# MULTI-SENSOR CLASSIFICATION VIA SPARSITY-BASED REPRESENTATION WITH LOW-RANK INTERFERENCE

*Minh Dao* [1], *Nasser M. Nasrabadi* [2], *Trac D. Tran* [1]

[1] Department of Electrical and Computer Engineering, The Johns Hopkins University.
[2] U.S. Army Research Laboratory.

## ABSTRACT

In this paper, we propose a general collaborative sparse representation framework for multi-sensor classification which exploits correlation as well as complementary information among homogeneous and heterogeneous sensors while simultaneously extracting the low-rank interference term. Specifically, we observe that incorporating the noise or interfered signal as a low-rank component is essential in a multi-sensor problem when multiple co-located sources/sensors simultaneously record the same physical event. We further extend our frameworks to kernelized models which rely on sparsely representing a test sample in terms of all the training samples in a feature space induced by a kernel function. A fast and efficient algorithm based on alternative direction method is proposed where its convergence to optimal solution is guaranteed. Extensive experiments are conducted on a real data set for a multi-sensor classification problem focusing on discriminating between human and animal footsteps. Results are compared with the conventional classifiers and existing sparsity-based representation methods to verify the effectiveness of our proposed models.

***Index Terms***— Multi-sensor, sparse representation, low-rank, kernel, classification.

## 1. INTRODUCTION

Multi-sensor classification has been an active research topic within the context of various practical applications, such as medical image analysis, remote sensing, and military target/threat detection [1, 2, 3]. These applications normally face the scenerio where data sampling is performed simultaneously from multiple co-located sources/sensors, yet within a small spatio-temporal neighborhood, recording the same physical event. This commonplace scenario allows exploitation of the complementary features within the related signal sources to improve the resulting classification performance. Conventional classifiers such as sparse logistic regression (SLR) [4] and support vector machine (SVM) [5] have been employed to jointly classify signals from multiple sensors/-sources, either at decision level [6] or feature level [3, 7].

Our recent work in [8] introduced a multi-sensor fusion framework which was developed from the sparse representation-based classification (SRC) theory, an algorithmic advance based on the assumption that all of the samples belonging to the same class lie approximately in the same low-dimensional subspace and first proposed for face recognition [9]. In [8], the related information from multiple sensors are fused via a joint sparsity constraint not only within observations of each sensor but also among all sensors. The model is also extended to account for sparse noise and empirically shown to outperform powerful classifiers widely used in machine learning in a multi-sensor border patrol classification problem to discriminate between human and human-animal footsteps.

In this paper, we propose a collaborative multi-sensor sparse representation method for classification, which also incorporates simultaneous structured-sparsity constraints, demonstrated via a row-sparse coefficient matrix, both within and across multiple sensors. However, instead of considering the noise as a sparse term as in [8], we study the case of presenting noise/interference as a low-rank signal. This scenario is normally observed when the recorded data is the superimpositions of target signals with interferences which can be signals from external sources, the underlying background that is inherently anchored in the data, or any pattern noise that remains stationary during signal collection. These interferences normally have correlated structure and appear as a low-rank signal-interference/noise. In a veritable manner, the model with the low-rank interference may be more appropriate for the multi-sensor dataset since the sensors are spatially co-located and data samples are temporally recorded, thus any interference from external sources will have similar effect on all the multiple sensor measurements. The significant improvement in classification performance conducted on the same multi-sensor datasets as in [8] further verifies this low-rank assumption. Moreover, we extend the framework further to cover the integration of a group sparse regularization into our model and the utilization of the sparsity-based representation in the kernel induced feature space, both yielding one more layer of classification robustness.

## 2. MULTI-SENSOR SPARSE REPRESENTATION WITH LOW-RANK INTERFERENCE

Consider a multi-sensor system containing $M$ sensors (so-called $M$ tasks or modalities) used to solve a $C$-class classification problem. For each sensor $m = 1, ..., M$, we de-

This work is partially supported by National Science Foundation under Grant CCF-1117545 and CCF-1422995, Army Research Office under Grant 60219-MA, and Office of Naval Research under Grant N00014-12-1-0765.

