the learned dictionary $D$ both have the powerful reconstruction capability and discriminative capability of $A$, the discrimination dictionary learning model can be established as Eq. (6).

$$J(D,X)=\arg\min_{(D,X)}\left\{r(A,D,X)+\lambda_1\|X\|_1+\lambda_2 f(X)\right\} \tag{6}$$

where $r(A,D,X)$ is the discriminative fidelity term, $\|X\|_1$ is the sparsity constraint term, $f(X)$ is a discrimination constraint imposed on the coefficient matrix $X$, $\lambda_1$ and $\lambda_2$ are scalar parameters.

### 2.2.1 Discriminative fidelity term $r(A,D,X)$

In this part, we study the discriminative fidelity term $r(A,D,X)$ based on the fisher discrimination criterion. Suppose $X_l=[X_{l,1},X_{l,2},\cdots,X_{l,j},\cdots,X_{l,K}]$, where $X_{l,j}$ is the coding coefficient of $A_l$ over the sub-dictionary $D_j$, then we try to denote the representation of $D_k$ to $X_l$ as $R_k=D_k X_{l,k} X_l$. Since $A_l$ can be well represented by the dictionary $D$, we get $A_l$ by the coding coefficient $X$ and the dictionary $D$ as Eq. (7).

$$A_l\approx DX_l=D_1X_{l,1}+...+D_lX_{l,l}+...+D_KX_{l,K}=R_1+...+R_i+...+R_KD \tag{7}$$

Since $D_l$ is associated with the $l$-th class, we expect that $A_l$ should be well represented by $D_l$ but not by $D_j$, where $j\neq l$. It means that $X_{l,l}$ has some significant coefficients to minimize $\|A_l-D_lX_{l,l}\|_F^2$, while $X_{l,j}$ have nearly zero coefficients to minimize $\|D_lX_{l,j}\|_F^2$. Therefore, the discriminative fidelity term can be defined as Eq. (8).

$$r(A_l,D,X_l)=\|A_l-DX_l\|_F^2+\|A_l-D_lX_{l,l}\|_F^2+\sum_{j=1,j\neq l}^{K}\|D_jX_{l,j}\|_F^2 \tag{8}$$

### 2.2.2 Discriminative coefficient term $f(X)$

In order to make the learning dictionary $D$ discriminative for all the samples in $A$, we minimize the within-class scatter of $X$, denoted by $S_W(X)$ as Eq. (9), and maximize the between-class scatter of $X$, denoted by $S_B(X)$ as Eq. (10), based on the fisher discrimination criterion.

$$S_W(X)=\sum_{j=1}^{K}\sum_{x_l\in X_j}(x_l-n_j)(x_l-n_j)^T \tag{9}$$

$$S_B(X)=\sum_{j=1}^{K}p_j(n_j-n)(n_j-n)^T \tag{10}$$

where $n_j$ is the mean vector of $X_j$, $n$ is the mean vector of $X$, and $p_j$ is the number of samples in class $A_j$. After that, we define $f(X)$ as $tr(S_W(X))-tr(S_B(X))$, where $tr(\cdot)$ means the trace of a matrix. Since the discriminative coefficient term $f(X)$ is non-convex and unstable, we try to bring an elastic term $\|X\|_F^2$ into $f(X)$ to make it strictly convex. Then, the discriminative coefficient term $f(X)$ can be represented as Eq. (11),

$$f(X)=tr(S_W(X))-tr(S_B(X))+\eta\|X\|_F^2 \tag{11}$$

where $\eta$ is a scalar constant.

Based on the definition of $r(A,D,X)$ and $f(X)$ above, the fisher discrimination dictionary learning model can be obtained by incorporating Eq. (8) and (11) into Eq. (6).

$$J(D,X)=\arg\min_{(D,X)}\left\{\sum_{j=1}^{K} r(A_j,D,X_j)+\lambda_1\|X\|_1+\lambda_2(tr(S_W(X)-S_B(X))+\eta\|X\|_F^2)\right\} \tag{12}$$

Although the objective function $J(D,X)$ in Eq. (12) is not jointly convex to $(D,X)$, it is convex respect to each of $D$ and $X$ when the other is fixed. This optimization problem can be divided into two sub-problems by optimizing $D$ and $X$ alternatively: Fixed $D$, then updating $X$; and fixed $X$, then updating $D$. When the dictionary $D$ is fixed, the objective function in Eq. (12) is reduced to be a sparse representation problem to compute $A=[A_1,A_2\cdots A_j\cdots,A_K]$ and it can be resolved by the Iterative Projection Method (IPM). However, when $X$ is fixed, the objective function in Eq. (12) is a quadratic programming problem and it can be efficiently solved by updating each dictionary atom one by one.

## 3 The framework of multi-task joint sparse representation classification based on FDDL for moving vehicle classification

The main content of this research is the target classification and recognition based on acoustic and seismic sensor network, the main object is the moving vehicle. Considering the success of multi-feature fusion method and fisher discrimination dictionary learning in running vehicle classification task, we proposed a multi-task joint sparse representation based on fisher discrimination dictionary learning method for vehicle classification. To analyze the performance of this proposed method, we extend our approaches for vehicle classification task based on the acoustic and seismic signal of running vehicles. The vehicle classification framework is shown in Fig. 1, which has the following major components.

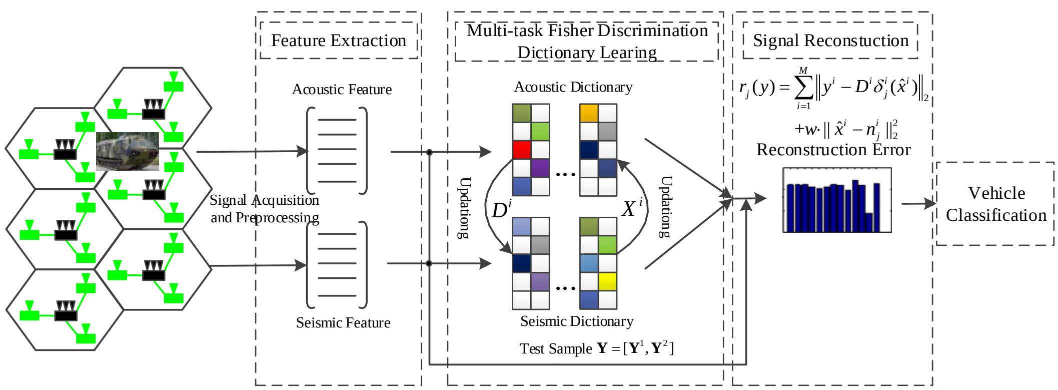



**Figure 1:** The framework of multi-task joint sparse representation based on FDDL for vehicle classification

### *3.1 Pre-processing*

The raw acoustic and seismic signals of vehicles are gathered from the multiple heterogeneous sensor nodes in complex scenes using sensor networks. However, it is inevitable that the signals will be disturbed by the ambient noise and some other uncertain conditions, so the pre-processing must be performed before the signal is extracted to remove unnecessary noise and other uncertainties. In the procedure of pre-processing, considering the useful event series span a short period of time when the vehicles are close to the sensor nodes, Constant False Alarm Rate (CFAR) algorithm is used to detect whether the vehicles are present and finally the useful event series are converted to frames. The main procedures include, (1) signal pre-emphasis: to enhance the high-frequency part of the signal; (2) sub-frame processing: the sound signal is divided into short-term stationary signal; (3) windowing process: used to reduce the truncation effect, so that the signal can have the smooth transition to zero; (4) endpoint detection: used to intercept the starting and ending point of the acoustic and seismic signal.

