2013 IEEE INTERNATIONAL WORKSHOP ON MACHINE LEARNING FOR SIGNAL PROCESSING, SEPT. 22–25, 2013, SOUTHAMPTON, UK

SPARSE CODING WITH ANOMALY DETECTION

Amir Adler[⋆]*, Michael Elad*[⋆]*, Yacov Hel-Or* [†]*, Ehud Rivlin*[⋆]

[⋆] Computer Science Department, Technion, Haifa 32000, Israel
[†] Computer Science Department, Interdisciplinary Center, Herzelia 46150, Israel

ABSTRACT

We consider the problem of simultaneous sparse coding and anomaly detection in a collection of data vectors. The majority of the data vectors are assumed to conform with a sparse representation model, whereas the anomaly is caused by an unknown subset of the data vectors - the outliers - which significantly deviate from this model. The proposed approach utilizes the Alternating Direction Method of Multipliers (ADMM) to recover simultaneously the sparse representations and the outliers components for the entire collection. This approach provides a unified solution both for jointly sparse and independently sparse data vectors. We demonstrate the usefulness of the proposed approach for irregular heartbeats detection in Electrocardiogram (ECG) and specular reflectance removal from natural images.

Index Terms— sparse coding, anomaly detection, ADMM, arrythmia detection, specular reflectance removal.

1. INTRODUCTION

Anomaly detection is the problem of detecting patterns that significantly deviate from an expected model. This problem has numerous applications such as fraud detection for banking and businesses, intrusion detection for network security, fault detection for production systems, health problems detection for biomedical systems and more, see [1] for a review. In this paper we assume that the expected behavior of the data vectors is to conform with a sparse representation model [2], and address the problem of simultaneous sparse coding and anomaly detection. This problem can be applied to three different tasks: 1) anomaly detection within sparsely represented data vectors. 2) removal of interference from sparsely represented data vectors. 3) dictionary learning in the presence of outliers. In this paper we address the first two tasks, whereas the latter is beyond the scope of this work.

Related work. Joint-sparse coding was addressed by [3, 4] for cases in which *all* data vectors are contaminated by either a sparse or a sparsely-represented interference. Anomaly detection in video was addressed by [5] which proposed a sparse reconstruction cost to measure the normality of events, with respect to a dictionary with various spatio-temporal structures. This problem was addressed also by [6], which combined online dictionary learning with an objective function that measures the normality of events. The work of [7] utilized sparse representations to analyze stochastic processes over graphs for anomaly detection in SmartGrids.

Contributions. The contributions of this paper are two-fold: 1) A unified formulation for the problem of simultaneous sparse coding and anomaly detection is provided for jointly sparse as well as for independently sparse data vectors; and a numerical solver is provided for both cases. 2) the proposed approach is demonstrated to detect irregular heartbeats in ECG and remove specular reflections from natural images.

Organization. Section 2 reviews sparse representations, section 3 formulates the problem, section 4 explains the proposed approach, and section 5 demonstrates its performance.

2. SPARSE REPRESENTATION MODELING

Sparse representation modeling [2] assumes that a signal (data vector) $\mathbf{y} \in \mathbb{R}^N$ can be described as $\mathbf{y} \approx \mathbf{D}\mathbf{x}$, where $\mathbf{D} \in \mathbb{R}^{N \times M}$ is a *dictionary* and $\mathbf{x} \in \mathbb{R}^M$ is sparse. Therefore, $\mathbf{y}$ is represented by a linear combination of *few* columns (atoms) of $\mathbf{D}$. The recovery of the sparse representation, termed *sparse coding*, can be obtained by solving the following problem:

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{D}\mathbf{x}\|_2^2 \text{ s.t. } \|\mathbf{x}\|_0 \leq T_0, \qquad (1)$$

where $\|\mathbf{x}\|_0$ is the l_0 pseudo-norm that counts the number of non-zero entries of $\mathbf{x}$, and T_0 is the maximum number of non-zero entries. Problem (1) can be augmented for a collection of signals:

$$\hat{\mathbf{X}} = \arg\min_{\mathbf{X}} \|\mathbf{Y} - \mathbf{D}\mathbf{X}\|_F^2 \text{ s.t. } \|\mathbf{X}\|_0 \leq LT_0, \qquad (2)$$

where $\mathbf{Y} \in \mathbb{R}^{N \times L}$ contains L signals $\{\mathbf{y}_i \in \mathbb{R}^N\}_{i=1}^L$, $\mathbf{X} \in \mathbb{R}^{M \times L}$ contains L sparse representations $\{\mathbf{x}_i \in \mathbb{R}^M\}_{i=1}^L$ and $\|\mathbf{X}\|_0$ counts the number of non-zero entries of $\mathbf{X}$. This type of model is referred to as the Single Measurement Vector

A. Adler is the recipient of the 2011 Google European Doctoral Fellowship in Multimedia, and this research is partly supported by this Google Fellowship and partly by ERC Grant agreement no. 320649.