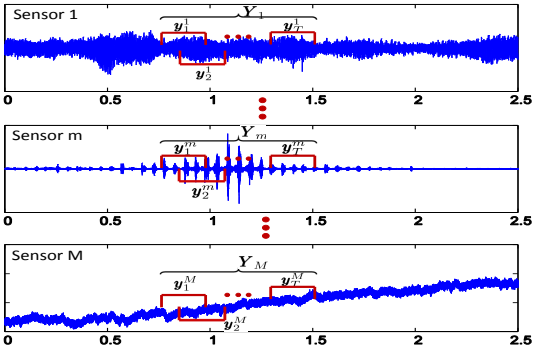


**Fig. 1**: Multi-sensor sample construction.

note $\boldsymbol{D}^m = [\boldsymbol{D}_1^m, \boldsymbol{D}_2^m, ..., \boldsymbol{D}_C^m]$ as an $N \times P$ dictionary, consisting of $C$ sub-dictionaries $\boldsymbol{D}_c^m$'s with respect to $C$ classes. Here, each class sub-dictionary $\boldsymbol{D}_c^m = [\boldsymbol{d}_{c,1}^m, \boldsymbol{d}_{c,2}^m, ..., \boldsymbol{d}_{c,P_c}^m] \in \mathbb{R}^{N \times P_c}$, $c = 1, ..., C$, represents a set of training data from the $m$-th sensor labeled with the $c$-th class; $N$ is the feature dimension of each sample; and $P_c$ is the number of training samples for class $c$, resulting in a total of $P = \sum_{c=1}^{C} P_c$ samples in the dictionary $\boldsymbol{D}^m$. Given a test sample set $\boldsymbol{Y}$ collected from $M$ sensors $\boldsymbol{Y} = [\boldsymbol{Y}^1, \boldsymbol{Y}^2, ..., \boldsymbol{Y}^M]$, where each sample subset $\boldsymbol{Y}^m$ from the sensor $m$ consists of $T$ observations $\boldsymbol{Y}^m = [\boldsymbol{y}_1^m, \boldsymbol{y}_2^m, ..., \boldsymbol{y}_T^m] \in \mathbb{R}^{N \times T}$, we would like to decide which class the sample $\boldsymbol{Y}$ belongs to. In our application, each observation is one local segment of the test signal, where segments are obtained by simultaneously partitioning the test signal of each sensor into $T$ (overlapping) segments, as shown in the Fig. 1.

In [8], a multi-sensor joint-sparse representation (MS-JSR) method is developed by imposing shared row-sparsity constraints on all the coefficient matrices $\boldsymbol{A}^m$'s in the linear representation of the sample subset $\boldsymbol{Y}^m$ over the training dictionary $\boldsymbol{D}^m$: $\boldsymbol{Y}^m = \boldsymbol{D}^m \boldsymbol{A}^m$, i.e. enforcing sparse coefficient vectors obtained from all sensors to share similar sparsity patterns, hence promoting the common class association. The model is then extended to deal with the presence of sparse noise, such as impulsive noise or wind noise, which provides an improvement in the classification performance.

### 2.1. Multi-sensor Joint-Sparse Representation with Low-rank Interference (MS-JSR+L)

In this section, we study a new multi-sensor model that is capable of coping with the dense and large but correlated noise, so-termed low-rank interference. This scenario often happens when there are external sources interfering with the recording process of all sensors. Since all the sensors are mounted onto a common sensing platform, recording the same physical events simultaneously, similar interference sources are picked up across all the sensors promoting large but low-rank corruption. These interference sources may include sound and vibration from a car passing by, a helicopter hovering nearby, or interference from any radio-frequency source. Furthermore, in many situations the recorded data may contain not only the signal of interest but also the intrinsic background that normally stays stationary, hence raising a low-rank background interference. Our proposed multi-sensor joint sparse representation with low-rank interference (MS-JSR+L) model is expected to tackle this problem by extracting the low-rank component while collaboratively taking advantage of having correlated information from different sensors.