### *3.2 Feature extraction*

The event time series are widely used for classification purposes. However, the acoustic and seismic signals in the time domain often change rapidly over time and appear to be not stationary. Thus, large quantities of methods for feature extraction have been developed in the frequency domain, since they can be considered to be quasi-stationary and analyzed using Fourier transform. Among them, MFCC [Sahidullah and Saha (2013); Kinnunen, Saeidi, Sedlak et al. (2012)] is more widespread used owing to its robustness, for considering the variation of human ear's critical bandwidths with frequency.

The main procedures of MFCC include: (a) Fast Fourier Transform (FFT) [Lü and Peng (2104)], it transforms the signal from time domain to frequency domain; (b) Mel Filtering, the Mel filter banks are actually composed of triangle filter banks which take advantage of the similar properties with human ear. Through the Mel filtering, we can get the Mel spectral coefficients; (3) Taking the Logarithm, the aim of taking the logarithm of the Mel spectral coefficients is to compress the dynamic range of the spectrum, and at the same time the multiplicative noise is removed; (4) Discrete Consine Transform (DCT), it converts the log Mel spectrum to time domain and the results which called the Mel frequency cepstral coefficients are the features we need. In the procedure of this paper, we use MFCC to extract the multi-dimensional frequency spectrum features of target vehicles.

### *3.3 Multi-task joint sparse representation classification based on FDDL*

By exploiting the correlation as well as complementary information among different heterogeneous sensors, we construct a multi-task fisher discrimination dictionary based on multi-feature signals, which makes the discriminative information of the acoustic and seismic features in both the representation residuals and the representation coefficients can be used for vehicle classification tasks. Then output the multi-task fisher discrimination dictionary for vehicle classification tasks.

### *3.4 Vehicle classification*

After learning a multi-task fisher discrimination dictionary by using multi-feature signals, whose atoms have correspondences to the subject class labels, the vehicle classification problem is transferred to represent the test samples sparsely using the learned dictionary $D$ . Finally, we get more accurate classification results by minimizing the decision fusion function.

## 4 Classification models

### *4.1 Sparse representation classification based on FDDL*

In this part, we study the single-task sparse representation classification method based on fisher discrimination dictionary. In the sparse representation model, we assume that the training samples belonging to the same class approximately lie on a low-dimensional subspace. Then the test sample $y \in R^{n*1}$, which belongs to the $j$-th class, will approximately liner subspace spanned by the fisher discrimination atom associated with the $j$-th class as Eq. (13),

$$y = Dx \tag{13}$$

where $x$ is the sparse coefficient vector.

For the reason that the number of training samples in the vehicle classification is relatively small, the query samples of this class may not be able to be faithfully represented by the sub-dictionary $D_j$. Thus, we choose to represent the test sample $Y$ over the whole fisher discrimination dictionary $D$. Then the sparse coding coefficients $\hat{x}$ can be represented by the Eq. (14),

$$\hat{x} = \arg\min_x \| y - Dx \|_2^2 + \gamma \| x \|_p \tag{14}$$

where $\gamma$ is a constant, $\|\cdot\|_p$ denotes $\ell_p$-norm, $p = 1$ or $2$. In the training stage of fisher discrimination dictionary learning, we have enforced the class-specific representation residuals to be discriminative.

Thus, in the testing stage, the residual $\| y - D_j \hat{x}_j \|_2^2$ should be small while $\| y - D_i \hat{x}_i \|_2^2$, $i \neq j$, should be big. In addition, the representation vector $\hat{x}$ is close to $n_j$ but far from the mean vectors of other classes. Therefore, the class label of $y$ can be defined as Eq. (15) for final classification decision.

$$r_j(y) = \left\| y - D\delta_j(\hat{x}) \right\|_2 + w \cdot \| \hat{x} - n_j \|_2^2 \qquad j \in \{1, 2, \cdots, K\} \tag{15}$$

where the former term in Eq. (15) is the reconstruction error of class $j$, the latter term is the distance between the coefficient vector $\hat{x}$ and the learned mean vector $n_j$ of class $j$, and $w$ is a preset weight to balance the contribution of the two terms. $\delta_j$ denotes the characteristic function that selects the coefficients associated with the $j$-th class. Finally, get the identity of $y$ by minimizing the residual error as Eq. (16).

$$\text{identity}(y) = \arg\min_j r_j(y) \tag{16}$$

### *4.2 Multi-task joint sparse representation classification based on FDDL*

In the previous section, we employed single sensor sparse representation models for classification. However, in the scenario where an event is captured by multiple heterogeneous sensors, multiple observations from different sensors are available for classification tasks. By exploiting the correlation as well as the complementary information among different sensors, we can potentially improve the classification accuracy.

In this section, we consider a multi-task $K$ -class classification problem into account, and suppose have a training set of $h$ samples, in which each sample has $M$ different feature modalities. Given the training samples $A^i = \left[A_1^i, A_2^i, \cdots, A_j^i, \cdots, A_K^i\right], i \in \{1, 2, \cdots, M\}$, where $A_j^i = \left[A_{j,1}^i, A_{j,2}^i, \cdots, A_{j,h_j}^i\right] \in R^{m \times h_j}$ denotes the sub-set of the training samples of the $i$-th sensor corresponding to the $j$-th class. For the reason that $h_j$ is the number of training sample for the $j$-th class, the total samples can be represented as $h = \sum_{j=1}^{K} h_j$. In the testing phase, given a testing sample $y$ consisting of $M$ tasks $[y^1, y^2, \cdots, y^i, \cdots, y^M]$, then the ultimate goal of this paper is to focus on how to determine which category $y$ belongs to.

For each sensor $i = 1, 2, \cdots, M$ we denote $D^i = [D_1^i, D_2^i, \cdots, D_j^i, \cdots, D_K^i]$ as a discriminative dictionary corresponding to $M$ sensors respectively, consisting of $K$ sub-dictionaries $D_j^i$ with respect to $K$ classes. Here, each sub-dictionary $D_j^i = [D_{j,1}^i, D_{j,2}^i, \cdots, D_{j,h_j}^i] \in R^{m \times h_j}$ represents a set of training data from the $i$-th sensor labeled with the $j$-th class, where $D_{j,l}^i$ is the $l$-th dictionary atom respect to the $i$-th sensor with the $j$-th class label. For the testing sample $y = [y^1, y^2, \cdots y^i \cdots, y^M]$, suppose $y^1$ belongs to the $i$-th class, then it can be reconstructed under the Sparse Representation Classification model as Eq. (17),

$$y^1 = D^1 x^1 + z^1 \tag{17}$$

where $x^1$ is a sparse matrix associated with the over-complete dictionary $D^1$, and $z^1$ is a small noise matrix.

Similarly, for the reason that $y^1$ and $y^2, y^3, \cdots, y^i, \cdots, y^M$ represent the same event and belong to the same class, thus it also can be reconstructed by the sub-dictionary $D^2, D^3, \cdots, D^i, \cdots, D^M$ and its relative sparse coding matrix $x^2, x^3, \cdots, x^i, \cdots, x^M$ shown as Eq. (18),

$$\begin{aligned} y^2 &= D^2 x^2 + z^2 \\ y^3 &= D^3 x^3 + z^3 \\ &\vdots \\ y^i &= D^i x^i + z^i \\ &\vdots \\ y^M &= D^M x^M + z^M \end{aligned} \tag{18}$$

where $x^2, x^3, \cdots, x^i, \cdots, x^M$ has the same nonzero coefficient term as $x^1$.

