Vehicle recognition in acoustic sensor networks using multiple kernel sparse representation over learned dictionaries

Rui Wang¹, Wenming Cao² and Zhihai He³

Abstract
Sparse representation–based classification and kernel methods have emerged as important methods for pattern recognition. In this work, we study the problem of vehicle recognition using acoustic sensor networks in real-world applications. To improve the recognition accuracy with noise sensor data collected from challenging sensing environments, we develop a new method, called multiple kernel sparse representation–based classification, for vehicle recognition. In the proposed multiple kernel sparse representation–based classification method, acoustic features of vehicles are extracted and mapped into a high-dimensional feature space using a kernel function, which combines multiple kernels to obtain linearly separable samples. To improve the recognition accuracy, we incorporate dictionary learning method K-singular value decomposition into the multiple kernel sparse representation–based classification framework. The vehicle recognition from acoustic sensor network is then formulated into an optimization problem. Our extensive experimental results demonstrate that the proposed multiple kernel sparse representation–based classification method with learned dictionaries outperforms other existing methods in the literature on vehicle recognition from complex acoustic sensor network datasets.

Keywords
Multiple kernel sparse representation–based classification, K-singular value decomposition, vehicle recognition, sensor networks

Date received: 6 October 2016; accepted: 27 February 2017

Academic Editor: Pedro P Rodrigues

Introduction
Wireless sensor networks (WSNs) have gained much attention recently due to advances in micro-electromechanical systems (MEMS) technology which has facilitated the development of smart sensors.¹ WSNs consist of a large number of sensor nodes whose position need not be pre-determined.² The protocols and algorithms of WSNs have self-organization capabilities. At the same time, sensor nodes have the ability to carry out simple computations locally and transmit required data to nodes responsible for the fusion instead of sending raw data, which truly compress the power consumption of the networks. For the powerful self-organization and fault-tolerant capacity, a wide variety

¹School of Communication and Information Engineering, Shanghai University, Shanghai, China
²College of Information Engineering, Shenzhen University, Shenzhen, China
³Department of Electrical and Computer Engineering, University of Missouri, Columbia, MO, USA

Corresponding author:
Rui Wang, School of Communication and Information Engineering, Shanghai University, 333 Nanchen Road, Shanghai 200444, China.
Email: rwang@shu.edu.cn

Creative Commons CC-BY: This article is distributed under the terms of the Creative Commons Attribution 4.0 License (http://www.creativecommons.org/licenses/by/4.0/) which permits any use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (http://www.uk.sagepub.com/aboutus/openaccess.htm).
of applications are being envisioned for sensor networks, including habitat monitoring, health monitoring, battlefield surveillance, and target tracking.

Detection and classification of objects moving through the sensor field is an important task in many field applications. For vehicle recognition, early work focused on military vehicle detection for battlefield surveillance. Nowadays, vehicle detection has become an important task for traffic monitoring and management. Recognition of different vehicle types, such as cars, motorbikes, buses, or trucks, provides detailed traffic statistics and useful information about road utilization. Since many characteristics of the vehicle can be inferred from the sound it generates, it is feasible to recognize the type of the moving vehicle in acoustic sensor networks. However, how to improve the robustness of vehicle recognition algorithms from sensor networks within complex and noisy environments remains a critical and challenging issue in practice.

The problem of vehicle classification in acoustic sensor networks is essentially a pattern recognition problem. In recent years, various classification methods have been proposed in this field to adapt to different situations and improve the recognition rate, such as maximum likelihood (ML), k-nearest neighbor (k-NN), support vector machines (SVMs), and decision tree (DT) methods. Wright et al. found that sparse representation-based classification (SRC) performed well on face recognition, especially in noisy and cluttered environments. Mei and Ling proposed a robust visual tracking and vehicle classification approach using sparse representation and demonstrated its effectiveness on a vehicle tracking and classification task using outdoor infrared (IR) video sequences. Kernel methods are effective in machine learning to solve real-world problems with nonlinear data structures. Gao et al. applied kernel sparse representation–based classification (KSRC) to map the data into a high-dimensional space to image classification and face recognition and achieved state-of-the-art performance. However, the success of kernel methods often depends on the choice of an appropriate kernel and features. Specific kernel functions are proposed for particular applications, such as text documents categorizing and computational biology. So, it is critical to select the kernel function that fits the samples. Instead of selecting one specific kernel function, multiple kernel learning methods which learn the kernel from the samples by a linear combination of base kernels is more effective, especially for complex scenarios with different sources or modalities.

The sparse representation classification model above assumes a data sample can be represented as a sparse combination of several atoms from the pre-specified or non-adaptive dictionaries, which cannot represent a given class of signals efficiently. To address this issue, recent research has focused on designing dictionaries using learning methods. Engan et al. introduced the method of optimal directions (MOD) to find a dictionary and a sparse matrix which minimize the representation error. This method suffers from the relatively high complexity. To train a generic dictionary for sparse signal representation, Aharon et al. developed the K-singular value decomposition (K-SVD) algorithm which updated the dictionary atom-by-atom in a simple process rather than using a matrix inversion.

