MULTI-SENSOR JOINT KERNEL SPARSE REPRESENTATION FOR PERSONNEL DETECTION

Nam H. Nguyen, Nasser M. Nasrabadi and Trac D. Tran

Johns Hopkins University and U.S. Army Research Laboratory

ABSTRACT

In this paper, we propose a novel nonlinear technique for multi-sensor classification, which relies on sparsely representing a test sample in terms of all the training samples in a feature space induced by a kernel function. Our approach simultaneously takes into account the correlations as well as the complementary information between the homogeneous/heterogeneous sensors, while also considering the joint sparsity within each sensor's multiple observations in the feature space. This approach can be seen as a generalized model for representing a multi-task and multivariate Lasso in the feature space, where the data from all the sensors representing the same physical events are jointly represented by a sparse linear combination of the training data. Extensive experiments are conducted on real data sets and the results are compared with the conventional discriminative classifiers to verify the effectiveness of the proposed method in the application of automatic border patrol, where it is required to discriminate between human and animal footsteps.

1. INTRODUCTION

Multi-sensor fusion have received considerable amount of attentions over the past few years for both military and non-military tasks (e.g. [1]). A particular interest in multi-sensor fusion is classification, where the ultimate question is how to take advantage of related information from different sources (tasks) measured from the same physical event in order to achieve an improvement in the classification performance. A variety of approaches have been proposed in the literature to answer this question (e.g. [2], [3]). These methods mostly fall into two categories: decision in - decision out (DI-DO) and feature in - feature out (FI-FO) [1]. In [2], the authors investigated the DI-DO method on vehicle classification problem using data collected from acoustic and seismic sensors. They proposed to perform local classification (decision) for each sensor signal by a conventional method such as the Support Vector Machine (SVM). These local decisions are then incorporated via a Maximum A Posterior (MAP) estimator to make the final classification decision, thus named the DI-DO method. In [3], the FI-FO method is studied for vehicle classification using both visual and acoustic sensors. They proposed a method to extract temporal gait patterns from both sensor signals in order to use them as inputs to an SVM classifier.

In this paper, we study a non-linear multi-sensor classification problem in a kernel induced feature space, where we aim to perform discrimination between human and non-human footsteps. Our experimental setup is as follows: the footstep events on the field are recorded by a set of four sensors consisting of two acoustic and two seismic sensors simultaneously. Our ultimate goal is to detect whether the events involve human or human and animal footsteps. We investigate a sparse signal representation technique which has been successfully employed in a large number of discriminative applications (e.g. [4], [5]). We propose in this paper a multivariate kernel sparse representation for classification (MV-kerSRC) which extends the conventional multivariate sparse representation method (MV-SRC) to the kernel induced feature space. Various extensions of the MV-kerSRC are also proposed in this paper to effectively incorporate information from different sensors, including multitask MV-kerSRC (MTMV-kerSRC) and MV-kerSRC with composite kernels [6]. The MTMV-kerSRC imposes joint-sparsity constraints both within each task (for multiple observations) and across multiple tasks, while the MV-kerSRC with composite kernels efficiently combines kernels dedicated for each sensor via stack or summation operations as shown in Section 3.2.

The remainder of this paper is organized as follows. Section 2 introduces our proposed MV-KerSRC method. We present in Section 3 various extensions of the MV-kerSRC, which include MTMV-kerSRC and MV-kerSRC with composite kernels. Extensive experiments are shown in Section 4 and conclusions are drawn in Section 5.

2. KERNEL SPARSE REPRESENTATION

a) Sparse representation: Assume we are given a dictionary representing C distinct classes $\boldsymbol{X} = [\boldsymbol{X}_1, \boldsymbol{X}_2, ..., \boldsymbol{X}_C] \in \mathbb{R}^{n \times p}$, where the j-th class sub-dictionary $\boldsymbol{X}_j$ has p_j training samples $\{\boldsymbol{x}_{j,k}\}_{k=1,...,p_j}$. To label a test sample, it is often assumed $\boldsymbol{y} \in \mathbb{R}^n$ can be represented by a subset of the

This work has been supported in part by the Army Research Office (ARO) under Grant 60291-MA and National Science Foundation (NSF) under Grant CCF-1117545.