We also define $x = \left[x^1, x^2, \cdots x^i \cdots, x^M\right]$, then $x$ is a sparse matrix with only $q_j$ nonzero rows. Based on this idea, we combined the Fisher Discriminative Dictionary Learning model in Eq. (12) and established the multi-task joint sparse representation classification based on FDDL for fisher dictionary learning as Eq. (19),

$$J(D, X) = \arg\min_{(D,X)} \sum_{i=1}^{M} \left\{ \sum_{j=1}^{K} r(A_j^i, D^i, X_j^i) + \lambda_1 (tr(S_W(X^i) - S_B(X^i)) + \eta \| X^i \|_F^2) \right\} + \lambda_2 \|X\|_1 \tag{19}$$

where $\lambda_1$ , $\lambda_2$ and $\eta$ are three scalar constants, the discriminative coefficient term $\lambda_1(tr(S_W(X^i)-S_B(X^i))+\eta\|X^i\|_F^2)$ and the discriminative fidelity item $r(A_j^i, D^i, X_j^i)$ are as the same definitions as the above.

By updating the discriminative dictionary $D$ and the sparse coding matrix $X$ iteratively, the learned multi-task fisher discriminative dictionary $D$ can be obtained, as same the specific solution as FDDL. Then, we focus on the sparse representation classification of $y$ through the multi-task fisher discriminative dictionary $D$ . By using the learned dictionary $D$ , the original signal can be reconstructed, and the moving vehicle type can be classified by the minimal residuals as Eq. (20). The specific classification decision function is shown as Eq. (21),

$$r_j(y)=\sum_{i=1}^{M}\left\|y^i-D^i\delta_j^i(\hat{x}^i)\right\|_2+w\cdot\|\hat{x}^i-n_j^i\|_2^2 \quad j\in\{1,2,\cdots,K\} \quad i\in\{1,2,\cdots,M\} \tag{20}$$

$$\text{identity}(y)=\arg\min_j r_j(y) \tag{21}$$

where $\delta_j^i$ is a unit matrix corresponding to the $j$-th class. Tab. 1 gives the algorithm of multi-task joint sparse representation classification based on FDDL.

**Table 1:** Multi-task joint sparse representation classification based on FDDL

1. Input the training sample $A^i=\left[A_1^i, A_2^i, \cdots, A_j^i, \cdots, A_K^i\right]$ , testing sample $y=[y^1, y^2, \cdots y^i \cdots, y^M]$ , $i\in\{1,2,\cdots,M\}$ .

2. Initialize the sub-dictionary $D_j^i$ over the training sample feature data $A_j^i$ .

3. Fix the discriminative dictionary $D^i$ , solve the problem of sparse coding by Iterative Projection Method (IPM) to update sparse coding matrix $X^i$ .

4. Fix the sparse coding matrix $X^i$ , solve the quadratic programming problem to update the learning dictionary $D^i$ .

5. Return to step 3 until convergence or the maximal iteration number are reached.

6. Compute the residuals.

7. Output the identity of $y$ .

$\text{identity}(y)=\arg\min_j r_j(\mathrm{y})$

## 5 Experimental analysis

In this section, in order to evaluate the performance of the proposed method, and verify the superiority compared to other classification methods, we conduct our extension experiments on the sensor data set collected from a real word wireless sensor networks (WDSN) in Twenty-Nine Palms, CA in November 2001. In this experiment, the acoustic, seismic and infrared sensors were deployed at the Marine Corps Air Ground Combat

Center. The sensor data set is available at http://www.ecs.umass.edu/~mduarte/Software.html [Duarte and Hu (2004)]. It contains the acoustic, seismic and infrared information of two types of military vehicles, namely Assault Amphibian Vehicle (AAV) and Dragon Wagon (DW). The original time series data are collected from eighteen sensor nodes distributed over three preset running routes, as shown in Fig. 2.

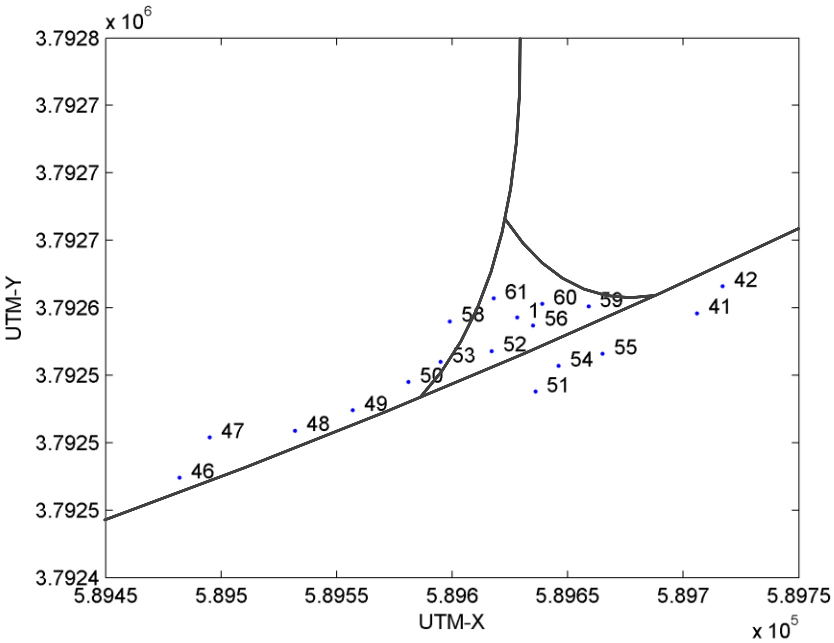



**Figure 2:** Sensor field layout

In the scenario where the vehicle running event is captured by multiple heterogeneous, multiple observations are available for vehicle classification from different sensors. In this paper, we propose the vehicle classification framework based on our proposed method in the acoustic and seismic sensor networks. The vehicle classification through the acoustic and seismic sensor networks contains three stages: pre-processing steps, feature extraction and vehicle classification.

### *5.1 Pre-processing steps*

The origin acoustic and seismic signals of vehicle are gathered from multiple heterogeneous sensor nodes at a rate of 4960 Hz in complex scenes, so it is inevitable that the signs will be disturbed by the noise and some other uncertain conditions during the experiment. In this experiment, it is needed to extract the actual event from the run series to reduce the accidental errors. However, the run time might be several minutes in length, the event series will be much shorter, as it only spans the short period of time when the target vehicle is close to the node. Thus, we use the CFAR detection algorithm to determine whether the vehicle is present in the region or not. All experiments in this paper are run on a desktop PC with Intel(R) Core(TM) i5-2467M 1.60 GHz CPU and 4

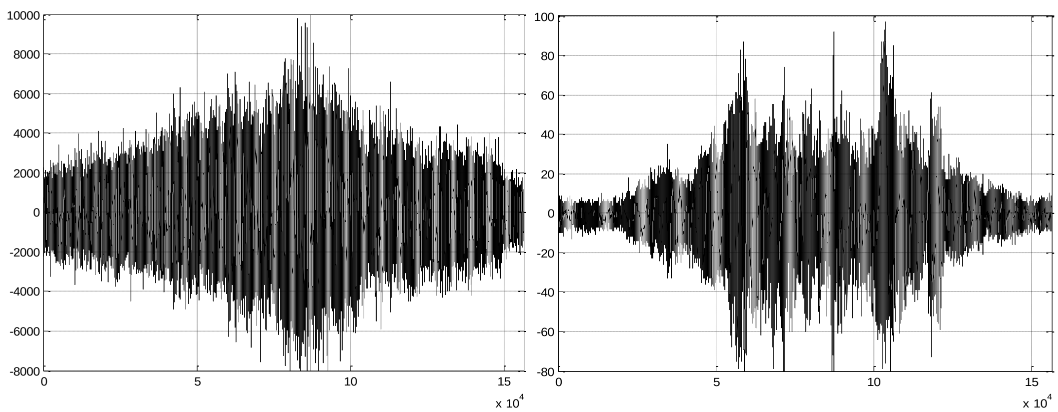
GB memory, and the sensor data sets was captured by the Defense Advanced Research Program in the DARPA/IXOs SensIT program through a truly distributed wireless distribution sensor networks.