In this article, we try to combine the dictionary learning method with the sparse representation to solve the multi-sensor vehicle classification problem, focusing on recognizing different types of vehicles. The dataset contains acoustic recordings observed at each individual sensor in a real-world experiment carried out at the city of Twenty-Nine Palms, CA, in November 2001. The features of vehicles are extracted using Mel-frequency cepstral coefficients (MFCC), which has proven to be efficient in acoustic signal recognition. Chitra and Sumalatha used MFCC to extract the sound features of the emergency vehicles and performed the classification and identification task using SVM. This approach achieved increased accuracy and reduced time delay for the emergency. Matthias and Rainer presented a mobile sound classification system which extracted 13 MFCC from the data collected by the microphone and classified the sound using the neural networks to recognize sounds of emergency vehicles in road traffic.

In this work, we study whether this set of features can effectively be applied to vehicle recognition in transportation applications. Since the acoustic signals gathered in a real-world setting is inherently complex, it is difficult to choose the best kernel function. It is better to have a set of kernel functions and let the algorithm select the best subset of kernels. Therefore, we propose multiple kernel sparse representation–based classification (MKSRC), which combines several possible kernels as the kernel function and optimizes the multiple kernel weights while training the KSRC to adapt to different cases. Meanwhile, in contrast to previous approaches on sparse representation in which the dictionary is fixed by the training samples, in this article, we update the dictionary by K-SVD to adapt to complex scenes.

The major contribution of this article lies in the following aspects. First, we have developed a new classification algorithm based on MKSRC and successfully applied it to vehicle recognition from acoustic sensor networks. Second, we have developed a new and effective multi-kernel weight update scheme based on gradient descend, enabling our multi-kernel representation to fit different input source characteristics. This source-adaptive representation scheme has demonstrated its unique advantages by our experiments. Third, we have proposed a K-SVD method for dictionary update instead of using fixed dictionary obtained from the
training samples as in existing methods. This new method is able to handle the classification tasks within different and complex environments.

The remainder of this article is organized as follows. In section “Framework of vehicle recognition,” we present the framework of vehicle recognition. Section “Sparse representation models” explains the sparsity models and MKSRC. Our dictionary learning method is presented in section “Dictionary learning methods.” Experimental results and performance comparisons are provided in section “Experimental results.” Finally, conclusions and discussions on future work are provided in section “Conclusion.”

Framework of vehicle recognition

Figure 1 shows our vehicle recognition framework using MKSRC, which has the following major components:

1. **Pre-processing.** The raw acoustic signals of vehicles are gathered from an acoustic sensor network. This pre-processing step is important for noise reduction. We use the constant false alarm rate (CFAR) detection method. After the CFAR detection, useful event series are converted to frames and the default frame increment is half length of the frame as shown in Figure 1.

2. **Feature extraction.** An appropriate feature extraction method is important for classification. We use MFCC acoustic features. Acoustic signals often change quickly over time. Compared with linear prediction cepstral coefficients (LPCC), MFCC is more extensively used because of its robustness.

3. **Dictionary learning and classification.** We construct an over-complete dictionary based on the MFCC features and map it to a high-dimensional feature space using the kernel function which consists of multiple kernels \( \phi_1, \phi_2, \ldots, \phi_p \) with weights \( \alpha_1, \alpha_2, \ldots, \alpha_p \). To establish the best possible representations for each member in this set with sparsity constraints, we update the dictionary columns by K-SVD in an iterative manner. Then, the object recognition problem becomes determining the best sparse representation of the test sample using the learned dictionary which has the minimum representation error.

Sparse representation models

**SRC**

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

Suppose we have \( l \) classes of objects, and let \( D = [D_1, D_2, \ldots, D_l] \in \mathbb{R}^{m \times n} \) be a set of \( n \) training samples in \( l \) classes, where \( D_i = [d_{i,1}, d_{i,2}, \ldots, d_{i,n}] \in \mathbb{R}^{m \times n} \) contains \( m \)-dimensional features of training samples in the \( i \)th class. Suppose the test sample \( y \in \mathbb{R}^{m \times 1} \) can be sparsely represented by the over-complete dictionary \( D \), we can obtain the sparse vector \( x \) by solving the following optimization problem

\[
\min \|x\|_0 \quad \text{s.t.} \quad y = Dx
\]  

where \( \|x\|_0 \) is the \( l_0 \) norm which stands for the number of the non-zero coefficients of the matrix. While equation (1) is a NP-hard combinational optimization, Candes proved that \( l_0 \) norm can be substituted by \( l_1 \) norm in equation (2) as an approximate solution if the solution of equation (1) is sparse enough

\[
\min \|x\|_1 \quad \text{s.t.} \quad y = Dx
\]  