Mathematically, each set of measurements $\boldsymbol{Y}^m$ collected from sensor $m$ are composed of a linear representation of the sub-dictionary $\boldsymbol{D}^m$'s atoms and an interference component $\boldsymbol{L}^m$: $\boldsymbol{Y}^m = \boldsymbol{D}^m \boldsymbol{A}^m + \boldsymbol{L}^m$. By concatenating the interference matrices $\boldsymbol{L} = [\boldsymbol{L}^1, \boldsymbol{L}^2, ..., \boldsymbol{L}^M]$ and the coefficient matrices $\boldsymbol{A} = [\boldsymbol{A}^1, \boldsymbol{A}^2, ..., \boldsymbol{A}^M]$, $\boldsymbol{L}$ becomes a low-rank component while $\boldsymbol{A}$ should be sparse at full row level. The coefficient matrix $\boldsymbol{A}$ and low-rank interference component $\boldsymbol{L}$ can be recovered jointly by solving the optimization problem

$$\begin{aligned} \underset{\boldsymbol{A},\boldsymbol{L}}{min} \quad & \|\boldsymbol{A}\|_{1,q} + \lambda_L \|\boldsymbol{L}\|_* \\ s.t. \quad & \boldsymbol{Y}^m = \boldsymbol{D}^m \boldsymbol{A}^m + \boldsymbol{L}^m \quad (m = 1, ..., M), \end{aligned} \tag{1}$$

where the norm $\|\boldsymbol{A}\|_{1,q}$ with $q > 1$, defined as $\|\boldsymbol{A}\|_{1,q} = \sum_{i=1}^{P} \|\boldsymbol{a}_{i,:}\|_q$ with $\boldsymbol{a}_{i,:}$'s being rows of the matrix $\boldsymbol{A}$, encourages shared sparsity patterns within each sensor and across multiple sensors [10, 11, 12]; the nuclear matrix norm $\|\boldsymbol{L}\|_*$ is a convex-relaxation version of the rank defined as the sum of all singular values of the matrix $\boldsymbol{L}$ [13, 14]; and $\lambda_L > 0$ is a weighting parameter balancing the two regularization terms.

Once the solution $\{\widehat{\boldsymbol{A}}, \widehat{\boldsymbol{L}}\}$ of (1) is computed, the class label of $\boldsymbol{Y}$ is decided by the minimal residual rule

$$\text{Class}(\mathbf{Y}) = \underset{c=1,...,C}{\text{argmin}} \sum_{m=1}^{M} \left\| \boldsymbol{Y}^m - \boldsymbol{D}_c^m \widehat{\boldsymbol{A}}_c^m - \widehat{\boldsymbol{L}}_c^m \right\|_F^2, \tag{2}$$

where $\|\cdot\|_F$ is the Frobenious norm of a matrix, and $\boldsymbol{D}_c^m$, $\widehat{\boldsymbol{A}}_c^m$, and $\widehat{\boldsymbol{L}}_c^m$ are the induced matrices associated with the $c$-th class and $m$-th sensor, respectively. This step can be interpreted as collaboratively assigning the class label of $\boldsymbol{Y}$ to the class that can best represent all sample subset $\boldsymbol{Y}^m$'s.

The model in (1) has the capability to extract the low-rank approximation in $\boldsymbol{L}$ while promoting sparsity at row level in the concatenated matrix $\boldsymbol{A}$ at the same time. Moreover, it is inherently robust in the case of existing outliers in the data samples. In other words, many cases of the sparse noise such as nonzero-row sparse matrix (corruptions at certain frequency bands) or nonzero-column sparse matrix (sensor failures) can also be viewed as a low-rank structure and recovered by the nuclear norm minimization.