### *5.2 Feature extraction*

Since the acoustic and seismic sensor data were recorded at a rate of 4960 Hz by microphones equipped on multiple heterogeneous sensor nodes, it is inevitable that the signals will be disturbed by the noise and some other uncertain conditions during the experiment. In order to reduce the accidental error of experiment, we classify the sensor databases by increasing the amount of test data and calculating the means of multiple tests.

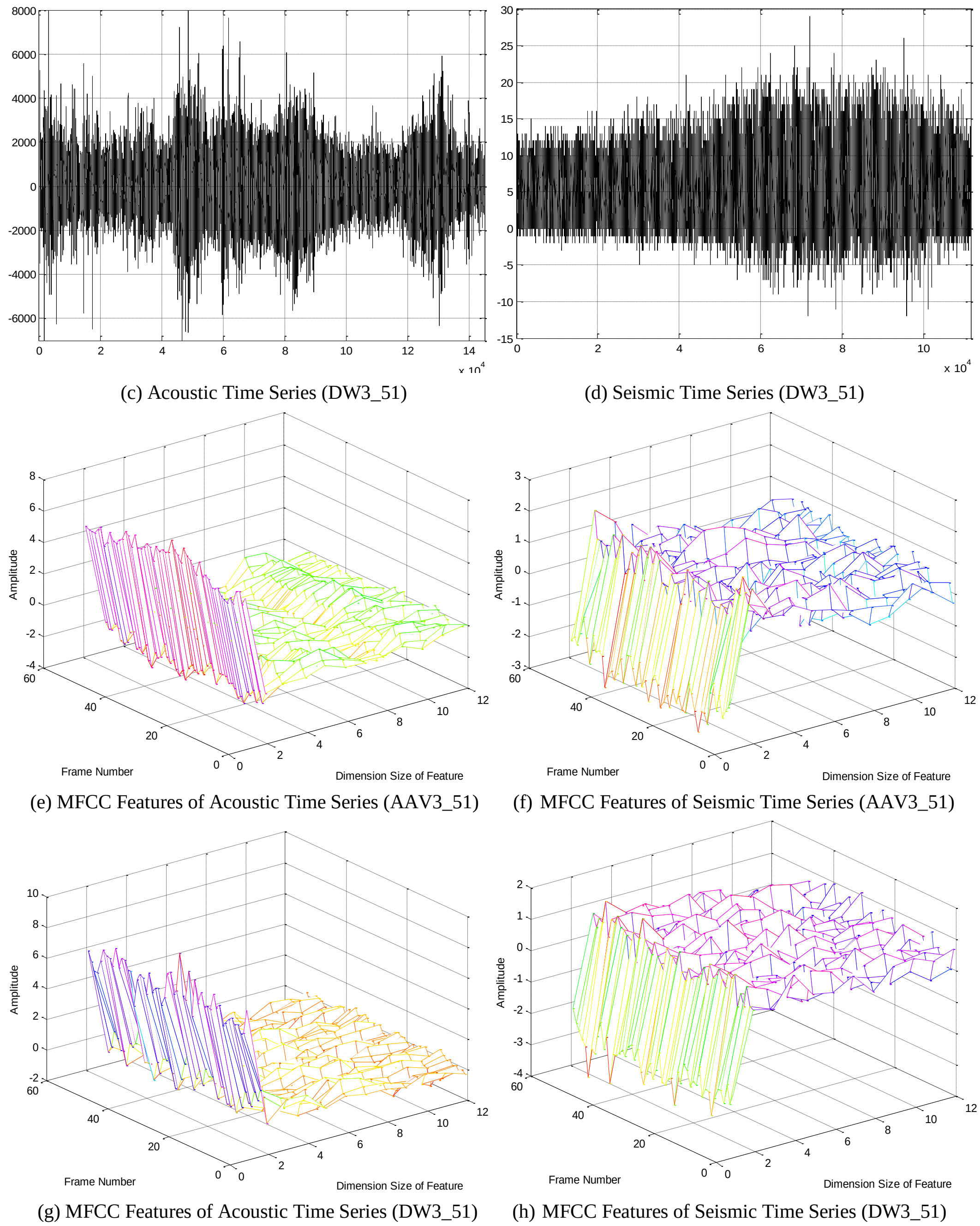
First of all, we choose the acoustic and seismic sensor data as shown in Figs. 3(a), 3(b), 3(c) and 3(d) collected by the nodes of the forty-one to sixty when the two kinds of military vehicles run from the third to eleventh, namely AAV3_41~AAV11_60 and DW3_41~DW11_60, and part data in the noise environment, so we have 450 sets of sensor data as the data source to assess feature extraction and classification tasks. To draw the useful events from raw time series data, we use constant false alarm rate (CFAR) detection algorithm to mark times according to high energy values.

The multi-dimensional frequency spectrum features as shown in Figs. 3(e), 3(f), 3(g) and 3(h) are extracted from the event time series for classification purposes using MFCC [Sahidullah and Saha (2013); Kinnunen, Saeidi, Sedlak et al. (2012)] algorithm.




(a) Acoustic Time Series (AAV3_51)

(b) Seismic Time Series (AAV3_51)




(c) Acoustic Time Series (DW3_51)

(d) Seismic Time Series (DW3_51)

(e) MFCC Features of Acoustic Time Series (AAV3_51)

(f) MFCC Features of Seismic Time Series (AAV3_51)

(g) MFCC Features of Acoustic Time Series (DW3_51)

(h) MFCC Features of Seismic Time Series (DW3_51)

**Figure 3:** Sample time series and features extracted by MFCC

### *5.3 Vehicle classification*

After feature extraction by MFCC, the multiple dimensional frequency spectrum features of vehicles are used for the proposed classification method to improve classification accuracy and reduce time complexity for vehicle classification tasks. We selected 75, 90, 105, 120, 135, 150 sets of acoustic and seismic sensor data, respectively and arbitrarily, as the training data to learn an acoustic fisher discrimination dictionary and a seismic fisher discrimination dictionary, and to classify the target vehicles. To speed up the process of multi-task fisher discrimination dictionary learning model, while ensuring that the classification efficiency is not reduced, the maximal iteration number is set 25, and the size of dictionary is set 30.

In this part, we demonstrate the effectiveness of using single-sensor and multi-task respectively for vehicle classification problem, particularly, in classifying the types of the moving vehicles. More specifically, we use the acoustic, the seismic and both of the acoustic and seismic sensor data as the samples data for the FDDL and our proposed methods to achieve the purpose of vehicle classification tasks, respectively. To achieve more reliable vehicle classification results, in this paper, we choose 300 arbitrary sets of acoustic and seismic sensor data as testing data to classify the type of target vehicles and get the classification rates by running 50 times classification procedure.