In fact, since real-world data are often noisy, it may not be possible to express the test sample exactly as a sparse combination of the training samples. The model in equation (1) can be rewritten as

\[
\min \|x\|_0 \quad \text{s.t.} \quad y = Dx + z
\]  

where \( z \in \mathbb{R}^{m \times 1} \) is a noise term with bounded energy \( \|z\|_2 < \epsilon \). Thus, the sparse presentation model can be modified as

\[
\min \|x\|_1 \quad \text{s.t.} \quad \|y - Dx\|_2 \leq \epsilon
\]  

and generally, model (4) is transformed to the optimization problem about \( J(x) \) as follows

\[
\min _{x} J(x) = \frac{1}{2} \|y - Dx\|_2^2 + \lambda \|x\|_1
\]  

where the first part is the residual and the parameter \( \lambda > 0 \) is a scalar regularization parameter that balances the sparsity of the solution and fidelity of the approximation to \( y \). The solution of equation (5) is already available, and we use orthogonal matching pursuit (OMP) to solve the sparse minimization problem. After sparsely coding \( y \) on \( D \) via \( l_1 \)-norm minimization, we obtain the solution denoted by \( \hat{x} \), and the classification result can be obtained by computing the residuals for each class

\[
r_i(y) = \|y - D\delta_i(\hat{x})\|_2
\]
A Mercer kernel is called a Mercer kernel if it satisfies the Mercer’s condition: continuous, symmetric, and positive semi-definite. To ensure an ideal classification performance, the vector $D_i$ in the dictionary should be uncorrelated. However, in practice, the underlying similarities between the extracted features of different vehicles may result in some correlation among $D_i$, and affect the final result. Similar to kernel methods in SVM, our kernel method is used to deal with the linearly non-separable problems using nonlinear mapping.

A kernel is a feature mapping function $\phi$ which maps the test sample $y$ and the dictionary $D$ from the input space $\chi$ to a high-dimensional kernel feature space $F$

$$y \rightarrow \phi(y)$$

$$D = [d_{1,1}, d_{1,2}, \ldots, d_{i,n}] \rightarrow$$

$$D' = [\phi(d_{i,1}), \phi(d_{i,2}), \ldots, \phi(d_{i,n})]$$

In SRC, the test sample can be sparsely represented by the training samples in the input space $\chi$. Similarly, we arrive at a kernel sparse–based representation of the test sample in the kernel feature space $F$

$$\min_x \|x\|_0 \quad \text{s.t.} \quad \phi(y) = D'x$$

Likewise, the optimization problem in equation (5) can be mapped in the high-dimensional space $F$ as follows

$$\min_x J(x) = \frac{1}{2} \|\phi(y) - D'x\|_2^2 + \lambda \|x\|_1$$

However, since the mapping function $\phi$ is unknown, the solution of KSRC is based on the selected kernel function. When the dictionary $D$ is fixed, equation (10) can be rewritten as

$$\min_x J(x) = \frac{1}{2} \|\phi(y) - D'x\|_2^2 + \lambda \|x\|_1$$

$$= \frac{1}{2} k(y, y) + \frac{1}{2} x^T K_{DD} x - x^T K_D y + \lambda \|x\|_1$$

$$= L(x) + \lambda \|x\|_1$$

where $L(x) = \frac{1}{2} k(y, y) + \frac{1}{2} x^T K_{DD} x - x^T K_D y$. $K_{DD}$ is a $k \times k$ matrix with $K_{DD} = k(D, D)$, and $K_D(x)$ is a

**Algorithm 1.** Sparse representation–based classification algorithm.

1. Normalize the columns of $D$ to have unit $l_2$-norm.
2. Code $y$ over $D$ via $l_1$-norm minimization
   $$x = \arg \min_\lambda \|y - Dx\|_2^2 + \lambda \|x\|_1$$
   where $\lambda$ is a positive scalar.
3. Compute the residuals $r_i = \|y - D_i x\|_2$
   $\hat{x}_i$ is the coefficient vector associated with class $i$.
4. Output the identity of $y$ as $\text{identity}(y) = \arg \min \{ r_i \}$

where $\delta_i(\hat{x}_i)$ selects only the non-zero coefficients belonging to class $i$. We then classify $y$ to the object class that minimizes the residual error.

**The kernel method**

To ensure an ideal classification performance, the vector $D_i$ in the dictionary should be uncorrelated. However, in practice, the underlying similarities between the extracted features of different vehicles may result in some correlation among $D_i$, and affect the final result. Similar to kernel methods in SVM, our kernel method is used to deal with the linearly non-separable problems using nonlinear mapping.