978-1-4799-1180-6/13/$31.00

(SMV), since each signal is assumed to be a single measurement associated with a unique non-zero pattern of its sparse representation (i.e. a unique combination of atoms). The case in which all the sparse representations share the same non-zero pattern is referred to as the Multiple Measurement Vector (MMV) [8] or joint-sparsity model, as illustrated in Fig. 1. For the MMV case, the following optimization problem recovers more accurately the sparse representations, by exploiting the joint-sparsity property:

$$\hat{\mathbf{X}} = \arg\min_{\mathbf{X}} \|\mathbf{Y} - \mathbf{D}\mathbf{X}\|_F^2 \text{ s.t. } \|\mathbf{X}\|_{0,2} \leq T_0, \quad (3)$$

where $\|\mathbf{X}\|_{0,2} = \sum_j \mathcal{I}(\|\mathbf{X}(j,:)\|_2)$ counts the number of non-zero rows, $\mathbf{X}(j,:)$ is the j-th row of $\mathbf{X}$ and $\mathcal{I}$ is the indicator function:

$$\mathcal{I}(a) = \begin{cases} 1 & \text{if } |a| > 0 \\ 0 & \text{otherwise} \end{cases}.$$

Note that problems (1)-(3) are NP-hard and their solutions can be approximated using convex relaxations: the l_1 norm $\|\mathbf{x}\|_1 = \sum_i |x_i|$ often replaces $\|\mathbf{x}\|_0$, and the $l_{1,p}$ norm $\|\mathbf{X}\|_{1,p} = \sum_j \|\mathbf{X}(j,:)\|_p$ often replaces $\|\mathbf{X}\|_0$ with $p = 1$ and $\|\mathbf{X}\|_{0,2}$ with $p = 2$.

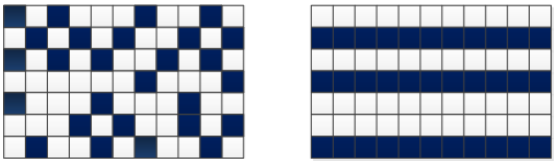

Fig. 1. The non-zeros (dark squares) of the sparse representations matrix $\mathbf{X}$ for the SMV (left) and MMV (right) models.

3. PROBLEM FORMULATION

Let $\mathbf{Y} \in \mathbb{R}^{N \times L}$ be a collection of signals that are well approximated by a sparse representations model, excluding a small number of signals - the outliers - which significantly deviate from this model. The collection $\mathbf{Y}$ is described as follows:

$$\mathbf{Y} = \mathbf{D}\mathbf{X} + \mathbf{E} + \mathbf{V}, \quad (4)$$

where $\mathbf{D}$ is assumed known, $\mathbf{E}$ has few non-zero columns that equal to the deviation of each outlier from the sparse representations model, and $\mathbf{V}$ is a low-energy noise component ($\|\mathbf{V}\|_F^2$ is small compared to $\|\mathbf{Y}\|_F^2$).
Our objective is to detect the outliers in $\mathbf{Y}$ and recover the sparse representations. For the SMV case this objective can be obtained by solving the following problem:

$$\begin{aligned} \{\mathbf{X}, \mathbf{E}\} = \arg\min_{\mathbf{X},\mathbf{E}} & \|\mathbf{Y} - \mathbf{D}\mathbf{X} - \mathbf{E}\|_F^2 \\ \text{s.t. } & \|\mathbf{X}\|_0 \leq L T_0 \\ & \|\mathbf{E}\|_{2,0} \leq K_0, \end{aligned} \quad (5)$$

where $\|\mathbf{E}\|_{2,0} = \sum_i \mathcal{I}(\|\mathbf{E}(:,i)\|_2)$ counts the number of non-zero columns in $\mathbf{E}$, $\mathbf{E}(:,i)$ is the i-th column of $\mathbf{E}$, and K_0 is the maximum number of non-zero columns (i.e. outliers). Problem (5) encourages a solution in which $\mathbf{X}$ is sparse, however, for the outliers that cannot be represented exclusively by $\mathbf{D}$, it permits non-zero columns in $\mathbf{E}$. For the MMV case the objective can be obtained by solving the following problem:

$$\begin{aligned} \{\mathbf{X}, \mathbf{E}\} = \arg\min_{\mathbf{X},\mathbf{E}} & \|\mathbf{Y} - \mathbf{D}\mathbf{X} - \mathbf{E}\|_F^2 \\ \text{s.t. } & \|\mathbf{X}\|_{0,2} \leq T_0 \\ & \|\mathbf{E}\|_{2,0} \leq K_0, \end{aligned} \quad (6)$$

where the constraints ensure at most T_0 non-zero rows in $\mathbf{X}$ and at most K_0 non-zero columns in $\mathbf{E}$.