training samples in $\boldsymbol{X}$. Mathematically, $\boldsymbol{y}$ is written as $\boldsymbol{y} = [\boldsymbol{X}_1, \boldsymbol{X}_2, ..., \boldsymbol{X}_C]\boldsymbol{w} = \boldsymbol{X}\boldsymbol{w}$ where $\boldsymbol{w} \in \mathbb{R}^p$ is the unknown coefficient vector. The above assumption implies that only the entries of $\boldsymbol{w}$ that are associated with the class of the test sample $\boldsymbol{y}$ are non-zeros, and thus, $\boldsymbol{w}$ is the sparse vector. Taking this prior into account, many methods have been proposed to find the coefficient vector w efficiently such as ℓ_1-norm minimization [7].

b) Kernel sparse representation: Sparse representation has been widely known as an efficient method for classification when the above assumption is valid. However, in various practical applications, this assumption might not hold due to the complex structure of the dataset. In this paper, we show empirically that kernel methods can be a solution to overcome this issue. In fact, classifiers such as SVM have been extensively validated to perform better in the kernel domain. The reason is that if the classes in the dataset are not linearly separable, then the kernel methods can be used to project the data onto a feature space, in which the classes become linearly separable [8]. In this paper, we extend the linear sparse representation to the kernel domain. Denote $\kappa : \mathbb{R}^n \times \mathbb{R}^n \mapsto \mathbb{R}$ as the kernel function, defined as the inner product $\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) = \langle \phi(\boldsymbol{x}_i), \phi(\boldsymbol{x}_j) \rangle$, where $\phi : \boldsymbol{x} \mapsto \phi(\boldsymbol{x})$ is an implicit mapping that maps the vector $\boldsymbol{x}$ onto a higher dimensional space, possibly infinite. A commonly used kernel is the Radial Basis Function (RBF) kernel $\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp(-\|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2^2 / \sigma^2)$ with σ used to control the width of the RBF [8].

Let $\boldsymbol{y} \in \mathbb{R}^n$ be an unlabeled test sample of interest and $\boldsymbol{\phi}(\boldsymbol{y})$ be its representation in the feature space. The kernel sparse representation of the test sample $\boldsymbol{y}$ in terms of the training samples $\{\boldsymbol{x}_i\}_{i=1}^p$ can be formulated as

$$\boldsymbol{\phi}(\boldsymbol{y}) = [\boldsymbol{\phi}(\boldsymbol{X}_1), \boldsymbol{\phi}(\boldsymbol{X}_2), ..., \boldsymbol{\phi}(\boldsymbol{X}_C)]\, \boldsymbol{w} = \boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{w}, \quad (1)$$

where columns of the matrix $\boldsymbol{\Phi}(\boldsymbol{X})$ are the representations of the training samples in the feature space and the vector $\boldsymbol{w} \in \mathbb{R}^p$ is assumed to be sparse. This equation implies that in the feature space, the test sample $\boldsymbol{y}$ can be represented as the sparse linear combination of the training samples.

c) Multivariate kernel sparse representation (MV-kerSRC): Joint sparse representation [4] can also be naturally extended to the feature space. Let $\boldsymbol{Y} = [\boldsymbol{y}_1, \boldsymbol{y}_2, ..., \boldsymbol{y}_d] \in \mathbb{R}^{n \times d}$ be a set of d observations representing multiple measurements obtained by a sensor and $\boldsymbol{\Phi}(\boldsymbol{Y}) = [\boldsymbol{\phi}(\boldsymbol{y}_1), \boldsymbol{\phi}(\boldsymbol{y}_2), ..., \boldsymbol{\phi}(\boldsymbol{y}_d)]$ be their representation in the feature space. We have

$$\boldsymbol{\Phi}(\boldsymbol{Y}) = [\boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{w}_1, \boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{w}_2, ..., \boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{w}_d] = \boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{W}, \quad (2)$$

where $\boldsymbol{W} \in \mathbb{R}^{p \times d}$ is the unknown coefficient matrix. In the joint kernel sparse representation, the sparse coefficient vectors $\{\boldsymbol{w}_i\}_{i=1,...,d}$ share the same support. This row-sparse coefficient matrix $\boldsymbol{W}$ can be recovered by the following optimization, which we call multivariate kernel Lasso (MV-kerLasso)

$$\min_{\boldsymbol{W}} \|\boldsymbol{\Phi}(\boldsymbol{Y}) - \boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{W}\|_F^2 + \lambda \|\boldsymbol{W}\|_{1,q}, \quad (3)$$