A kernel is called a Mercer kernel if it satisfies the Mercer’s condition: continuous, symmetric, and positive semi-definite. Suppose $x$ and $x'$ are two vectors in $\chi$ and $\phi$ is a nonlinear function realizing the mapping from the input space $\chi$ to the feature space $F$. Usually, a Mercer kernel $k$ can be expressed as

$$k(x, x') = \langle \phi(x), \phi(x') \rangle$$

where $\langle \cdot, \cdot \rangle$ denotes dot product. It transforms the dot product calculation in the high-dimensional feature space to the kernel function in the input space avoiding curse of dimensionality. Then, we can focus on the kernel function instead of $\phi$. The linear kernel, polynomial kernels, and Gaussian radial basis function (RBF) kernels are commonly used in kernel function design.
Algorithm 2. Multiple kernel sparse representation–based classification algorithm.

Input. The dictionary \( D \) composed by \( t \) training samples. A test sample \( y \). The number of kernel function \( p \).
1. Initialize the kernel weights \( \alpha_m = 1/\sqrt{p} \).
2. Compute the coefficient \( \bar{x} = \arg \min \left\{ \frac{1}{2} \| y - D\bar{x} \|^2 + \lambda \| \bar{x} \| \right\} \) with the kernel function \( k(a, b) = \sum_m \alpha_m^2 k_m(a, b) \).
3. Update the kernel weights by \( (\alpha_m^2)^{n+1} = (\alpha_m^2)^n - \epsilon \nabla_{\alpha_m} J \).
4. Go back to step 2 until the condition of convergence \( |\alpha_m^{n+1} - \alpha_m^n| \leq q \) is met.
5. Compute the residual \( r_j(y) = \frac{1}{2} \| y - D\delta_j(\bar{x}) \|_2^2 \), for \( j = 1, \ldots, l \).

Output. Identity \( (y) = \arg \min_j \{ r_j \} \)

\( k \times 1 \) vector with \( \{K_D(\bar{x})\}_i = k(D_i, \bar{x}) \). By selecting appropriate kernel functions, we can solve the optimization problem.

\[
\frac{\partial J(\alpha_m^2)}{\partial \alpha_m} = \frac{1}{2} k_m(y, y) + \frac{1}{2} \bar{x}^T k_m(d_{1, 1}, y) + \frac{1}{2} \bar{x}^T k_m(d_{1, 2}, y) + \cdots + \frac{1}{2} \bar{x}^T k_m(d_{1, n}, y) - \bar{x}^T \left[ \sum_m \alpha_m^2 k_m(d_{1, 1}, y) \right] + \lambda \| \bar{x} \|_1
\]

\( MKSRC \)

In this article, to determine the best kernel for our classification task, we propose MKSRC which assumes that sample \( a \) can be mapped into several different feature spaces by nonlinear mapping functions \( \phi_m(a) \), \( m = 1, 2, \ldots, p \) with different weights \( \alpha_m \) to achieve the best performance in classification. The corresponding kernel function for samples \( a \) and \( b \) is defined as

\[
k(a, b) = \sum_m \alpha_m^2 k_m(a, b)
\]

where \( k_m(a, b) = \phi_m(a)^T \phi_m(b) \) and the kernel weights have to satisfy \( \sum_m \alpha_m^2 = 1 \), \( \alpha_m \geq 0 \). Within this MKSRC framework, the optimization problem in equation (5) can be mapped into a high-dimensional feature space \( F \). The parts related to the kernel can be described as follows

\[
k(y, y) = \sum_m \alpha_m^2 k_m(y, y)
\]

\[
K_{DD} = \begin{bmatrix}
\sum_{m=1}^p \alpha_m^2 k_m(d_{1, 1}, d_{1, 1}) & \sum_{m=1}^p \alpha_m^2 k_m(d_{1, 1}, d_{1, 2}) & \cdots & \sum_{m=1}^p \alpha_m^2 k_m(d_{1, 1}, d_{1, n}) \\
\sum_{m=1}^p \alpha_m^2 k_m(d_{1, 2}, d_{1, 1}) & \sum_{m=1}^p \alpha_m^2 k_m(d_{1, 2}, d_{1, 2}) & \cdots & \sum_{m=1}^p \alpha_m^2 k_m(d_{1, 2}, d_{1, n}) \\
\vdots & \vdots & \ddots & \vdots \\
\sum_{m=1}^p \alpha_m^2 k_m(d_{1, n}, d_{1, 1}) & \sum_{m=1}^p \alpha_m^2 k_m(d_{1, n}, d_{1, 2}) & \cdots & \sum_{m=1}^p \alpha_m^2 k_m(d_{1, n}, d_{1, n})
\end{bmatrix}
\]

\[
K_D(y) = \begin{bmatrix}
\sum_{m=1}^p \alpha_m^2 k_m(d_{1, 1}, y) \\
\sum_{m=1}^p \alpha_m^2 k_m(d_{1, 2}, y) \\
\vdots \\
\sum_{m=1}^p \alpha_m^2 k_m(d_{1, n}, y)
\end{bmatrix}
\]

\[
k(x, x') = \langle \phi(x), \phi(x') \rangle
\]