4. THE PROPOSED APPROACH

The solutions to problems[1] (5) and (6) can be approximated by solving the following unconstrained convex problem:

$$\min_{\mathbf{X},\mathbf{E}} \frac{1}{2}\|\mathbf{Y} - \mathbf{D}\mathbf{X} - \mathbf{E}\|_F^2 + \alpha\|\mathbf{X}\|_{1,p} + \beta\|\mathbf{E}\|_{2,1} \quad (7)$$

where $p = 1$ for the SMV case, $p = 2$ for the MMV case and α, β are a small positive scalars. In addition, $\|\mathbf{E}\|_{2,1} = \sum_i \|\mathbf{E}(:,i)\|_2$ is the $l_{2,1}$ norm which provides a convex relaxation to $\|\mathbf{E}\|_{2,0}$, and was applied in [9] for robust non-negative matrix factorization. We propose to solve problem (7) with the Alternating Direction Method of Multipliers (ADMM) [10], and in the following subsections we describe this method and its application to our problem.

4.1. Alternating Direction Method of Multipliers

ADMM is a numerical method for solving problems of the following form:

$$\min_{\mathbf{X},\mathbf{Z}} f(\mathbf{X}, \mathbf{Z}) \text{ s.t. } \mathbf{A}\mathbf{X} + \mathbf{B}\mathbf{Z} = \mathbf{C}, \quad (8)$$

where $\mathbf{X}, \mathbf{Z}, \mathbf{A}, \mathbf{B}, \mathbf{C}$ are matrices, and the objective function is either separable $f(\mathbf{X}, \mathbf{Z}) = g(\mathbf{X}) + h(\mathbf{Z})$ or bi-convex. ADMM solves (8) by minimizing its Augmented-Lagrangian:

$$\begin{aligned} & \mathcal{L}_A(\mathbf{X}, \mathbf{Z}, \mu, \mathbf{M}) = f(\mathbf{X}, \mathbf{Z}) \\ & + < \mathbf{M}, \mathbf{A}\mathbf{X} + \mathbf{B}\mathbf{Z} - \mathbf{C} > + \frac{\mu}{2}\|\mathbf{A}\mathbf{X} + \mathbf{B}\mathbf{Z} - \mathbf{C}\|_F^2, \end{aligned} \quad (9)$$

where $\mathbf{M}$ is a Lagrange multiplier and μ is a penalty coefficient that controls the penalty level of deviation from the equality constraint. The minimization of $\mathcal{L}_A(\mathbf{X}, \mathbf{Z}, \mu, \mathbf{M})$

[1] The observant reader may notice that problem (5) is actually separable, implying that we can solve for each column of X independently from the others. Nevertheless, we choose in this paper a joint solver for two reasons: (i) Giving a unified view of the two problems (5 and 6); and (ii) Our approach loses nothing in terms of complexity nor elegance when compared to the independent sparse coding tasks.

is performed iteratively, while alternating between the minimizations of $\mathbf{X}$ and $\mathbf{Z}$:

$$\mathbf{X}^{k+1} = \arg\min_{\mathbf{X}} \mathcal{L}_A(\mathbf{X}, \mathbf{Z}^k, \mu^k, \mathbf{M}^k) \quad (10)$$

$$\mathbf{Z}^{k+1} = \arg\min_{\mathbf{Z}} \mathcal{L}_A(\mathbf{X}^{k+1}, \mathbf{Z}, \mu^k, \mathbf{M}^k) \quad (11)$$

$$\mathbf{M}^{k+1} = \mathbf{M}^k + \mu^k(\mathbf{A}\mathbf{X}^{k+1} + \mathbf{B}\mathbf{Z}^{k+1} - \mathbf{C}) \quad (12)$$

$$\mu^{k+1} = \rho\mu^k; \rho > 1. \quad (13)$$

ADMM can be extended to more than two variables, and its convergence properties are analyzed in [10].