where the ℓ_1/ℓ_q-norm imposed on the matrix $\boldsymbol{W}$ promotes the shared sparse pattern across multiple columns of $\boldsymbol{W}$: $\|\boldsymbol{W}\|_{1,q} \triangleq \sum_{i=1}^p \|\boldsymbol{w}_i\|_q$ where $\boldsymbol{w}_i$'s are rows of $\boldsymbol{W}$. The optimization (3) is implicitly solved in the feature space, which can be efficiently done using the kernel trick. That means, we do not need to explicitly express the data in the feature space, rather we only evaluate the kernel functions at the training points. In fact, by reformulating (3) and replacing each dot product by a kernel function, we can write (3) as

$$\min_{\boldsymbol{W}} \left(\mathrm{tr}(\boldsymbol{W}^T \boldsymbol{K}_{\boldsymbol{XX}} \boldsymbol{W}) - 2\, \mathrm{tr}(\boldsymbol{K}_{\boldsymbol{XY}} \boldsymbol{W}) \right) + \lambda \|\boldsymbol{W}\|_{1,q}, \quad (4)$$

where $\boldsymbol{K}_{\boldsymbol{XX}} \in \mathbb{R}^{p \times p}$ is the kernel matrix whose (i, j) entry is $\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j)$, and $\boldsymbol{k}_{\boldsymbol{XY}} \in \mathbb{R}^p$ is the matrix whose i-th entry is $\kappa(\boldsymbol{x}_i, \boldsymbol{y}_j)$. The above optimization is a quadratic convex programming due to the positive definiteness of the kernel matrix $\boldsymbol{K}_{\boldsymbol{XX}}$, thus it can be optimized via several existing algorithms in the literature (e.g. [9]).

Once the matrix $\widehat{\boldsymbol{W}}$ is obtained, the class label of the test sample $\boldsymbol{Y}$ is determined by computing the error residual between the test sample and it approximation from each class in the feature space. The label is given to the class with smallest residual

$$r_c(\boldsymbol{Y}) = \left\| \boldsymbol{\Phi}(\boldsymbol{Y}) - \boldsymbol{\Phi}(\boldsymbol{X}_c)\widehat{\boldsymbol{W}}_c \right\|_F$$
$$= \left(\mathrm{tr}(\boldsymbol{K}_{\boldsymbol{YY}}) - 2\, \mathrm{tr}(\widehat{\boldsymbol{W}}_c^T \boldsymbol{K}_{\boldsymbol{X}_c\boldsymbol{Y}}) + \mathrm{tr}(\widehat{\boldsymbol{W}}_c^T \boldsymbol{K}_{\boldsymbol{X}_c\boldsymbol{X}_c} \widehat{\boldsymbol{W}}_c) \right)^{1/2},$$

where $\boldsymbol{W}_c$ is the coefficient matrix of $\boldsymbol{W}$ associated with c-th class.

3. EXPLOIT INFORMATION FROM DIFFERENT SENSORS IN KERNEL SPARSE REPRESENTATION

In this section, we briefly describe our experimental setup and explain why we need to do more than just joint kernel sparse representation. Our database is from the Army Research Laboratory which conducted footstep data collection using four sensors including two acoustic and two seismic sensors. The goal is to discriminate between human and human-animal footsteps. In our experimental setup, we divide the signal (either acoustic or seismic signal) into ten overlapping segments (multiple observations) to capture local signal information. A natural question arises as whether we can improve the classification performance by incorporating information from all the sensors simultaneously. In this section, we give an affirmative answer by proposing two methods to exploit the complementary information between the different sensors. We experimentally show that the methods that exploit the coupling information from all the sensors potentially improve the classification accuracy.

Throughout this section, we use similar notations to define test samples and its dictionaries. Let $\boldsymbol{Y}^a \in \mathbb{R}^{n_a \times d}$ and $\boldsymbol{Y}^s \in \mathbb{R}^{n_s \times d}$ represent the test samples associated with the acoustic and seismic signals, respectively. Each $\boldsymbol{Y}^a$ or $\boldsymbol{Y}^s$ consists of d test segments. In addition, let $\boldsymbol{X}^a = [\boldsymbol{X}^a_1, ..., \boldsymbol{X}^a_C] \in \mathbb{R}^{n_a \times p}$ and $\boldsymbol{X}^s = [\boldsymbol{X}^s_1, ..., \boldsymbol{X}^s_C] \in \mathbb{R}^{n_s \times p}$ be the training dictionaries associated with acoustic and seismic signals, respectively. We also note that the segment dimensions n_a and n_s could be different to emphasize the flexibility of our approach.