Here, according to Lemma 2 in Chapelle et al., \( \alpha_m^2 \) it is possible to differentiate \( J \) with respect to \( \alpha_m^2 \) as if \( \bar{x} \) does not depend on \( \alpha_m^2 \) (\( \bar{x} \) is the vector \( x \) where the extreme value in \( J \) is attained.). We have

\[
\frac{\partial J}{\partial \alpha_m} = \frac{1}{2} \bar{x}^T \left[ \sum_m \alpha_m^2 k_m(d_{1, 1}, y) \right] - \bar{x}^T \left[ \sum_m \alpha_m^2 k_m(d_{1, 1}, y) \right] + \lambda \| \bar{x} \|_1
\]

Using the gradient descent method, we can update kernel weights \( (\alpha_m^2)^{n+1} = (\alpha_m^2)^n - \epsilon \nabla_{\alpha_m} J \) and solve the optimization problem. The following table summarizes the proposed MKSRC method.

Under the condition of convergence, \( q \) plays a critical role in determining the complexity of the algorithm and it should be adjusted based on the application requirements. Besides, \( \delta_j(\bar{x}) \) is a vector whose only non-zero entries are the entries in \( x \) that are associated with class \( j \). By computing the residual, we can classify the test sample \( y \) into the group which minimizes residual error.

Dictionary learning methods

The key component in sparse representation is to construct an over-complete dictionary. It is crucial to choose an appropriate dictionary. One can use pre-determined dictionaries, such as undecimated...
wavelets,\textsuperscript{29} steerable wavelets,\textsuperscript{30} and curvelets.\textsuperscript{31} Their major advantage is simplicity and low complexity. However, their performance largely depends on the specific characteristics of the target signal. In this article, to address this issue, we introduce a dictionary learning method to update the over-complete dictionary and represent the signals sparsely. Dictionary learning has been widely used recently in many signal processing applications, such as image compression enhancement\textsuperscript{32} and classification tasks.\textsuperscript{33}

To update an over-complete dictionary $D$ in equation (5), a typical dictionary learning algorithm performs a two-stage procedure iteratively: sparse coding and dictionary update. Sparse coding is the process of computing the representation coefficients $x$ based on the given signal and the present dictionary. The dictionary is updated in stage 2 to reduce the representation error.

The MOD introduced by Engan et al.\textsuperscript{17,34} is one of the first methods to implement sparsification.\textsuperscript{15} Like other learning methods, MOD alternates two steps, with a sparse coding stage that uses OMP followed by an update of the dictionary. The aim of MOD is to find a dictionary $D$ and a sparse matrix $x$ that minimize the representation error

$$\arg\min_{x,D} ||y-Dx||^2_F \quad \text{s.t.} \quad ||x||_0 \leq \varepsilon$$ (17)

where $\{x_i\}$ represents the columns of $x$ and the notation $||x||_F$ represents the Frobenius norm, defined as

$$||x||_F = \sqrt{\sum_{j}^{} x_{j}^2}$$

Suppose that the sparse matrix $x$ is fixed, and we can find the appropriate $D$ that minimizes the above error by taking the derivative of equation (17) with respect to $D$. To achieve $(y-Dx)x^T = 0$, we use the following iterative method

$$D^{(n+1)} = yx^{(n)}x^T (x^{(n)}x^{(n)}x^T)^{-1}$$ (18)

In this article, we propose to use K-SVD to update the dictionary which is more efficient than MOD. The K-SVD algorithm is based on the SVD process ($K$ is the number of columns in $D$) and the dictionary update is performed atom-by-atom in a simple and efficient process.\textsuperscript{15} Assuming that $D$ and $x$ are fixed, to update the $k$th column $d_k$ in $D$, the object function can be rewritten as

$$||y-Dx||^2_2 = ||y - \sum_{j=1}^m d_jx_j^T||^2_2 = ||y - \sum_{j=1}^m d_jx_j^T - d_kx_k^T||^2_2$$ (19)

where $x_j^T$ stands for the $j$th row in $x$. In the above expression, we aim to the update both $d_k$ and $x_k^T$ by a simple rank-1 approximation\textsuperscript{15} of the pre-computed error term

$$E_k = y - \sum_{j \neq k} d_jx_j^T$$ (20)