4.2. Sparse Coding with Anomaly Detection

In order to apply ADMM to solve problem (7), we add an auxiliary variable Z and an equality constraint as follows:

$$\min_{\mathbf{X},\mathbf{E},\mathbf{Z}} \frac{1}{2}\|\mathbf{Y} - \mathbf{D}\mathbf{X} - \mathbf{E}\|_F^2 + \alpha\|\mathbf{Z}\|_{1,p} + \beta\|\mathbf{E}\|_{2,1} \text{ s.t. } \mathbf{Z} = \mathbf{X}. \quad (14)$$

The Augmented-Lagrangian of this problem is given by:

$$\mathcal{L}_p(\mathbf{X}, \mathbf{Z}, \mathbf{E}, \mathbf{M}, \mu) = \frac{1}{2}\|\mathbf{Y} - \mathbf{D}\mathbf{X} - \mathbf{E}\|_F^2 + \alpha\|\mathbf{Z}\|_{1,p} + \beta\|\mathbf{E}\|_{2,1} + < \mathbf{M}, \mathbf{Z} - \mathbf{X} > + \frac{\mu}{2}\|\mathbf{Z} - \mathbf{X}\|_F^2. \quad (15)$$

The main stages of the ADMM-based solution are summarized in Algorithm 1, and in the following we describe the update equations of $\mathbf{X}^{k+1}, \mathbf{Z}^{k+1}, \mathbf{E}^{k+1}$. The update equation of $\mathbf{X}^{k+1}$ is closed-form (and derived in the Appendix):

$$\mathbf{X}^{k+1} = (\mathbf{D}^T\mathbf{D} + \mu^k\mathbf{I})^{-1}(\mathbf{D}^T(\mathbf{Y} - \mathbf{E}^k) + \mathbf{M}^k + \mu^k\mathbf{Z}^k). \quad (16)$$

The update equation of $\mathbf{Z}^{k+1}$ for the SMV case is obtained from:

$$\mathbf{Z}^{k+1} = \arg\min_{\mathbf{Z}} \alpha\|\mathbf{Z}\|_{1,1} + < \mathbf{M}^k, \mathbf{Z} - \mathbf{X}^{k+1} > + \frac{\mu^k}{2}\|\mathbf{Z} - \mathbf{X}^{k+1}\|_F^2, \quad (17)$$

which can be simplified to :

$$\mathbf{Z}^{k+1} = \arg\min_{\mathbf{Z}} \frac{1}{2}\|\mathbf{P} - \mathbf{Z}\|_F^2 + \gamma\|\mathbf{Z}\|_{1,1}, \quad (18)$$

where $\mathbf{P} = \mathbf{X}^{k+1} - \frac{1}{\mu^k}\mathbf{M}^k$ and $\gamma = \frac{\alpha}{\mu^k}$. The solution to problem (18) is the element-wise soft thresholding operator [10]:

$$\mathbf{Z}^{k+1} = \mathcal{S}_\gamma(\mathbf{P}), \quad (19)$$

where:

$$\mathcal{S}_\gamma(a) = \begin{cases} a - \gamma & \text{if } a > \gamma \\ 0 & \text{if } |a| \leq \gamma \\ a + \gamma & \text{if } a < -\gamma \end{cases}.$$

Algorithm 1 Sparse Coding with Anomaly Detection

Solve: $\min_{\mathbf{X},\mathbf{E}} \frac{1}{2}\|\mathbf{Y} - \mathbf{D}\mathbf{X} - \mathbf{E}\|_F^2 + \alpha\|\mathbf{X}\|_{1,p} + \beta\|\mathbf{E}\|_{2,1}$.
Input: signals $\mathbf{Y} \in \mathbb{R}^{N\times L}$, Dictionary $\mathbf{D} \in \mathbb{R}^{N\times M}$.
Mode: $p = 1$ for SMV or $p = 2$ for MMV.
Initialize: set $k = 0, \mathbf{Z}^0, \mathbf{E}^0, \mathbf{M}^0, \mu^0, \rho, \epsilon$.
Repeat Until Convergence:

1. $\mathbf{X}^{k+1} = \arg\min_{\mathbf{X}} \mathcal{L}_p(\mathbf{X}, \mathbf{Z}^k, \mathbf{E}^k, \mathbf{M}^k, \mu^k)$.
2. $\mathbf{Z}^{k+1} = \arg\min_{\mathbf{Z}} \mathcal{L}_p(\mathbf{X}^{k+1}, \mathbf{Z}, \mathbf{E}^k, \mathbf{M}^k, \mu^k)$.
3. $\mathbf{E}^{k+1} = \arg\min_{\mathbf{E}} \mathcal{L}_p(\mathbf{X}^{k+1}, \mathbf{Z}^{k+1}, \mathbf{E}, \mathbf{M}^k, \mu^k)$.
4. $\mathbf{M}^{k+1} = \mathbf{M}^k + \mu^k(\mathbf{Z}^{k+1} - \mathbf{X}^{k+1})$.
5. $\mu^{k+1} = \rho\mu^k$.
6. $k = k + 1$.
7. Stop if $\frac{\|\mathbf{Z}^k - \mathbf{X}^k\|_F^2}{\|\mathbf{X}^k\|_F^2} < \epsilon$

Output: $\mathbf{X}^k, \mathbf{E}^k$.