### 2.2. Multi-sensor Group-Joint-Sparse Representation with Low-rank Interference (MS-GJSR+L)

The idea of adding group structure has been intensively studied [15, 16] and theoretically as well as empirically proved to better represent signals in discriminative applications. This concept is critically beneficial for classification tasks where multiple measurements not necessarily represent the same signals but rather come from the same set of classes. We tentatively apply this concept into the MS-JSR+L model by simultaneously incorporating the group-and-row sparse representation among all sensors together with the regularization

of low-rank interference, which is termed MS-GJSR+L

$$\begin{aligned} \underset{\boldsymbol{A},\boldsymbol{L}}{min} \quad & \|\boldsymbol{A}\|_{1,q} + \lambda_G \sum_{c=1}^{C} \|\boldsymbol{A}_c\|_F + \lambda_L \|\boldsymbol{L}\|_* \\ s.t. \quad & \boldsymbol{Y}^m = \boldsymbol{D}^m \boldsymbol{A}^m + \boldsymbol{L}^m \ (m = 1, ..., M), \end{aligned} \tag{3}$$

where $\boldsymbol{A}_c = [\boldsymbol{A}_c^1, \boldsymbol{A}_c^2, ..., \boldsymbol{A}_c^M]$ is the concatenation of all sub-coefficient matrices $\boldsymbol{A}_c^m$'s induced by the labeled indices corresponding to class $c$; the group regularizer defined by the second term in (3) tends to minimize the number of active classes (groups) in the same matrix $\boldsymbol{A}$; and $\lambda_G \geq 0$ is the weighting parameter of the group constraint. In succession, the model promotes group-sparsity and row-sparsity within a group at the same time, in parallel with extracting the low-rank interference appearing in all measurements all together. Once the solutions of coefficient matrix and low-rank term are recovered, the class label of $\boldsymbol{Y}$ is decided by the same function (1).

### 2.3. Multi-sensor Kernel Group-Joint Sparse Representation with Low-rank Interference (MS-KerGJSR+L).

In this section, we further exploit information from different sensors in the kernel sparse representation to improve classification results. The kernel representation has been well known to yield a significant improvement in discriminative tasks since the kernel-based methods implicitly exploit the higher-order non-linear structure of the testing data which may not be well separated via linear classifiers [17, 18]. Throughout this section, we use similar notations to define test samples and their dictionaries, while the assumption is that the representations in the nonlinear kernel-induced space of test samples within a sensor and among different sensors share the same support sets and the noise/interference in the feature kernel space is still low-rank. The kernelized model of (3) (namely MS-KerGJSR+L) can be written as

$$\begin{aligned} \underset{\boldsymbol{A},\boldsymbol{L}}{min} \quad & \|\boldsymbol{A}\|_{1,q} + \lambda_G \sum_{c=1}^{C} \|\boldsymbol{A}_c\|_F + \lambda_L \|\boldsymbol{L}_\phi\|_* \\ s.t. \quad & \Phi(\boldsymbol{Y}^m) = \Phi(\boldsymbol{D}^m)\boldsymbol{A}^m + \boldsymbol{L}_\phi^m \ (m = 1, ..., M), \end{aligned} \tag{4}$$

where $\Phi$ is an implicit mapping that maps any set of vectors onto a higher dimensional space, possibly infinite, and $\boldsymbol{L}_\phi$ is the additive low-rank interference in the kernal feature domain. Note that in general the mapping function $\Phi$ is not explicitly defined, but rather characterized by the kernel function $\kappa$, defined as the inner product of two vectors: $\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) = \langle \Phi(\boldsymbol{x}_i), \Phi(\boldsymbol{x}_j) \rangle$. The kernel function used in our application is the radial basis function (RBF) Gaussian kernel $\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp(-\|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2^2 / \eta^2)$ with $\eta$ used to control the width of the RBF [19]. The classification assignment for $\boldsymbol{Y}$ in the kernelized model (4) is slightly modified from (2) with attentive manipulations of involving kernel functions.

## 3. ALGORITHM

In this section, we provide a fast and efficient algorithm based on the alternating direction method of multipliers (ADMM) [20] to solve the convex optimization problem (3). The optimization (1) can be similarly solved by setting $\lambda_G = 0$ and the algorithm for MS-KerGJSR+L in (4) can be modified by evaluating the kernel functions at the training points. The augmented Lagrangian function of (3) is defined as