3.1. Kernel sparse representation via multi-task multivariate kernel Lasso

As in the joint kernel sparse representation, the test samples from each sensor can be represented by their training samples

$$\boldsymbol{\Phi}(\boldsymbol{Y}^a) = \boldsymbol{\Phi}(\boldsymbol{X}^a)\boldsymbol{W}^a \qquad \text{and} \qquad \boldsymbol{\Phi}(\boldsymbol{Y}^s) = \boldsymbol{\Phi}(\boldsymbol{X}^s)\boldsymbol{W}^s,$$

where $\boldsymbol{W}^a$ and $\boldsymbol{W}^s$ are row-sparse coefficient matrices associated with the acoustic and seismic signals, respectively. We recall that the coefficient matrix can be seen as the discriminative feature for classification. Thus, by incorporating information from both sensors, we propose to solve $\boldsymbol{W}^a$ and $\boldsymbol{W}^s$ simultaneously via the following convex optimization

$$\min_{\boldsymbol{W}^a, \boldsymbol{W}^s} \Big(\|\boldsymbol{\Phi}(\boldsymbol{Y}^a) - \boldsymbol{\Phi}(\boldsymbol{X}^a)\boldsymbol{W}^a\|_F^2 + \|\boldsymbol{\Phi}(\boldsymbol{Y}^s) - \boldsymbol{\Phi}(\boldsymbol{X}^s)\boldsymbol{W}^s\|_F^2 \Big) + \lambda \|[\boldsymbol{W}^a \,, \boldsymbol{W}^s]\|_{1,q} \,. \tag{5}$$

It is clear from (5) that the information from both sensors are integrated into the sparse classification via the shared sparsity pattern of the matrices $\boldsymbol{W}^a$ and $\boldsymbol{W}^s$. This optimization is called multi-task multivariate kernel Lasso (MTMV-KerLasso). It is worth noticing that though in this section we investigate the multi-task multivariate kernel sparse representation for classification (MTMV-KerSRC) with only two different sensors, there is of course nothing preventing us from generalizing the problem to multiple sensors. Generally, if there are D sensors recording the same event, we will incorporate sensors' information by solving

$$\min_{\boldsymbol{W}'} \sum_{i=1}^{D} \left\|\boldsymbol{\Phi}(\boldsymbol{Y}^i) - \boldsymbol{\Phi}(\boldsymbol{X}^i)\boldsymbol{W}^i\right\|_F^2 + \lambda \|\boldsymbol{W}'\|_{1,q} \,, \tag{6}$$

where $\boldsymbol{Y}^i$ is the test sample from the i-th sensor and $\boldsymbol{X}^i$ is the i-th dictionary associated with the i-th sensor; and $\boldsymbol{W}' \in \mathbb{R}^{p \times Dd}$ is the big matrix consisting of matrices $\{\boldsymbol{W}^i\}_{i=1,...,D}$: $\boldsymbol{W}' = [\boldsymbol{W}^1, ..., \boldsymbol{W}^D]$. Again, (6) can be reformulated to account for kernel matrices

$$\min_{\boldsymbol{W}'} \sum_{i=1}^{D} \Big(\mathrm{tr}(\boldsymbol{W}^{i^T} \boldsymbol{K}_{\boldsymbol{X}^i \boldsymbol{X}^i} \boldsymbol{W}^i) - 2\,\mathrm{tr}(\boldsymbol{K}_{\boldsymbol{X}^i \boldsymbol{Y}^i} \boldsymbol{W}^i) \Big) + \lambda \|\boldsymbol{W}'\|_{1,q} \,. \tag{7}$$

This optimization can be efficiently solved via the alternating direction method. More detailed description of the algorithm can be found in [10].