The update equation of $\mathbf{Z}^{k+1}$ for the MMV case is given by:

$$\mathbf{Z}^{k+1} = \arg\min_{\mathbf{Z}} \frac{1}{2}\|\mathbf{P} - \mathbf{Z}\|_F^2 + \gamma\|\mathbf{Z}\|_{1,2}, \quad (20)$$

which results in a row-shrinkage operator (as proved in [4]):

$$\mathbf{Z}^{k+1}(j,:) = \begin{cases} \frac{\|\mathbf{P}(j,:)\|_2 - \gamma}{\|\mathbf{P}(j,:)\|_2}\mathbf{P}(j,:) & \text{if } \gamma < \|\mathbf{P}(j,:)\|_2 \\ 0 & \text{otherwise} \end{cases}, \quad (21)$$

where $\mathbf{P}(j,:)$ is the *j*-th row of $\mathbf{P}$.
The update equation of $\mathbf{E}^{k+1}$ is obtained from:

$$\mathbf{E}^{k+1} = \arg\min_{\mathbf{E}} \frac{1}{2}\|\mathbf{Y} - \mathbf{D}\mathbf{X}^{k+1} - \mathbf{E}\|_F^2 + \beta\|\mathbf{E}\|_{2,1}, \quad (22)$$

which results in a column-shrinkage operator (similar to the derivation of (21)):

$$\mathbf{E}^{k+1}(:,i) = \begin{cases} \frac{\|\mathbf{Q}(:,i)\|_2 - \beta}{\|\mathbf{Q}(:,i)\|_2}\mathbf{Q}(:,i) & \text{if } \beta < \|\mathbf{Q}(:,i)\|_2 \\ 0 & \text{otherwise} \end{cases},$$

where $\mathbf{Q} = \mathbf{Y} - \mathbf{D}\mathbf{X}^{k+1}$ and $\mathbf{Q}(:,i)$ is the *i*-th column of $\mathbf{Q}$.

5. PERFORMANCE EVALUATION

The purpose of this section is to show the usefulness of the proposed approach, by demonstrating[2] it on two very different real life problems: The SMV mode of Algorithm 1 is utilized to detect irregular heartbeats in ECG signal; and the MMV mode of Algorithm 1 is utilized for the image processing task of specular reflectance removal from natural images. The simulations were performed on an i7 quad-core computer with 8GB of RAM memory.

[2]All the results in this paper are reproducible with a MATLAB package that is freely available for distribution.

5.1. Arrhythmia Detection in ECG Signals

Irregular heartbeats, know as *Arrhythmia*, is a collection of several types of abnormal cardiac electrical activity. Arrhythmia is detected by analyzing ECG, which is a non-invasive technique for monitoring cardiac electrical activity. The durations of ECG recordings often reach 24 hours, which promoted research efforts for automatic Arrhythmia detection algorithms, see for example [11, 12]. In this experiment we focused on the detection of one type of Arrhythmia event: a Premature Ventricular Contraction (PVC), which is demonstrated in Fig. 2. Sparse representations have been proposed by [13] for ECG source separation problems, which motivated the utilization of the proposed approach for Arrhythmia detection: given an ECG signal that contains mostly normal heartbeats, the key idea is to decompose the signal into all possible N-samples windows (on the order of 1 second duration) and train a dictionary that will provide a sparse representation for these windows. Note that due to the multiplicity and periodicity of normal heartbeats, their corresponding windows are highly repetitive, and constitute the majority among all windows. The dictionary is expected to enable an accurate sparse representation for the windows that correspond to normal heartbeats, due to their high contribution to the training stage. However, the windows that correspond to Arrhythmia events are not expected to be accurately represented by this dictionary, due to their significant deviation from the normal heartbeats waveforms and their low contribution to the training stage (due to rareness of such events). Therefore, a possible strategy for Arrhythmia detection is to solve problem (7) for the SMV case, since each window is expected to be sparsely represent by a different combination of dictionary atoms, and mark columns of $\mathbf{E}$ with an l_2-norm above a threshold τ as irregular heartbeats locations.