$$\begin{aligned} \mathcal{L}(\boldsymbol{A}, \boldsymbol{L}, \boldsymbol{Z}) &= \|\boldsymbol{A}\|_{1,q} + \lambda_G \sum_{c=1}^{C} \|\boldsymbol{A}_c\|_F + \lambda_L \|\boldsymbol{L}\|_* \\ &+ \sum_{m=1}^{M} \Big[ \langle \boldsymbol{Y}^m - \boldsymbol{D}^m \boldsymbol{A}^m - \boldsymbol{L}^m, \boldsymbol{Z}^m \rangle + \frac{\mu}{2} \|\boldsymbol{Y}^m - \boldsymbol{D}^m \boldsymbol{A}^m - \boldsymbol{L}^m\|_F^2 \Big], \end{aligned} \tag{5}$$

where $\boldsymbol{Z} = [\boldsymbol{Z}^1, ..., \boldsymbol{Z}^M]$ is the multiplier for the smoothness constraints, and $\mu$ is a positive penalty parameter. The algorithm then minimizes $\mathcal{L}(\boldsymbol{A}, \boldsymbol{L}, \boldsymbol{Z})$ with respect to one variable at a time by keeping others fixed and then updating the variables sequentially and is formally presented in Algorithm 1.

**Inputs**: Matrices $\boldsymbol{Y}$ and $\boldsymbol{D}$, weighting parameters $\lambda_G$ and $\lambda_L$.
**Initializations**: $\boldsymbol{A}_0 = \boldsymbol{0}, \boldsymbol{L}_0 = \boldsymbol{0}, j = 0$.
**While** not converged **do**
1. Solve for $\boldsymbol{L}_{j+1}$: $\boldsymbol{L}_{j+1} = \text{argmin}_{\boldsymbol{L}} \ \mathcal{L}(\boldsymbol{A}_j, \boldsymbol{L}, \boldsymbol{Z}_j)$
2. Solve for $\boldsymbol{A}_{j+1}$: $\boldsymbol{A}_{j+1} = \text{argmin}_{\boldsymbol{A}} \ \mathcal{L}(\boldsymbol{A}, \boldsymbol{L}_{j+1}, \boldsymbol{Z}_j)$
3. Update the multiplier:
$\boldsymbol{Z}_{j+1}^m = \boldsymbol{Z}_j^m + \mu(\boldsymbol{Y}^m - \boldsymbol{D}^m \boldsymbol{A}_{j+1}^m - \boldsymbol{L}_{j+1}^m), (m = 1, ..., M)$
4. $j = j + 1$.
**end while**
**Outputs**: $(\widehat{\boldsymbol{A}}, \widehat{\boldsymbol{L}}) = (\boldsymbol{A}_j, \boldsymbol{L}_j)$.

Algorithm 1: ADMM for MS-GJSR+L.

The first optimization subproblem which updates variable $\boldsymbol{L}$ can be recasted as

$$\begin{aligned} \boldsymbol{L}_{j+1} &= \underset{\boldsymbol{L}}{\text{argmin}} \ \lambda_L \|\boldsymbol{L}\|_* + \frac{\mu}{2} \sum_{m=1}^{M} \left\| \boldsymbol{L}^m - (\boldsymbol{Y}^m - \boldsymbol{D}^m \boldsymbol{A}_j^m - \frac{\boldsymbol{Z}_j^m}{\mu}) \right\|_F^2 \\ &= \underset{\boldsymbol{L}}{\text{argmin}} \ \lambda_L \|\boldsymbol{L}\|_* + \frac{\mu}{2} \|\boldsymbol{L} - \boldsymbol{P}\|_F^2, \end{aligned} \tag{6}$$

where $\boldsymbol{P}^m = (\boldsymbol{Y}^m - \boldsymbol{D}^m \boldsymbol{A}_j^m - \frac{1}{\mu} \boldsymbol{Z}_j^m)$ and $\boldsymbol{P} = [\boldsymbol{P}^1, ..., \boldsymbol{P}^M]$. The proximal minimization in (6) can be solved via the well-known singular value thresholding (SVT) operator [13].