3.2. Kernel sparse representation via composite kernels

Another way to integrate information between different sensors is through a composite kernel [6], which efficiently combines kernels dedicated for each sensor. In this paper, we propose to combine kernels dedicated for acoustic and seismic sensors in an efficient manner and use the new kernel for our multivariate kernel Lasso optimization in (4). Let $\boldsymbol{x}^a_i$ and $\boldsymbol{x}^s_i$ be segment samples extracted from acoustic and seismic sensor signals, respectively. Various kernel combinations include

Stacking kernel: Denote $\boldsymbol{x}_i$ as a new segment sample obtained by stacking two corresponding samples $\boldsymbol{x}^a_i$ and $\boldsymbol{x}^s_i$ in the original domain to obtain a longer vector: $\boldsymbol{x}_i = [\boldsymbol{x}^a_i; \boldsymbol{x}^s_i]$. The stacking kernel is defined as $\kappa(\boldsymbol{x}_i, \boldsymbol{x}_i) = \langle \boldsymbol{\phi}(\boldsymbol{x}_i), \boldsymbol{\phi}(\boldsymbol{x}_i) \rangle$.

Weighted summation kernel: By concatenating the two corresponding samples $\boldsymbol{x}^a_i$ and $\boldsymbol{x}^s_i$ in the feature space as $\boldsymbol{\phi}(\boldsymbol{x}_i) = [\beta_1 \boldsymbol{\phi}(\boldsymbol{x}^a_i); \beta_2 \boldsymbol{\phi}(\boldsymbol{x}^s_i)]$, we have the new kernel

$$\kappa(\boldsymbol{x}_i, \boldsymbol{x}_i) = \langle \boldsymbol{\phi}(\boldsymbol{x}_i), \boldsymbol{\phi}(\boldsymbol{x}_i) \rangle = \alpha_1 \kappa(\boldsymbol{x}^a_i, \boldsymbol{x}^a_i) + \alpha_2 \kappa(\boldsymbol{x}^s_i, \boldsymbol{x}^s_i),$$

where β_1 and β_2 are positive real-valued parameters, which are freely defined by user to control the weights associated with each kernel.

4. EXPERIMENTAL RESULTS

We use a set of four sensors consisting of two acoustic and two seismic sensors to record footstep data over two days. The test subjects are human and human leading animal. A total of 132 round-trip runs were conducted in two days including 60 runs for the first day and another 72 runs for the second day. The collected footstep data sets, named DEC09 and DEC10, corresponds to two different days December 09 and 10, where DEC09 consists of 30 human and 30 human-animal signals and DEC10 consists of 36 human and 36 human-animal signals. More detailed experimental setup is described in [10].

4.1. Two-class problem

First, we demonstrate the classification effectiveness of our proposed MV-kerSRC method presented in Section 2 on the DEC09 and DEC10 data sets. In particular, we carry experiments to perform discrimination tasks between the human and human-animal footsteps. In this experiment, we use the DEC09 data for training and the DEC10 data for testing. The experiments are performed on each sensor separately, where for each sensor, the corresponding training dictionary $\boldsymbol{X}$ is constructed from all the cepstral feature segments extracted from the 60 training signals. In our experiments, ten overlapping segments (d=10) are taken from each individual sensor signal to capture local signal information. Therefore, the training dictionary $\boldsymbol{X}$ is of size 500×600 and the associated observation $\boldsymbol{Y}$ is of size 500×10, where $n = 500$ is the feature dimension.

Table 1. Classification accuracy (%) for the two-class problem with training samples taken from DEC09.

Methods	H	HA	OA
MV-SRC acoustic sensor 1	**88.89**	41.67	65.28
MV-kerSRC acoustic sensor 1	**88.89**	**100.00**	**94.44**
MV-SRC acoustic sensor 2	80.55	50.00	65.28
MV-kerSRC acoustic sensor 2	**94.44**	**97.22**	**95.83**
MV-SRC seismic sensor 1	66.67	66.67	66.67
MV-kerSRC seismic 1	**80.56**	**91.67**	**86.11**
MV-SRC seismic sensor 2	66.67	61.11	63.89
MV-kerSRC seismic sensor 2	**91.66**	**72.22**	**81.94**
MV-kerSRC with stacking kernel	**100**	94.44	97.22
MV-kerSRC with weighted kernel	97.22	**97.22**	97.22
MTMV-kerSRC	**100**	**97.22**	**98.61**

Table 2. Classification accuracy (%) for the two-class problem with training samples taken from DEC10.