We validated our approach using the MIT-BIH Arrhythmia Database [14] that contains a collection of 30 minutes fully annotated ECG recordings, sampled at 360Hz. We analyzed ECG recording $\#109$, which includes[3] 40 PVC events and 2492 regular heartbeats, by extracting all possible 256-samples windows, leading to initial signal collection dimensions of $256 \times 647,745$. Due to normal sinus rhythm variations in this recording between 77 to 101, this collection was divided into six segments of five minutes that were processed independently: the dimension of all windows in a segment was reduced from 256 to 32 by projection onto the 32 leading PCA basis vectors of the segment, and an over-complete dictionary $\mathbf{D} \in \mathbb{R}^{32\times 128}$ was trained using the K-SVD [15] algorithm for each segment. The SMV mode of Algorithm 1 was employed for each segment with the following parameters: $\mu^0 = 1.0, \rho = 1.25, \alpha = 1.0, \beta = 2.6, \epsilon = 0.0025$, and Arrhythmia events were detected as column in $\mathbf{E}$ with l_2-norm above a threshold $\tau = 0.1$ (the processing time for the 30 minutes recording was 176 seconds). Fig. 3 depicts $\|\mathbf{E}(:,i)\|_2$ for the first 15 minutes (formed by concatenation of the results from the first 3 segments), demonstrating accurate matching between most of the non-zero l_2-norm columns and the ground truth annotations of this recording. Due to the randomness of the initial dictionary used in the K-SVD algorithm, the entire experiment was repeated 10 times, resulting in an average of 97.18% true positive detections with standard deviation 1.89%, and an average of 2.82% false negatives with standard deviation 1.89%. Additional 13 non-PVC events were detected on average, which corresponded to noise and waveform distortions.

[3] http://www.physionet.org/physiobank/database/html/mitdbdir/records.htm

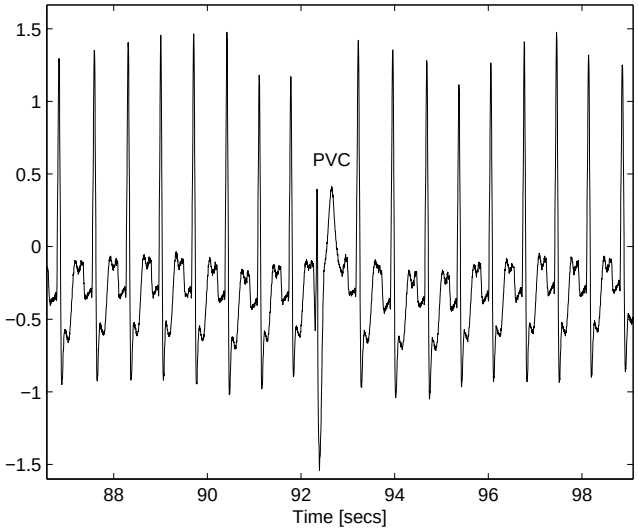

Fig. 2. A Premature Ventricular Contraction (PVC) event interrupts a series of normal heartbeats in ECG recording $\#109$ from the MIT-BIH Arrhythmia Database [14].

5.2. Specular Reflectance Removal From Natural Images

The reflection of light from surfaces is generally associated with two main components [16]: diffuse and specular. The diffuse component scatters light uniformly in all directions, regardless of the angle of incident light, whereas the specular component scatters light in a direction that depends on the angle of incident light. Light energy due to specular reflections is often concentrated, causing strong bright regions to appear in the image, as demonstrated in Fig. 5 (left column). These bright regions can cause computer vision algorithms such as segmentation, shape from shading, stereo, and motion detection to produce errors. Therefore, there has been significant interest in specular reflectance removal, see [17] for a review. Given a collection of images of a diffuse object, photographed from the same view-point and under varying illumination conditions, the diffuse component of the light is well approximated by a low-dimensional subspace [18, 19] that can be computed from the PCA basis of the images. However, specular components are not represented by this subspace.

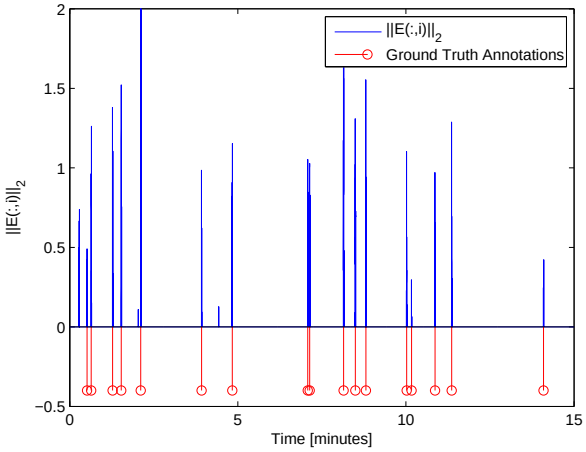

Fig. 3. Arrhythmia detection: columns of $\mathbf{E}$ with an l_2-norm above $\tau = 0.1$ indicate an ECG anomaly.