The second subproblem to update $\boldsymbol{A}$ can be re-written as

$$\begin{aligned} \boldsymbol{A}_{j+1} = \underset{\boldsymbol{A}}{\text{argmin}} & \ \|\boldsymbol{A}\|_{1,q} + \lambda_G \sum_{c=1}^{C} \|\boldsymbol{A}_c\|_F \\ & + \frac{\mu}{2} \sum_{m=1}^{M} \left\| \boldsymbol{D}^m \boldsymbol{A}^m - (\boldsymbol{Y}^m - \boldsymbol{L}_{j+1}^m - \frac{1}{\mu} \boldsymbol{Z}_j^m) \right\|_F^2. \end{aligned} \tag{7}$$

This subproblem in (7) is a convex utility function. Unfortunately, its closed-form solution is not easily determined. Therefore, we do not solve for an exact solution of (7) but approximate the third term by its Taylor expansion at $\boldsymbol{A}_j^m$ up to the second derivative order and use separable structure properties of $\|\cdot\|_F^2$ and $\|\cdot\|_{1,q}$ to simplify it to

$$\begin{aligned} (\boldsymbol{A}_{j+1})_c = \underset{\boldsymbol{A}_c}{\text{argmin}} & \|\boldsymbol{A}_c\|_{1,q} + \lambda_G \|\boldsymbol{A}_c\|_F + \frac{\mu}{2\theta} \|\boldsymbol{A}_c - ((\boldsymbol{A}_j)_c - \theta(\boldsymbol{T}_j)_c)\|_F^2 \\ & (\forall c = 1, 2, ..., C), \end{aligned} \tag{8}$$

where $\theta$ is a positive proximal parameter, and $\boldsymbol{T}_j = [\boldsymbol{T}_j^1, ..., \boldsymbol{T}_j^M]$ with $\boldsymbol{T}_j^m = (\boldsymbol{D}^m)^T (\boldsymbol{D}^m \boldsymbol{A}_j^m - (\boldsymbol{Y}^m - \boldsymbol{L}_{j+1}^m - \frac{1}{\mu} \boldsymbol{Z}_j^m))$ being the gradient at $\boldsymbol{A}_j^m$ of the Taylor expansion. The explicit solution of (8) can then be solved via the proximal operators for both group-sparsity and row-sparsity as can be modified from [15]. Furthermore, Algorithm 1 is guaranteed to provide the global optimum of the convex program (3) as stated in the following proposition (see [21] for the detailed proof).

| Set | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|
| Sensors | $S_{1-2}$ | $S_{5-7}$ | $S_{1-4}$ | $S_{1-7}$ | $S_{1-2,5-9}$ | $S_{1-9}$ |

**Table 1**: List of sensor combinations.

**Fig. 2**: Four acoustic sensors (left), seismic sensor (middle left), PIR sensor (middle right) and ultrasound sensor (right).

***Proposition 1***: *If the proximal parameter $\theta$ satisfies the condition: $\max_{1\leq m\leq M}\{\sigma_{max}((\boldsymbol{D}^m)^T\boldsymbol{D}^m)\} < \frac{1}{\theta}$, where $\sigma_{max}(\cdot)$ is the largest eigenvalue of a matrix, then $\{\boldsymbol{A}_j, \boldsymbol{L}_j\}$ generated by algorithm 1 for any value of the penalty coefficient $\mu$ converges to the optimal solution $\{\widehat{\boldsymbol{A}}, \widehat{\boldsymbol{L}}\}$ of (3) as $j \to \infty$.*

## 4. EXPERIMENTAL RESULTS

We verify the effectiveness of the proposed methods via solving a classification problem on multi-sensor data focusing on discriminating between human and human-animal footsteps. The experimental setup is as follows: a set of nine sensors including four acoustic ($S_{1-4}$), three seismic ($S_{5-7}$), one passive infrared (PIR) ($S_8$), and one ultrasonic ($S_9$) sensors (visuallized in Fig. 2) are used to measure the same physical event simultaneously on the field. The ultimate goal is to detect whether the event involves human or human leading animal footsteps. The data collection and signal description can be found in detail as in [8].

The system allows to evaluate the classification results based on various combinations of recording sources, not just multiple sensors of the same brands but also sensors of different signal types. For all methods, 15 combination sets of sensors are processed and compared, in which the first 9 sets are conducted using one single sensor separately, corresponding to $S_1$, $S_2$, ... $S_9$. The next 6 sets combine multiple sensors into various scenarios as listed in Table 1.