**Experimental results**

In this section, we evaluate the performance of the proposed method on the dataset collected from a real-world wireless sensor networks (WDSN) in the city of Twenty-Nine Palms, CA, in November 2001. This dataset is available at http://www.ecs.umass.edu/~mduarte/Software.html.\textsuperscript{2} It contains the acoustic, seismic, and IR information of two types of military vehicles, namely, Assault Amphibian Vehicle (AAV) and Dragon Wagon (DW). The original time series data are collected from 18 sensor nodes on 3 routes, as shown in Figure 2. Each node has three types of sensors: microphone, geophone, and polarized IR sensor. These sensors are able to cover a field of about 900 $\times$ 300 $m^2$, which consists of an east–west road, a south–north road, and an intersection. Each record in the dataset represents a vehicle passing by at a constant speed. Note that the Doppler effect will cause change in the frequencies of the measured signal. Similar to Duarte and Hu,\textsuperscript{2} we do not consider this Doppler effect since the relative speed between the moving vehicles and sensor nodes is stable and relatively slow.

**Feature extraction**

In this experiment, we aim to recognize each vehicle using the acoustic data. The acoustic data were recorded at a rate of 4960 Hz by microphones equipped
on sensor nodes. First, we choose the data collected from the 3rd to 11th runs (AAV3–AAV11 and DW3–DW11) as the data source to assess different feature extraction and classification methods. To detect the useful events in raw time series data, we use CFAR detection algorithm which is able to mark times with high energy values. Then, we use the MFCC method to extract features from the event time series for classification purposes.

Choosing an arbitrary set of the above data and comparing its 12-dimensional features extracted by MFCC with 50-dimensional features extracted by fast Fourier transform (FFT) which is calculated for every 512 point sample (every 103.2 ms for the current sample rate 4960 Hz), we can find the distinct difference between them in Figure 3(b) and (c) where the $x$-axes stand for the frame number and the $y$-axe is on behalf of the magnitude of each feature item. The MFCC features mainly concentrate in lower feature item compared with FFT ones.

**Dictionary learning**

For cross-validation, after feature extraction, we divide the acoustic features into two parts, one as test samples and the other as training samples. For vehicle recognition, we need to compute the sparse representation of the test samples using a specific dictionary. The initial dictionary consists of acoustic features of training samples. Then, to better fit the current dataset, we use the K-SVD approach to update the initial dictionary. Using the OMP algorithm, we can first get the corresponding sparse matrix of the test samples. The sparsity level, which stands for the number of non-zero
coefficients in the sparse matrix, will affect the recognition performance, as well as the computational complexity of the algorithm. Assumed that there are total 100 training samples in the dictionary, we demonstrate the sparse coding result of a test sample in Figure 4 with different sparsity levels \((K = 30 \text{ and } K = 60)\). Figure 5 shows the relationship between the sparsity and the time consumption of the sparse coding process with a single sample. The algorithm is written in MATLAB without optimization for speed, and ran on a laptop of Core i5 at 1.60 G with 4-GB memory. We can see that complexity increases significantly with the sparsity level.

To further study the impact of different dictionaries on the algorithm performance, we define the relative error of the sparse representation as

\[
\gamma = \frac{\|y - Dx\|_2}{\|y\|_2} \tag{21}
\]

Compared with the initial dictionary which has the original training samples, the dictionary updated by K-SVD (with eight times iteration) shows its significantly improved performance as we can see from Figure 6. The relative error decreases significantly with the sparsity level. This implies that, in practice, we need to choose the appropriate sparsity level and find a good tradeoff between the relative error and complexity.

**Classification methods**

To solve the vehicle recognition problem by the proposed MKSRC model, the features are mapped into a high-dimensional feature space with the kernel function. Here, we choose two common kernels, polynomial kernel (22), and Gaussian RBF kernel (23)

\[
k(a, b) = (a^T b + 1)^d \tag{22}
\]
\[ k(a, b) = \exp\left(\frac{-|a - b|^2}{2\sigma^2}\right) = \exp\left(-\beta |a - b|^2\right) \quad (23) \]

where \(d\) and \(\beta\) are the parameters related to the characteristics of acoustic features and they are tuned using cross validations. We compare our MKSRC method with existing classification methods SVM\(^2,36\) and SRC\(^10,37,38\).

In our experiments, there are 90 samples for each vehicle collected from 9 runs (3–11) of 10 sensor nodes (51–56 and 58–61), as well as 90 samples for the noise in the acquisition process. To validate the results of a classifier, we employ threefold cross validation with stratified partition of the samples. The classifier is trained three times and each time a different set is used as a validation set. Tables 1 and 2 present the detection, false alarm, and classification rates based on FFT and MFCC. Here, the detection rate is defined to be the ratio between the number of correct classification samples and the size of the class. The false alarm rate is defined to be the ratio between the number of incorrect classification samples and the total number of samples.
in other classes. Furthermore, to analyze the effect with different dictionaries, we list the classification result based on the dictionary updated by K-SVD in Table 3.