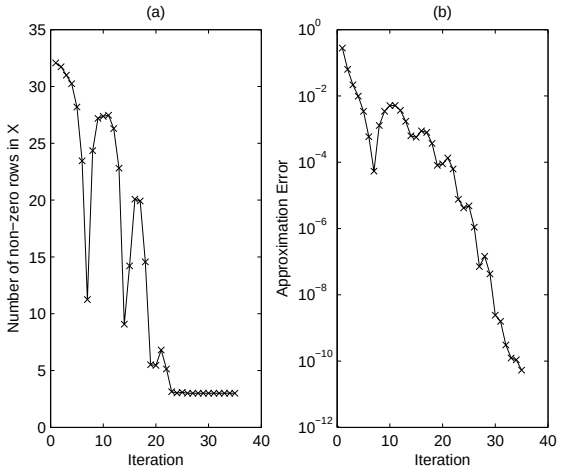

Fig. 4. Specular reflectance removal convergence: (a) number of non-zero rows in $\mathbf{X}$, (b) approximation error $\frac{\|\mathbf{Z}^k-\mathbf{X}^k\|_F^2}{\|\mathbf{X}^k\|_F^2}$.

Therefore, we can solve problem (7) with the MMV mode, in order to decompose the images $\mathbf{Y}$ (each column of $\mathbf{Y}$ is one column-stacked image) into diffuse components $\mathbf{DX}$ and specular components $\mathbf{E}$ as follows: the diffuse components of all images are expected to be jointly-sparse with respect to the PCA basis $\mathbf{D}$, whereas the specular components are assumed to appear in a subset of the images, thus, by minimizing $\|\mathbf{E}\|_{2,1}$ the columns of $\mathbf{E}$ would contain those parts of the images that do not conform with the joint-sparsity model. In our experiment we used a collection of 37 images (195×317 pixels) of a wrist watch, photographed from the same viewpoint and using 37 different illumination conditions. We computed the PCA basis of $\mathbf{Y} \in \mathbb{R}^{61,815\times 37}$ and used it as the dictionary $\mathbf{D}$. We further employed Algorithm 1 and set $p = 2, \alpha = 4.5, \beta = 0.5, \mu = 0.05, \rho = 1.15, \epsilon = 10^{-10}$. Fig. 4 presents convergence of the algorithm within 35 iterations (processing time was 22 seconds) to a joint-sparsity pattern with 3 non-zero rows (a 3-dimensional subspace). Fig. 5 presents specular reflectance removal results (best viewed in the electronic version of this paper) for five images: the obtained diffuse components (equal to $\mathbf{DX}(:,i)$, where i is the corresponding index of each input image) are clearly free of specular reflections.

6. CONCLUSIONS

Sparse coding and anomaly detection are important tasks, with numerous signal processing applications. This paper presented a unified approach for simultaneous sparse coding and anomaly detection for both jointly-sparse and independently-sparse signal models. The usefulness of the proposed approach was demonstrated for two challenging real-life problems: Arrhythmia detection in ECGs and specular reflectance removal from natural images. Due to the constantly growing number of signals that are well modeled by sparse representations, the proposed approach could be combined in many existing and emerging applications.

7. APPENDIX

The update equation for $\mathbf{X}^{k+1}$ is obtained by solving:

$$\min_{\mathbf{X}} \frac{1}{2}\|\mathbf{Y}-\mathbf{DX}-\mathbf{E}^k\|_F^2 + <\mathbf{M}^k, \mathbf{Z}^k-\mathbf{X}> + \frac{\mu^k}{2}\|\mathbf{Z}^k-\mathbf{X}\|_F^2. \tag{23}$$

The solution of (23) is computed from:

$$\frac{\partial}{\partial \mathbf{X}}\Big(\frac{1}{2}\mathbf{Tr}\{(\mathbf{Y}-\mathbf{DX}-\mathbf{E}^k)(\mathbf{Y}-\mathbf{DX}-\mathbf{E}^k)^T\} + \mathbf{Tr}\{(\mathbf{Z}^k-\mathbf{X})^T\mathbf{M}^k\} + \frac{\mu^k}{2}\mathbf{Tr}\{(\mathbf{Z}^k-\mathbf{X})(\mathbf{Z}^k-\mathbf{X})^T\}\Big) = 0, \tag{24}$$

which results in:

$$\mathbf{D}^T(\mathbf{Y}-\mathbf{DX}-\mathbf{E}^k) + \mathbf{M}^k + \mu^k(\mathbf{Z}^k-\mathbf{X}) = 0, \tag{25}$$

and the update equation is given by:

$$\mathbf{X}^{k+1} = (\mathbf{D}^T\mathbf{D}+\mu^k\mathbf{I})^{-1}(\mathbf{D}^T(\mathbf{Y}-\mathbf{E}^k) + \mathbf{M}^k + \mu^k\mathbf{Z}^k). \tag{26}$$

8. REFERENCES

[1] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM Computing Surveys*, vol. 41, no. 3, 2009.