Our three proposed methods, MS-JSR+L, MS-GJSR+L and MS-KerGJSR+L are processed through all 15 sensor sets to determine the joint coefficient matrix $\mathbf{A}$ and class label is then obtained by minimal error residual classifiers. The results are then compared with MS-JSR and MS-JSR+E developed in [8] and other popular powerful classifiers such as sparse logistic regression (SLR) [4], heterogeneous feature machine (HFM) [22], linear support vector machine (SVM) [19], and their kernelized versions to verify the effectiveness of the proposed methods. The classification rates defined as ratios of the total number of correctly classified samples to the total number of testing samples, expressed as percentages, are plotted in Fig. 3 and Fig. 4, corresponding to two testing data sets DEC09 and DEC10 collected in December 09 and 10, respectively. We tabulate the classification performance of all the proposed models as well as competing methods in Table 2. The first and second columns in each sub-table (for DEC09 and DEC10) describe the classification rates by using single sensor (SS) and multiple sensors (MS) (which average the rates of sets

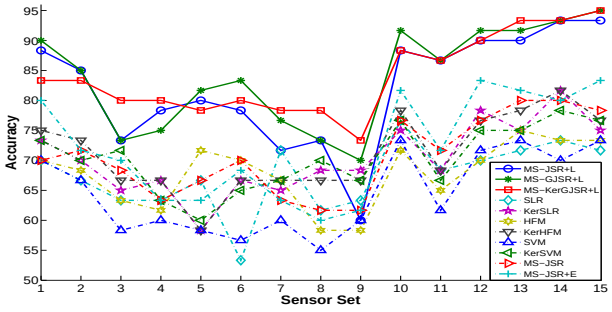


**Fig. 3**: Comparison of classification results - DEC09 for testing.

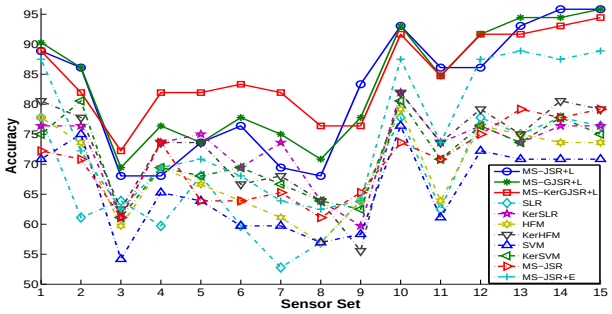


**Fig. 4**: Comparison of classification results - DEC10 for testing.

1-9 and 10-15, respectively), and the last column shows the overall results by averaging over all 15 sensor sets (AS).

The experimental results with particular datasets show that our proposed models outperform the other conventional classifiers and sparsity models developed in [8]. These results also reveal several critical observations: *(1)* the use of complementary information from multiple sensors significantly improves the classification results over just using a singular sensor; *(2)* appropriate structured regularizations (joint and group sparsity) bring more advantage in selecting the right classes, hence increasing the accuracy rate; *(3)* low-rank noise/signal-interference is a critical issue in a multi-sensor fusion problem and *(4)* the classification in feature space induced by a kernel function yields a compelling performance improvement. Nevertheless, although our techniques are only verified on the data set specifically geared for border patrol control, they are not restricted to this specific application. Rather, they can be applied to any set of classification or discrimination problems, where the data is collected from multiple co-located sensors.

| Methods | DEC09 for testing | | | DEC10 for testing | | |
|---|---|---|---|---|---|---|
| | SS | MS | AS | SS | MS | AS |
| SLR | 64.44 | 71.94 | 67.44 | 62.65 | 74.54 | 67.41 |
| Ker-SLR | 66.85 | 75.56 | 70.33 | 69.91 | 76.39 | 72.50 |
| HFM | 65.37 | 71.39 | 67.78 | 65.90 | 73.61 | 68.98 |
| Ker-HFM | 67.41 | 76.67 | 71.11 | 69.14 | 78.24 | 72.78 |
| SVM | 60.56 | 70.56 | 64.56 | 62.65 | 70.37 | 65.74 |
| Ker-SVM | 67.41 | 74.72 | 70.33 | 68.52 | 75.69 | 71.39 |
| MS-JSR | 66.30 | 77.22 | 70.67 | 66.36 | 75.93 | 70.19 |
| MS-JSR+E | 66.85 | 80.28 | 72.22 | 68.98 | 85.65 | 75.65 |
| MS-JSR+L | 76.48 | 90.28 | 82.00 | 75.77 | 91.67 | 82.13 |
| MS-GJSR+L | 78.70 | **91.67** | **83.89** | 77.47 | **92.36** | 83.43 |
| MS-KerGJSR+L | **79.44** | 90.56 | **83.89** | **80.25** | 90.74 | **84.44** |