The kernel parameters, $d, \beta$, are tuned by cross validations. In this work, we have obtained $d = 2, \beta = 0.0002$. For feature extraction, the MFCC method has a distinct advantage over FFT since it considers human auditory characteristics. For the choice of dictionary, we find that our method using updated dictionary outperforms other classifiers based on sparse representation. From Table 3, we can see that the proposed MKSRC achieves a high classification rate of 90.00%, outperforming other methods while achieving lower false alarm rates.

To further illustrate the performance of MKSRC, we introduce the normalized correlation between the sparse codes of SRC, KSRC, and MKSRC. We list the results in two classes (each contains 30 samples) in Figure 7, where the $x$-axes and $y$-axes both stand for the number of samples used and the first 30 (1–30) and the rest (31–60) are from the same class, respectively. The entry $(p, q)$ is the normalized correlation of the sparse codes between the acoustic test samples $p$ and $q$.

According to the definition of correlation, the normalized correlation of the sparse codes should be block-wise since the sparse codes belonging to the same class are more similar. In Figure 7, we find that MKSRC becomes more discriminative sparse codes since the correlation coefficients of the same class (the first 30 samples belong to AAV and the rest belong to DW) in MKSRC are generally higher than in SRC and KSRC, which facilitates the better classification performance.

**Conclusion**

In this work, we have studied the problem of vehicle recognition using acoustic sensor networks. We have
developed a new method, called MKSRC, for vehicle recognition. Acoustic features of vehicles are extracted and mapped into a high-dimensional feature space using the kernel function, which combines multiple kernels to obtain linearly separable samples. To improve the recognition accuracy, we incorporate dictionary learning into the MKSRC framework. By calculating the reconstructing error and updating the kernel weights, the objective vehicles will be recognized by solving the optimization problem. Our extensive experimental results demonstrate that the proposed MKSRC method with learned dictionaries outperforms other existing methods based on SVM, SRC, and KSRC in the literature on vehicle recognition from complex acoustic sensor network datasets. In our future work, we will focus on the self-adaption of the kernel parameters to further improve the recognition efficiency and robustness.

Declaration of conflicting interests
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding
The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the National Natural Science Foundation of China (NSFC) under grant nos 61301027 and 61375015, and the Zhejiang Provincial Natural Science Foundation of China (NSFC) under grant nos 61375015 and 61301027.

References
1. Yick J, Mukherjee B and Ghosal D. Wireless sensor network survey. Comput Netw 2008; 52(12): 2292–2330.
2. Duarte MF and Hu YH. Vehicle classification in distributed sensor networks. J Parallel Distir Com 2004; 64(7): 826–838.
3. D’Costa A and Sayeed AM. Collaborative signal processing for distributed classification in sensor networks. In: IPSN’03 proceedings of the 2nd international conference on information processing in sensor network, Palo Alto, CA, 22–23 April 2003, pp.193–208. Berlin: Springer.
4. Eom KB. Analysis of acoustic signatures from moving vehicles using time-varying auto regressive models. Multidim Syst Sign P 1999; 10(4): 357–378.
5. Thomas DW and Wilkins BR. The analysis of vehicle sounds for recognition. Pattern Recogn 1972; 4(4): 379–389.
6. Sokolov RT and Rogers JC. Removing harmonic signal nonstationarity by dynamic resampling. In: Proceedings of the IEEE international symposium on industrial electronics, Athens, 10–14 July 1995, vol. 1, pp.303–308. New York: IEEE.
7. Choe HC, Karlsen RE, Gerhart GR, et al. Wavelet-based ground vehicle recognition using acoustic signals. Proc SPIE 1996; 2762: 434–445.
8. Li D, Wong KD and Hu YH. Detection, classification and tracking of targets in distributed sensor networks. IEEE Signal Proc Mag 2002; 19(2): 17–29.
9. Ghasemzadeh H and Jafari R. Physical movement monitoring using body sensor networks: a phonological approach to construct spatial decision trees. IEEE T Ind Inform 2011; 7(1): 66–77.
10. Wright J, Yang AY and Ganesh A. Robust face recognition via sparse representation. IEEE T Pattern Anal 2009; 31(2): 210–227.
11. Mei X and Ling HB. Robust visual tracking and vehicle classification via sparse representation. IEEE T Pattern Anal 2011; 33(11): 2259–2272.
12. Gao S, Tsang IWH and Chia LT. Kernel sparse representation for image classification and face recognition. In ECCV’10 proceedings of the 11th European conference on computer vision: Part IV, Heraklion, 5–11 September 2010, vol. 6314, pp.1–14. Berlin: Springer.
13. Huma L, Craig S and John ST. Text classification using string kernels. J Mach Learn Res 2002; 2: 419–444.
14. Schölkopf B, Tsuda K and Vert J. Support vector machine applications in computational biology. In: Schölkopf B, Tsuda K and Vert J-P (eds) Kernel methods in computational biology. Cambridge, MA: The MIT Press, 2004, pp.71–92.
15. Rubinstein R, Bruckstein AM and Elad M. Dictionaries for sparse representation modeling. Proc IEEE 2010; 98(6): 1045–1057.
16. Tösic I and Frossard P. Dictionary learning. IEEE Signal Proc Mag 2011; 28(2): 27–38.
17. Engan K, Aase SO and Husoy JH. Method of optimal directions for frame design. In 1999 IEEE international conference on acoustic, speech, and signal processing, Phoenix, AZ, 15–19 March 1999, vol. 5, pp.2443–2446. New York: IEEE.
18. Aharon M, Elad M and Bruckstein AM. The K-SVD: an algorithm for designing of overcomplete dictionaries for sparse representation. IEEE Trans Signal Process 2006; 54(11): 4311–4322.
19. Chitra VPJ and Sumalatha MR. SVM-instance based approach to improve QoS parameters for time critical applications in WSN. In: 2012 fourth international conference on advanced computing (ICoAC), Chennai, India, 13–15 December 2012, pp.1–6. New York: IEEE.
20. Matthias M and Rainer B. Smartphone application for automatic classification of environmental sound. In: 2013 proceedings of the 20th international conference on mixed design of integrated circuits and systems (MIXDES), Gdynia, 20–22 June 2013, pp.512–515. New York: IEEE.
21. Gönen M and Alpaydın E. Multiple kernel learning algorithms. J Mach Learn Res 2011; 12: 2211–2268.
22. Gray AH Jr and Markel JD. Distance measures for speech processing. IEEE Trans Acoustics Speech Signal Process 1976; 24(5): 380–391.
23. Hasan MR, Jamil M, Rabbani MG, et al. Speaker identification using Mel frequency cepstral coefficients. In: 3rd international conference on electrical & computer engineering, Dhaka, Bangladesh, 28–30 December 2004, pp.565–570.
24. Candès EJ, Romberg J and Tao T. Robust Uncertainty Principles: exact Signal Reconstruction from Highly Incomplete Frequency Information. *IEEE T Inform Theory* 2004; 52(2): 489–509.