Methods	H	HA	OA
MV-SRC acoustic sensor 1	76.67	46.67	61.67
MV-kerSRC acoustic sensor 1	**100**	**90.00**	**95.00**
MV-SRC acoustic sensor 2	76.67	53.33	65.00
MV-kerSRC acoustic sensor 2	**100**	**76.67**	**88.33**
MV-SRC seismic sensor 1	50.00	66.67	58.33
MV-kerSRC seismic 1	**100**	**86.67**	**93.33**
MV-SRC seismic sensor 2	56.67	66.67	61.67
MV-kerSRC seismic sensor 2	**96.67**	**86.67**	**91.67**
MV-kerSRC with stacking kernel	**100**	93.33	96.67
MV-kerSRC with weighted kernel	**100**	93.33	96.67
MTMV-kerSRC	**100**	**96.67**	**98.34**

For each sensor, we solve the joint sparse recovery problem (4) in the original and kernel domains for each test sample, and then determine the class label by the minimal error residual. Values of the regularization parameter λ in these optimizations are selected via the cross-validation process. The classification performance is summarized in the first part of Table 1, where the first column refers to the methods and the sensor data used in our experiments, which include MV-SRC and MV-kerSRC for four sensors separately. The second and third columns describe the classification accuracy of human (H) and human-animal footsteps (HA) classification performance, and the last column is the overall accuracy (OA), which is simply computed by taking the average of the H and HA. In all our experiments, we employ the classical Gaussian kernel with parameter $\sigma = 2^4$, which we find to archieve the best performance for MV-kerSRC. As one can clearly see from Table 1, the MV-kerSRC method significantly outperforms MV-SRC counterpart for all the test sets. Specifically, the performance gain is roughly $10\% - 15\%$ for all four sensor data.

The performance consistency of the proposed MV-kerSRC is also validated on the other dataset in which we now use DEC10 for training and DEC09 for testing. Accordingly, we now have 72 training and 60 testing samples, respectively. The result is reported in the first part of Table 2. Again, it is clear from this table that the joint sparse representation method operates significantly better in the kernel feature space.

4.2. Combining different sensors

In this section, we discuss various extensions of the MV-kerSRC method in order to incorporate the coupling information between the different sensors. We perform experiments with the methods we proposed in Section 3 and show that combining information from both acoustic and seismic sensors will improve the classification performance substantially. In particularly, we solve the optimizations (4) and (6) with different composite kernels and assign class label by computing residuals once the coefficient matrix is obtained. For the weighted sum kernels, the weights associated with these kernels are set to one. This implies that acoustic and seismic kernels contribute equally for classification. In addition, for these two kernels, the sample $\boldsymbol{x}_i^a$ in Section 3.2 is set as a concatenation of the two acoustic samples $\boldsymbol{x}_i^a = [\boldsymbol{x}_i^{a_1}; \boldsymbol{x}_i^{a_2}]$ where $\boldsymbol{x}_i^{a_1}$ and $\boldsymbol{x}_i^{a_2}$ are samples from the first and second acoustic sensors, respectively. Similar construction is also applied for the sample $\boldsymbol{x}_i^s$ as $\boldsymbol{x}_i^s = [\boldsymbol{x}_i^{s_1}; \boldsymbol{x}_i^{s_2}]$. Classification performances of these methods are shown in the second part of Table 1. As one can see, by exploiting the complimentary information across sensors via the joint sparse representation method, the MTMV-kerSRC and MV-kerSRC with different composite kernels is roughly 4% improvement over the MV-kerSRC, which makes the overall classification accuracy of the MTMV-kerSRC up to nearly 99%.

To validate the efficiency of our proposed approach, we rerun the experiments where we use DEC10 for training and DEC09 for testing. As can be observed in the second part of Table 2, similar phenomenon occurs where MTVT-kerSRC outperforms all other methods.

4.3. Comparing with other methods

Our next experiment compares the proposed MTMV-kerSRC and MV-kerSRC with composite kernel with the current state-of-the-art classifiers such as the (kernel) sparse logistic regression (SLR) [11], SVM, and kernel SVM [8]. In the following experiments, all four sensors are used. In the conventional classifiers such as SVM and SLR, we incorporate information from multiple sensors by concatenating all $D = 4$ sensors' training dictionaries to form an elongated dictionary $\boldsymbol{X}_e \in \mathbb{R}^{Dn \times p}$. Atoms of the new dictionary $\boldsymbol{X}_e$ are considered as the training samples, which are then used to train the SVM and SLR classifiers. These classifiers are then used to

Table 3. Classification accuracy (%) for the two-class problem with training samples taken from DEC09.