**Table 2**: Summarized classification results.

## 5. REFERENCES


[1] D. L. Hall and J. Llinas, "An introduction to multisensor data fusion," *Proceedings of the IEEE*, vol. 85, no. 1, pp. 6–23, 1997.

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

[3] P. K. Varshney, "Multisensor data fusion," *Electronics and Communication Engineering Journal*, vol. 9, no. 6, pp. 245–253, 1997.

[4] L. Meier, S. V. D. Geer, and P. Bhlmann, "The group lasso for logistic regression," *Journal of the Royal Statistical Society: Series B*, vol. 70, no. 1, pp. 53–71, 2008.

[5] B. Waske and J. A. Benediktsson, "Fusion of support vector machines for classification of multisensor data," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 45, no. 12, pp. 3858–3866, 2007.

[6] M. Duarte and Y.-H. Hu, "Vehicle classification in distributed sensor networks," *Journal of Parallel and Distributed Computing*, vol. 64, no. 7, pp. 826–838, 2004.

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

[8] N. H. Nguyen, N. M. Nasrabadi, and T. D. Tran, "Robust multi-sensor classification via joint sparse representation," in *IEEE International Conference on Information Fusion (FUSION)*, 2011, pp. 1–8.

[9] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210–227, 2009.

[10] P. Zhao, G. Rocha, and B. Yu, "The composite absolute penalties family for grouped and hierarchical variable selection," *Annals of Statistics*, vol. 37, pp. 3469–3497, 2009.

[11] G. Obozinski, M. J. Wainwright, and M. I. Jordan, "Support union recovery in high-dimensional multivariate regression," *Annals of Statistics*, vol. 39, no. 1, pp. 1–47, 2011.

[12] H. Zhang, N. M. Nasrabadi, Y. Zhang, and T. S. Huang, "Joint dynamic sparse representation for multi-view face recognition," *Pattern Recognition*, vol. 45, no. 4, pp. 1290–1298, 2012.

[13] J. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM Journal on Optimization*, vol. 20, no. 4, pp. 1956–1982, 2010.

[14] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *Journal of ACM*, vol. 58, no. 3, pp. 1–37, 2011.

[15] P. Sprechmann, I. Ramírez, G. Sapiro, and Y. C. Eldar, "C-hilasso: A collaborative hierarchical sparse modeling framework," *IEEE Trans. on Signal Processing*, vol. 59, no. 9, pp. 4183–4198, 2011.

[16] Y. Suo, M. Dao, U. Srinivas, V. Monga, and T. D. Tran, "Structured dictionary learning for classification," *arXiv preprint arXiv:1406.1943*, 2014.

[17] G. Camps-Valls and L. Bruzzone, "Kernel-based methods for hyperspectral image classification," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 43, no. 6, pp. 1351–1362, 2005.

[18] S. Gao, I. W.-H. Tsang, and L.-T. Chia, "Kernel sparse representation for image classification and face recognition," in *Computer Vision–ECCV 2010*. Springer, 2010, pp. 1–14.

[19] J. Shawe-Taylor and N. Cristianini, *Kernel Methods for Pattern Analysis*. Cambridge University Press, 2004.

[20] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.

[21] M. Dao, N. H. Nguyen, N. M. Nasrabadi, and T. D. Tran, "Collaborative multi-sensor classification via sparsity-based representation," *arXiv preprint arXiv:1410.7876*, 2014.

[22] L. Cao, J. Luo, F. Liang, and T. S. Huang, "Heterogeneous feature machines for visual recognition," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1095–1102, 2009.