25. Tropp JA and Gilbert AC. Signal recovery from random measurements via orthogonal matching pursuit. *IEEE T Inform Theory* 2007; 53(12): 4655–4666.

26. Burges CJC. A tutorial on support vector machines for pattern recognition. *Data Min Knowl Disc* 1998; 2(2): 121–167.

27. Zhang L, Zhou WD, Chang PC, et al. Kernel sparse representation-based classifier. *IEEE T Signal Process* 2012; 60(4): 1684–1695.

28. Chapelle O, Vapnik V, Bousquet O, et al. Choosing multiple parameters for support vector machines. *Mach Learn* 2002; 46(1–3): 131–159.

29. Starck JL, Fadili MJ and Murtagh F. The undecimated wavelet decomposition and its reconstruction. *IEEE Trans Image Process* 2007; 16(2): 297–309.

30. Portilla J, Strela V, Wainwright MJ, et al. Image denoising using scale mixtures of Gaussians in the wavelet domain. *IEEE Trans Image Process* 2003; 12(11): 1338–1351.

31. Starck JL, Candès EJ and Donoho DL. The curvelet transform for image denoising. *IEEE Trans Image Process* 2002; 11(6): 670–684.

32. Elad M. Sparse and redundant representations: from theory to applications in signal and image processing. Berlin, Germany: Springer, 2010.

33. Mairal J, Bach F and Ponce J. Task-driven dictionary learning. *IEEE Trans Pattern Anal Mach Intell* 2012; 34(4): 791–804.

34. Engan K, Rao B and Kreutz-Delgado K. Frame design using FOCUSS with method of optimized directions (MOD). In: *Proceedings of the NORSIG’99*, Oslo, September 1999, pp.65–69, http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.719.8961&rep=rep1&type=pdf

35. Golub GH and Van Loan CF. *Matrix computations*, 3rd ed. Baltimore, MD: Johns Hopkins University Press, 1996.

36. Zhou CJ and Chen YZ. The research of vehicle classification using SVM and KNN in a ramp. In: *International forum on computer science-technology and applications*, USA, 25–27 December 2009, vol. 3, pp.25–27. New York: IEEE.

37. He R, Zheng WS and Hu BG. Two-stage nonnegative sparse representation for large-scale face recognition. *IEEE Trans Neural Netw Learn Syst* 2013; 24(1): 35–46.

38. Wang KY, Wang R, Feng YT, et al. Vehicle recognition in acoustic sensor networks via sparse representation. In: *2014 IEEE international conference on multimedia and expo workshops (ICMEW)*, Chengdu, China, 14–18 July 2014, pp.1–4. New York: IEEE.