[2] A.M. Bruckstein, D. Donoho, and M. Elad, "From sparse solutions of systems of equations to sparse modeling of signals and images," *SIAM Rev.*, vol. 51, no. 1, 2009.

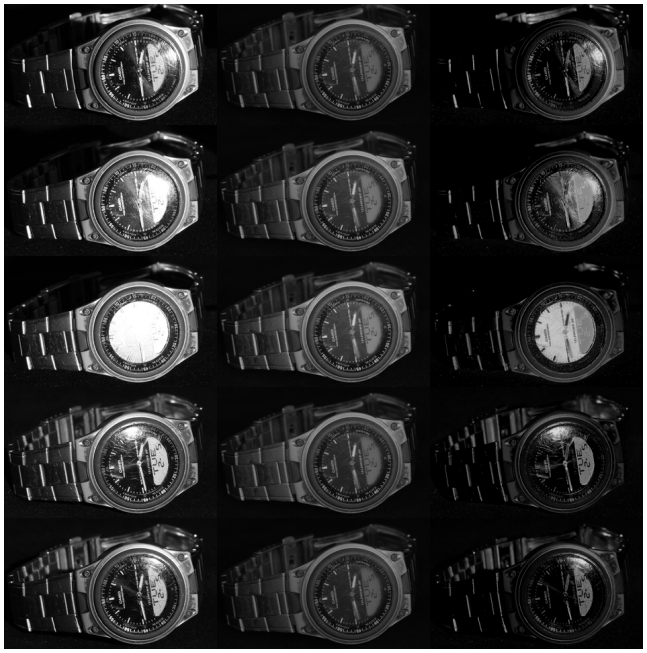

Fig. 5. Specular reflectance removal: input images (left), diffuse components (center) and specular components (right).

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

[4] S. Shekhar, V.M. Patel, M. Nasrabadi, and R. Chellappa, "Joint sparsity-based robust multimodal biometrics recognition," in *The 12th European Conference on Computer Vision (ECCV)*, 2012.

[5] Y. Cong, J. Yuan, and J. Liu, "Sparse reconstruction cost for abnormal event detection," in *Computer Vision and Pattern Recognition (CVPR)*, 2011.

[6] B. Zhao, L. Fei-Fei, and E.P. Xing, "Online detection of unusual events in videos via dynamic sparse coding," in *Computer Vision and Pattern Recognition (CVPR)*, 2011.

[7] M. Levorato and U. Mitra, "Fast anomaly detection in smartgrids via sparse approximation theory," in *Sensor Array and Multichannel Signal Processing Workshop (SAM)*, 2012.

[8] S.F. Cotter, B.D. Rao, K. Engan, and K. Kreutz-Delgado, "Sparse solutions to linear inverse problems with multiple measurement vectors," *Signal Processing, IEEE Transactions on*, vol. 53, no. 7, 2005.

[9] D. Kong, C. Ding, and H. Huang, "Robust nonnegative matrix factorization using l21-norm," in *The 20th ACM international conference on Information and knowledge management(CIKM)*, 2011.

[10] 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, 2010.

[11] T. Ince, S. Kiranyaz, and M. Gabbouj, "A generic and robust system for automated patient-specific classification of ecg signals," *Biomedical Engineering, IEEE Transactions on*, vol. 56, no. 5, 2009.

[12] C. Ye, B.V. Kumar, and M. T. Coimbra, "Heartbeat classification using morphological and dynamic features of ecg signals," *Biomedical Engineering, IEEE Transactions on*, vol. 59, no. 10, 2012.

[13] B. Mailhe, R. Gribonval, F. Bimbot, M. Lemay, P. Vandergheynst, and J. M Vesin, "Dictionary learning for the sparse modelling of atrial fibrillation in ecg signals," in *Internatioanl Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2009.

[14] G.B. Moody and R.G. Mark, "The impact of the mit-bih arrhythmia database," *Engineering in Medicine and Biology Magazine, IEEE*, vol. 20, no. 3, 2001.

[15] M. Aharon, M. Elad, and A.M. Bruckstein, "K-svd: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. on Signal Processing*, vol. 54, no. 11, 2006.

[16] R. Szeliski, *Computer Vision: Algorithms and Applications*, Springer-Verlag New York, Inc., 2010.

[17] A. Artusi, F. Banterle, and D. Chetverikov, "A survey of specularity removal methods," *Computer Graphics Forum*, vol. 30, no. 11, 2011.

[18] R. Basri and D.W. Jacobs, "Lambertian reflectance and linear subspaces," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 25, no. 2, 2003.

[19] A. Shashua, "On photometric issues in 3d visual recognition from a single 2d image," *International Journal of Computer Vision*, vol. 21, no. 1-2, 1997.