At the same time, the SVM [Seokhyeon, Yu and Taekwang (2017)], SRC [Mei and Ling (2011)], KSRC(GUASS) [Li (2016)], KSRC(POLY) [Kang and Kil (2015)], FDDL [Wang, Guo and Li (2017)], and MT-SRC [Nguyen, Nasrabadi and Tran (2011)] algorithms are also worked as references to this proposed method, and all of them utilize the acoustic signal to classify the types of the moving vehicles. Given the specific condition of the algorithms, the SVM algorithm is derived from Seokhyeon et al. [Seokhyeon, Yu and Taekwang (2017); Kachach and Cañas (2016)], where the optimization problem is solved by LIBSVM software package. The SRC algorithm is obtained from Mei et al. [Mei and Ling (2011)], in which the sparsity level is set to 0.7. While the Gauss-KSRC and Poly-KSRC algorithms are got from Li et al [Li (2016); Kang and Kil (2015)], which use Gaussian kernel and Polynomial kernel as their kernel function, and set 0.5 and 0.7 as their sparse level respectively. The LC-KSVD algorithm is proposed in Jiang et al. [Jiang, Zhe and Davis (2011)], in which the maximal iteration number is set 25 and the sparsity threshold is set 8. The MT-SRC algorithm is proposed in Nguyen et al. [Nguyen, Nasrabadi and Tran (2011)], in which the sparsity level is also set to 0.7.

#### *5.3.1 Classification performance*

**Single-sensor analysis:** Fig. 4 illustrates the trends of vehicle classification accuracy of the SVM, SRC, KSRC(GUASS), KSRC(POLY) and FDDL methods based on seismic sensor networks under different testing samples. It can be clearly seen from the figure that, under seismic signal, the SVM algorithm is superior to the SRC method for the vehicle classification performance. Especially, the classification accuracy of the SVM has a significant linear trend of growth with the increasing of the number of training samples. On the contrary, the classification accuracy of the SRC method shows a slight downward trend while the training samples increase.

Compared with the SRC algorithm, the KSRC algorithm has greatly improved the classification and recognition accuracy for the moving vehicle. However, regardless of the Polynomial kernel function or the Gaussian kernel function, the classification performance is slightly lower than the SVM algorithm.

Also, we know that, from the figure, whether it is under the acoustic or seismic signals of moving vehicles, compared with the SVM and SRC algorithm, the classification rates of the FDDL algorithm has been greatly improved, for the reason that the size of the over-complete dictionary in SRC is much larger than that of the fisher discrimination dictionary in FDDL, and these classification result are also superior to the KSRC methods based on Polynomial kernel and Gaussian kernel.

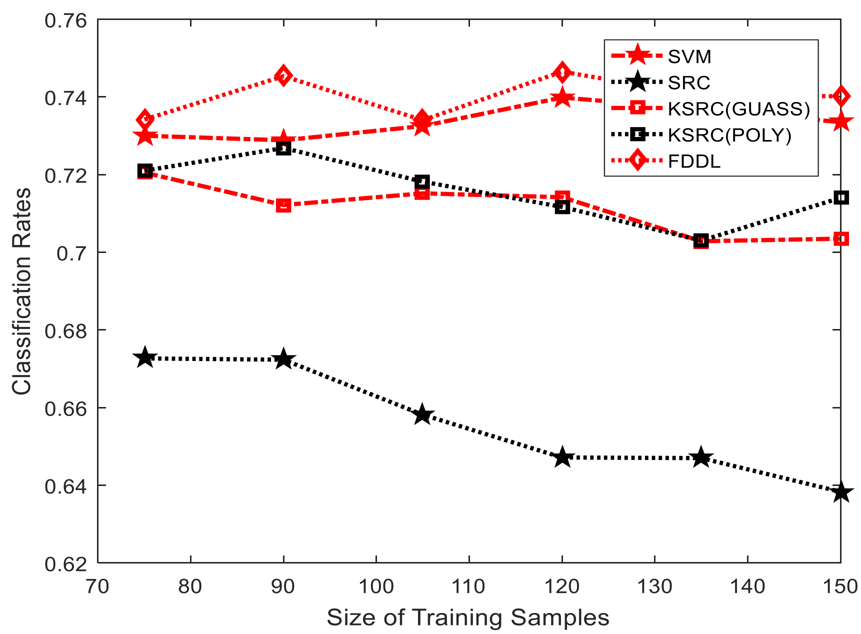



**Figure 4:** The trends of classification rates across various classification methods under seismic signals

The classification performance of the SVM, SRC, KSRC(GUASS), KSRC(POLY) and FDDL methods under acoustic signal are summarized in Tab. 2, where the detection rate, the false alarm rate and the classification accuracy is studied according to the different training samples size. Here, the detection rate is defined to be the ratio among the number of correct classification samples and the size of the class. The false alarm rate is defined to be the ratio among the number of incorrect classification samples and the total number of samples in other classes.

From the detection rates of noise in Tab. 2, we know that the FDDL algorithm can well recognize the background noise of the environment in acoustic sensor networks, with the ratio around 100%. It is also shown that the classification rates of the FDDL method based on acoustic signals (about 85.7%) is much too higher than the SVM, SRC and KSRC algorithms, and the classification rates of the FDDL algorithm is gradually increased with the increase number of training samples. All in all, the FDDL method shows prominently high performance.

**Multi-task analysis:** The trends of classification rates of the SVM, SRC, KSRC(GUASS), KSRC(POLY), FDDL, MT-SRC and our proposed methods are shown in Fig. 5. It shows the classification rate of the MT-SRC algorithm shows a slight upward trend with the increase number of the testing sample, compares to the SRC algorithm, whose classification rate shows a fluctuation trend or even downward trend. It is the truth that the multi-task feature fusion method utilizes the advantages of each signal feature to make the classification rates much greater than that of any kind of single sensor feature.