Methods	H	HA	OA
MTMV-KerSRC	**100**	**97.22**	**98.61**
MV-kerSRC with stacking kernel	**100**	94.44	97.22
MV-kerSRC with weighted kernel	97.22	**97.22**	97.22
SLR	80.56	97.22	88.93
Kernel SLR	80.56	**100**	90.28
SVM	80.56	89.19	84.87
Kernel SVM	77.78	**100**	88.89

Table 4. Classification accuracy (%) for the two-class problem with training samples taken from DEC10.

Methods	H	HA	OA
MTMV-kerSRC	**100**	**96.67**	**98.34**
MV-kerSRC with stacking kernel	**100**	93.33	96.67
MV-kerSRC with weighted kernel	**100**	93.33	96.67
SLR	80.00	63.33	71.67
Kernel SLR	80.00	80.00	80.00
SVM	80.00	66.67	73.33
Kernel SVM	86.67	66.67	76.67

test on each of the ten concatenated test segments and a voting scheme is employed to finally assign a class label for each test signal which consists of the ten segments. For the kernel versions, we use a RBF kernel with bandwidth selected via cross validation.

Table 3 compares the classification accuracy of the aforementioned approaches as well as the MTMV-SRC approach proposed in [10]. This method is similar to MTMV-kerSRC except it operates in linear domain. As one can clearly observe, our MTMV-KerSRC and MV-kerSRC with composite kernels perform significantly better than the conventional classifiers with roughly 20% performance gain. To further show the efficiency of our approach, we repeat the same experiments using DEC10 for training and DEC09 for testing. The classification performances are provided in Table 4. Similar behaviors can be observed in this table that our proposed approaches considerably outperform all the conventional classification methods. These experiments validate the potential use of our MTMV-KerSRC and MV-kerSRC with composite kernel models for multi-sensor classification.

5. CONCLUSION AND DISCUSSION

In this paper, we propose a new multi-sensor classification technique for personnel footstep detection which is based on sparse representation in a nonlinear feature space induced by a kernel function. The correlation between homogeneous and complementary information between heterogeneous sensors are efficiently exploited via a multi-task multivariate joint sparsity model. Various experimental results on the real dataset show that the kernelization of the joint sparsity-based algorithms significantly improve the classification over the linear version studied in [10]. that our proposed method can be applied to various other applications in which the data is collected from multiple sources (sensors).

6. REFERENCES

[1] M. E. Liggins, J. Llinas, and D. L. Hall, *Handbook of Multisensor Data Fusion: Theory and Practice*, 2nd ed. CRC Press, 2008.

[2] M. Duarte and Y.-H. Hu, "Vehicle classification in distributed sensor networks," *J. Paral. Distrib. Comput.*, vol. 64, no. 7, pp. 826–838, 2004.

[3] A. Klausner, A. Tengg, and B. Rinner, "Vehicle classification on multi-sensor smart cameras using feature- and decision-fusion," *IEEE Conf. Dist. Smart Cameras*, pp. 67–74, 2007.

[4] X.-T. Yuan and S. Yan, "Visual classification with multi-task joint sparse representation," in *IEEE Conf. Comput. Vision Patt. Recog. (CVPR)*, San Francisco, CA, USA, June 2010, pp. 3493–3500.

[5] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, Feb. 2009.

[6] G. Camps-Valls, L. Gomez-Chova, J. Muoz-Mari, J. Vila-Frances, and J. Calpe-Maravilla, "Composite kernels for hyperspectral image classification," *IEEE Geos. Remote Sens. Letters*, vol. 3, no. 1, pp. 93–97, Jan. 2006.

[7] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Roy. Statist. Soc. Ser. B*, vol. 58, no. 1, pp. 267–288, 1996.

[8] B. Scholkopf and A. J. Smola, *Learning with kernels: support vector machine, regularization, optimization, and beyond*. The MIT press, 2002.

[9] E. van den Berg and M. P. Friedlander, "Probing the pareto frontier for basis pursuit solutions," *SIAM J. Sci. Comput.*, vol. 31, no. 2, pp. 890–912, 2008.

[10] N. H. Nguyen, N. M. Nasrabadi, and T. D. Tran, "Robust multi-sensor classification via joint sparse representation," in *Proc. Inter. Conf. Inf. Fusion (FUSION)*, Chicago, IL, USA, Jul. 2011, pp. 1–8.

[11] L. Meier, S. V. D. Geer, and P. Bhlmann, "The group lasso for logistic regression," *J. Royal Statist. Soc.: Series B*, vol. 70, no. 1, pp. 53–71, 2008.