**Table 2:** The classification rates across various classification methods under acoustic signal (%)

| Classification Method | | Detection rates | | | False alarm rates | | | Classification rates |
|---|---|---|---|---|---|---|---|---|
| | | AAV | DW | Noise | AAV | DW | Noise | |
| SVM | 75 | 78.08 | 76.69 | 96.75 | 14.62 | 15.54 | 0.81 | 79.97 |
| | 90 | 80.23 | 79.31 | 96.37 | 13.18 | 13.79 | 0.91 | 81.98 |
| | 105 | 80.69 | 81.50 | 99.25 | 12.87 | 12.33 | 0.19 | 83.52 |
| | 120 | 82.88 | 82.92 | 97.12 | 11.41 | 11.38 | 0.72 | 84.80 |
| | 135 | 84.15 | 83.08 | 97.00 | 10.56 | 11.28 | 0.75 | 85.40 |
| | 150 | 83.69 | 83.50 | 96.75 | 10.87 | 11.00 | 0.81 | 85.35 |
| SRC | 75 | 78.19 | 79.27 | 98.12 | 17.26 | 16.67 | 0.47 | 81.95 |
| | 90 | 77.65 | 79.38 | 98.00 | 17.46 | 16.36 | 0.50 | 81.73 |
| | 105 | 80.31 | 78.96 | 97.50 | 15.69 | 16.51 | 0.63 | 82.65 |
| | 120 | 79.65 | 78.27 | 95.50 | 16.15 | 17.05 | 1.13 | 81.80 |
| | 135 | 80.00 | 78.73 | 96.00 | 15.62 | 16.90 | 1.00 | 82.22 |
| | 150 | 77.50 | 77.12 | 96.00 | 17.44 | 17.72 | 1.00 | 80.42 |
| KSRC(GUASS) | 75 | 80.12 | 77.58 | 100.00 | 13.26 | 14.95 | 0 | 81.67 |
| | 90 | 80.15 | 80.81 | 99.88 | 13.23 | 12.79 | 0.031 | 83.07 |
| | 105 | 82.15 | 81.27 | 100.00 | 11.90 | 12.49 | 0 | 84.15 |
| | 120 | 82.77 | 83.50 | 100.00 | 11.49 | 11.00 | 0 | 85.38 |
| | 135 | 82.04 | 82.54 | 100.00 | 11.97 | 11.64 | 0 | 84.65 |
| | 150 | 83.69 | 83.31 | 100.00 | 10.87 | 11.13 | 0 | 85.70 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **KSRC(POLY)** | **75** | 82.23 | 82.85 | 99.50 | 11.85 | 11.44 | 0.13 | 84.80 |
| | **90** | 84.00 | 82.27 | 98.38 | 10.67 | 11.82 | 0.41 | 85.17 |
| | **105** | 84.19 | 83.27 | 98.50 | 10.54 | 11.15 | 0.38 | 85.70 |
| | **120** | 83.69 | 82.65 | 99.63 | 10.87 | 11.56 | 0.093 | 85.37 |
| | **135** | 84.04 | 84.19 | 99.38 | 10.64 | 10.54 | 0.16 | 86.15 |
| | **150** | 83.69 | 83.19 | 99.38 | 10.87 | 11.21 | 0.16 | 85.57 |
| **FDDL** | **75** | 79.85 | 80.88 | 99.62 | 13.44 | 12.74 | 0.093 | 82.93 |
| | **90** | 91.96 | 81.15 | 100 | 12.03 | 12.56 | 0 | 84.02 |
| | **105** | 82.73 | 83.42 | 100 | 11.51 | 11.05 | 0 | 85.33 |
| | **120** | 84.00 | 83.46 | 99.38 | 10.67 | 11.03 | 0.16 | 85.82 |
| | **135** | 84.15 | 83.31 | 100 | 10.56 | 11.13 | 0 | 85.90 |
| | **150** | 83.85 | 83.77 | 99.75 | 10.77 | 11.82 | 0.062 | 85.93 |

**Table 3:** The classification rates across various classification methods (%)

| **Classification Method** | | **Detection rates** | | | **False alarm rates** | | | **Classification rates** |
|---|---|---|---|---|---|---|---|---|
| | | **AAV** | **DW** | **Noise** | **AAV** | **DW** | **Noise** | |
| **SVM(Acoustic)** | **75** | 78.08 | 76.69 | 96.75 | 14.62 | 15.54 | 0.81 | 79.97 |
| | **90** | 80.23 | 79.31 | 96.37 | 13.18 | 13.79 | 0.91 | 81.98 |
| | **105** | 80.69 | 81.50 | 99.25 | 12.87 | 12.33 | 0.19 | 83.52 |
| | **120** | 82.88 | 82.92 | 97.12 | 11.41 | 11.38 | 0.72 | 84.80 |
| | **135** | 84.15 | 83.08 | 97.00 | 10.56 | 11.28 | 0.75 | 85.40 |
| | **150** | 83.69 | 83.50 | 96.75 | 10.87 | 11.00 | 0.81 | 85.35 |
| **SRC(Acoustic)** | **75** | 78.19 | 79.27 | 98.12 | 17.26 | 16.67 | 0.47 | 81.95 |
| | **90** | 77.65 | 79.38 | 98.00 | 17.46 | 16.36 | 0.50 | 81.73 |
| | **105** | 80.31 | 78.96 | 97.50 | 15.69 | 16.51 | 0.63 | 82.65 |
| | **120** | 79.65 | 78.27 | 95.50 | 16.15 | 17.05 | 1.13 | 81.80 |
| | **135** | 80.00 | 78.73 | 96.00 | 15.62 | 16.90 | 1.00 | 82.22 |
| | **150** | 77.50 | 77.12 | 96.00 | 17.44 | 17.72 | 1.00 | 80.42 |
| **KSRC(GUASS)** | **75** | 80.12 | 77.58 | 100.00 | 13.26 | 14.95 | 0 | 81.67 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **(Acoustic)** | **90** | 80.15 | 80.81 | 99.88 | 13.23 | 12.79 | 0.031 | 83.07 |
| | **105** | 82.15 | 81.27 | 100.00 | 11.90 | 12.49 | 0 | 84.15 |
| | **120** | 82.77 | 83.50 | 100.00 | 11.49 | 11.00 | 0 | 85.38 |
| | **135** | 82.04 | 82.54 | 100.00 | 11.97 | 11.64 | 0 | 84.65 |
| | **150** | 83.69 | 83.31 | 100.00 | 10.87 | 11.13 | 0 | 85.70 |
| **KSRC(POLY) (Acoustic)** | **75** | 82.23 | 82.85 | 99.50 | 11.85 | 11.44 | 0.13 | 84.80 |
| | **90** | 84.00 | 82.27 | 98.38 | 10.67 | 11.82 | 0.41 | 85.17 |
| | **105** | 84.19 | 83.27 | 98.50 | 10.54 | 11.15 | 0.38 | 85.70 |
| | **120** | 83.69 | 82.65 | 99.63 | 10.87 | 11.56 | 0.093 | 85.37 |
| | **135** | 84.04 | 84.19 | 99.38 | 10.64 | 10.54 | 0.16 | 86.15 |
| | **150** | 83.69 | 83.19 | 99.38 | 10.87 | 11.21 | 0.16 | 85.57 |
| **FDDL(Acoustic)** | **75** | 79.85 | 80.88 | 99.62 | 13.44 | 12.74 | 0.09 | 82.93 |
| | **90** | 91.96 | 81.15 | 100.00 | 12.03 | 12.56 | 0 | 84.02 |
| | **105** | 82.73 | 83.42 | 100.00 | 11.51 | 11.05 | 0 | 85.33 |
| | **120** | 84.00 | 83.46 | 99.38 | 10.67 | 11.03 | 0.16 | 85.82 |
| | **135** | 84.15 | 83.31 | 100.00 | 10.56 | 11.13 | 0 | 85.90 |
| | **150** | 83.85 | 83.77 | 99.75 | 10.77 | 11.82 | 0.062 | 85.93 |
| **MT-SRC** | **75** | 84.54 | 85.23 | 100 | 10.31 | 9.85 | 0 | 86.90 |
| | **90** | 85.04 | 84.65 | 100 | 9.97 | 10.23 | 0 | 86.87 |
| | **105** | 85.77 | 85.46 | 100 | 9.49 | 9.69 | 0 | 87.53 |
| | **120** | 85.85 | 85.96 | 100 | 9.44 | 9.36 | 0 | 87.78 |
| | **135** | 85.85 | 85.96 | 100 | 9.44 | 9.36 | 0 | 87.78 |
| | **150** | 85.65 | 86.85 | 100 | 9.56 | 8.77 | 0 | 88.08 |
| **The Proposed Method** | **75** | 84.42 | 85.31 | 100 | 10.38 | 9.79 | 0 | 86.88 |
| | **90** | 85.73 | 86.85 | 100 | 9.51 | 8.77 | 0 | 88.12 |
| | **105** | 87.23 | 87.19 | 100 | 8.51 | 8.54 | 0 | 88.92 |
| | **120** | 87.85 | 87.04 | 100 | 8.10 | 8.64 | 0 | 89.12 |
| | **135** | 89.35 | 88.54 | 100 | 7.10 | 7.64 | 0 | 90.42 |
| | **150** | 88.81 | 89.04 | 100 | 7.46 | 7.31 | 0 | 90.40 |

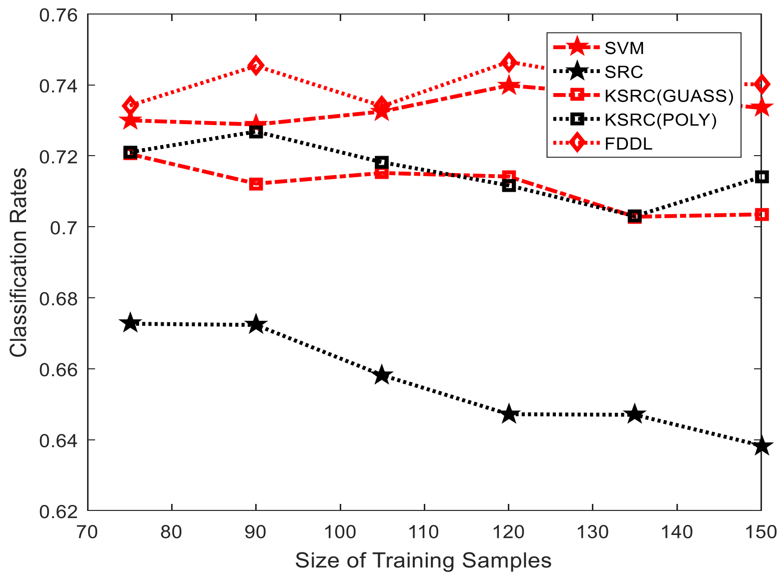



**Figure 5:** The trends of classification rates across various classification methods

Tab. 3 shows the specific value of detection rates, false alarm rates and classification rates across various classification methods. To further analyze the classification performance, we combine the results in Fig. 5 and Tab. 3, we can clearly see that the MT-SRC algorithm, which combines the feature of both acoustic and seismic signals, shows much higher classification rates (88.08%) than single acoustic or seismic signals for vehicle classification. In addition, compared to the FDDL algorithm, the MT-SRC algorithm also shows a considerable advantage. Therefore, it is significantly to study the target classification and recognition based on multi-task.

It is obviously observed that, from Tab. 3, our proposed approach shows an excellent ability of noise recognition, compared to the FDDL algorithm in which only acoustic or seismic data is used to perform the vehicle classification while conducting classification. Also, from the comparison result of the SRC and MT-SRC algorithms, we can see that the MT-SRC method makes full use of the advantages of the noise recognition rates under the seismic signal, and greatly improves the classification rates of both AAV and DW while making sure that the noise recognition rates are stable at around 100%. The classification rate of our proposed method based on the combination of acoustic and seismic signals (about 90%) is much better than that of the FDDL method under the acoustic signal (about 85%).

From Tab. 2 and Tab. 3, we note here that our proposed approach achieves a higher noise recognition rate in sparse noise environment, which makes fully use of the noise identification of the single signal while keeping the higher classification rate. In summary, our proposed method has made great progress in the classification of moving vehicles, which greatly improves the classification accuracy of single sensor classification algorithm.

Moreover, to show the superiority of our proposed model, we also study the SRC and MT-SRC algorithm in vehicle classification tasks. From the classification results in Tab. 3 and Fig.

5, we note that the performances of our proposed method are improved and algorithms are achieved higher classification rate of 90.40% superiorly than MT-SRC algorithm (88.08%), when the training sample is set to 150. These experiments validate the potential use of our proposed model for moving vehicles classification in acoustic and seismic sensor networks.

*5.3.2 Time complexity*

To further analyze the efficiency of our proposed method, we study the time complexity of this algorithm and compare it to other classification methods. In our proposed classification model scheme, the number of training samples are set 75, 90, 105, 120, 135, 150 respectively, much smaller than the dimension of the multi-dimensional frequency spectrum features (m=636). Fortunately, this matrix will not change in the iteration, and the inverse of it can be pre-computed.

Tab. 4 describes the running efficiency of various classification methods under single acoustic signal, seismic signal and also the combination of both acoustic and seismic signals.

**Table 4:** The running efficiency of various classification methods (s)

| **Classification** | **Size of training samples** | | | | | |
|---|---|---|---|---|---|---|
| **Method** | **75** | **90** | **105** | **120** | **135** | **150** |
| **SVM(Acoustic)** | 0.1272 | 0.1473 | 0.1739 | 0.1874 | 0.2117 | 0.2321 |
| **SRC(Acoustic)** | 22.97 | 35.17 | 50.93 | 70.75 | 94.78 | 124.2 |
| **FDDL(Acoustic)** | 17.83 | 23.10 | 29.68 | 36.11 | 43.45 | 52.14 |
| **KSRC(GUASS) (Acoustic)** | 92.63 | 95.39 | 99.02 | 102.3 | 106.1 | 110.4 |
| **KSRC(POLY) (Acoustic)** | 99.79 | 101.79 | 106.01 | 108.92 | 113.23 | 117.81 |
| **SVM(Seismic)** | 0.0297 | 0.0337 | 0.0380 | 0.0445 | 0.0490 | 0.0550 |
| **SRC(Seismic)** | 25.45 | 39.18 | 56.94 | 79.42 | 107.5 | 140.36 |
| **FDDL(Seismic)** | 20.22 | 26.13 | 33.31 | 38.87 | 49.29 | 58.16 |
| **MT-SRC** | 31.21 | 47.70 | 73.37 | 105.39 | 144.40 | 190.67 |
| **MT-FDDL** | 17.84 | 23.47 | 40.24 | 50.68 | 59.94 | 70.87 |

It is shown in Tab. 4 that the time complexity of the SVM algorithm under single acoustic signal as well as seismic signal is lower than that of the SRC algorithm, because the SRC algorithm needs to solve the problem of $\ell_p$-norm($p \leq 1$) sparse coding. Therefore, it is necessary to improve classification accuracy of the SRC method while reducing the time complexity of the algorithm. The time consumption of the KSRC method is very huge, but as the training sample increases, the time consumption remains essentially stable or only a slight growth. For the SRC method, the time consumption will increase as the training sample exponentially. Therefore, the KSRC method is more suitable for the classification task who has large training samples. From Tab. 4, we can clearly see that the FDDL algorithm achieves less time-consuming comparing to SRC algorithm under single acoustic and seismic signals, but still higher than the SVM algorithm.

Compared to the FDDL method, it can be seen that although the samples of our proposed method are the sum of the FDDL method under single acoustic and seismic signals, its time consumption is far less than the sum of the two in the process of moving vehicle classification. Moreover, our proposed method shows quite less time consumption comparing with the MT-SRC method. Therefore, we can conclude that our proposed method based on multi-feature fusion can greatly improve the classification rates of the vehicle classification while reducing the time complexity of the algorithm.

## 6 Conclusion

In this paper, we have addressed the moving vehicle classification problems in multi-sensor networks. Inspired by the current success of the discrimination dictionary learning approaches and multi-task joint covariate selection, we propose a novel multi-task joint sparse representation model based on fisher discrimination dictionary learning for vehicle classification to achieve more superior performance, where the data is collected from acoustic and seismic sensors. Our proposed method shows how to effectively exploit relationships among sensors measuring the same physical events. To further analyze the performance of our proposed approach, we extend it to handle the vehicle classification problems with sparse environmental noise and conduct comparison experiments with existing leading classification methods. Experimental results demonstrate that our method yields greatly accurate classification performance and noise recognition rate in terms of vehicle classification tasks.

**Acknowledgement:** This work was supported by National Natural Science Foundation of China (NSFC) under Grant No. 61771299, No. 61771322, No. 61375015, No. 61301027.

## References

**Chen, Y.; Nasrabadi, N.; Tran, T.** (2011): Sparse representation for target detection in hyperspectral imagery. *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 3, pp. 629-640.

**Cui, M.** (2015): Class-dependent sparse representation classifier for robust hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing,* vol. 53, no. 5, pp. 2683-2695.

**Duarte, M.; Hu, Y.** (2004): Vehicle classification in distributed sensor networks. *Journal of Parallel and Distributed Computing*, vol. 64, no. 7, pp. 826-838.

**Eom, K. B.** (1999): Analysis of acoustic signatures from moving vehicles using time-varying autoregressive models. *Multidimensional Systems and Signal Processing*, vol. 10, no. 4, pp. 357-378.

**Gao, S.; Tsang, I.; Chia, L.** (2010): Kernel sparse representation for image classification and face recognition. *Computer Vision-ECCV 2010 DBLP*, pp. 1-14.

**Gravina, R.; Alinia, P.; Ghasemzadeh, H.; Fortino, G.** (2017): Multi-sensor fusion in body sensor networks: State-of-the-art and research challenges. *Information Fusion*, vol. 35, pp. 68-80.

**Gu, S.; Zhang, L.; Zuo, W.; Feng, X.** (2014): Projective dictionary pair learning for pattern classification. *28th Annual Conference on Neural Information Processing Systems*, vol. 1, pp. 793-801.

**Jiang, Z.; Lin, Z.; Davis, L.** (2013): Label consistent K-SVD: Learning a discriminative dictionary for recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 35, pp. 2651-2664.

**Jiang, Z.; Zhe, L.; Davis, L.** (2011): Learning a discriminative dictionary for sparse coding via label consistent K-SVD. *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1697-1704.

**Kachach, R.; Cañas, J.** (2016): Hybrid three-dimensional and support vector machine approach for automatic vehicle tracking and classification using a single camera. *Journal of Electronic Imaging*, vol. 25, no. 3.

**Kang, D. H.; Kil, R. M.** (2015): Nonlinear filtering based on a network with gaussian kernel functions. *Neural Information Processing*, pp. 705-712.

**Kinnunen, T.; Saeidi, R.; Sedlak, F.; Kong, A. L.; Sandberg, J.** (2012): Low-variance multitaper MFCC features: a case study in robust speaker verification. *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 7, pp. 1990-2001.

**Klausner, A.; Tengg, A.; Rinner, B.** (2007): Vehicle classification on multi-sensor smart cameras using feature and decision fusion. *IEEE Conference on Distributed Smart Cameras*, pp. 67-74.

**Li, H.** (2016): Kernel polynomial representation for imaginary-time green's functions in continuous-time quantum monte carlo impurity solver. *Chinese Physics B*, vol. 25, no. 11, pp. 418-423.

**Lü, J.; Peng, Q.** (2014): The extraction of splice features based on fast fourier transform. *Beijing Ligong Daxue Xuebao/Transaction of Beijing Institute of Technology*, vol. 34, no. 2, pp. 207-210.

**Mairal, J.; Bach, F.; Ponce, J.** (2012): Task-driven dictionary learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 4, pp. 791-804.

**Mairal, J.; Bach, F.; Ponce, J.; Sapiro, G.; Zisserman, A.** (2008a): Supervised dictionary learning. *HAL-Inria*, pp. 1-8.

**Mairal, J.; Bach, F.; Ponce, J.; Sapiro, G.; Zisserman, A.** (2008b): Discriminative learned dictionaries for local image analysis. *Conference on Computer Vision and*

*Pattern Recognition*, pp. 1-8.

**Mei, X.; Ling, H.** (2011): Robust visual tracking and vehicle classification via sparse representation. *IEEE Transactions on Pattern analysis and Machine Intelligence*, vol. 33, no. 11, pp. 2259-2272.

**Nam, H.; Nasser, M.; Trac, D.** (2011): Robust multi-sensor classification via joint sparse representation. *Proceedings of the 14th International Conference on Information Fusion*, pp. 1-8.

**Nguyen, N.; Nasrabadi, N.; Tran, T.** (2011): Robust multi-sensor classification via joint sparse representation. *14th International Conference on Information Fusion*, pp. 1-8.

**Oh, T.; Chung, M.; Myung, H. (2017):** Accurate localization in urban environments using fault detection of GPS and multi-sensor fusion. *Robot Intelligence Technology and Applications 4*, vol. 447, pp. 43-53.

**Ramirez, I.; Sprechmann, P.; Sapiro, G.** (2010): Classification and clustering via dictionary learning with structured incoherence and shared features. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 23, no. 3, pp. 3501-3508.

**Sahidullah, M.; Saha, G.** (2013): A novel windowing technique for efficient computation of MFCC for speaker recognition. *IEEE Signal Processing Letters*, vol. 20, no. 2, pp. 149-152.

**Seokhyeon, J.; Yu, C.; Taekwang, J.** (2017): 21.6 A 12nW always-on acoustic sensing and object recognition microsystem using frequency-domain feature extraction and SVM classification. *IEEE International Solid-State Circuits Conference*, pp. 362-363.

**Shrivastava, A.; Patel, V. M.; Chellappa, R.** (2014): Multiple kernel learning for sparse representation-based classification. *IEEE Transactions on Image Processing*, vol. 23, no. 7, pp. 3013-3024.

**Tian, Y.; Dong, H. H.; Jia, L. M.; Wang, L. Y.; Li, S. Y.** (2014): Multi-sensor signature fusion algorithm for vehicle type classification. *Huanan Ligong Daxue Xuebao/Journal of South China University of Technology*, vol. 42, no. 3, pp. 52-58.

**Wang, R.; Guo, S.; Li, Y.; Zhang, Y.** (2017): Fisher discriminative dictionary learning for vehicle classification in acoustic sensor networks. *Journal of Signal Processing Systems*, vol. 86, no. 1, pp. 99-107.

**Wright, J.; Yang, A.; Ganesh, A.** (2009): Robust face recognition via sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210-227.

**Yan, Y.; Wang, H.; Xue, T.; Wang, X.** (2016): Optimal local sensor decision rule for target detection with channel fading statistics in multi-sensor networks. *Journal of the Franklin Institute*, vol. 354, no. 1, pp. 530-550.

**Yang, M.; Zhang, D.; Feng, X.** (2011): Fisher discrimination dictionary learning for sparse representation. *IEEE International Conference on Computer Visio*, pp. 543-550.

**Yuan, X.; Yan, S.** (2010): Visual classification with multi-task joint sparse representation. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 3493-3500.

**Zhang, B. C.; Lin, J. Q.; Chang, Z. C.; Yin, X. J.; Gao, Z.** (2016): The application of multi sensor data fusion based on the improved BP neural network algorithm. *Chinese Control and Decision Conference*, pp. 3842-3846.

**Zhang, H.; Nasrabadi, N. M.; Huang, T. S.; Zhang, Y.** (2011): Transient acoustic signal classification using joint sparse representation. *International Conference on Acoustics, Speech and Signal Processing*, pp. 2220-2223.