J Sign Process Syst (2015) 79:179–188
DOI 10.1007/s11265-014-0913-0

Sparse Coding with Anomaly Detection

Amir Adler · Michael Elad · Yacov Hel-Or · Ehud Rivlin

Received: 19 January 2014 / Revised: 26 May 2014 / Accepted: 5 June 2014 / Published online: 12 July 2014

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) as well as for specular reflectance and shadows removal from natural images.

Keywords Sparse coding · Anomaly detection · ADMM · Arrythmia detection · Specular reflectance removal · Shadows removal

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.

A. Adler (✉) · M. Elad · E. Rivlin
Computer Science Department, Technion, Haifa, Israel
e-mail: adleram@cs.technion.ac.il

M. Elad
e-mail: elad@cs.technion.ac.il

E. Rivlin
e-mail: ehudr@cs.technion.ac.il

Y. Hel-Or
Computer Science Department, Interdisciplinary Center, Herzelia, Israel
e-mail: toky@idc.ac.il

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, and 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 and shadows 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:

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

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:

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

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 (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:

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

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 $p = 2$.

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:

$$\mathbf{Y} = \mathbf{D}\mathbf{X} + \mathbf{E} + \mathbf{V}, \tag{4}$$

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:

$$\begin{aligned} \{\mathbf{X}, \mathbf{E}\} = \arg\min_{\mathbf{X},\mathbf{E}} &\|\mathbf{Y} - \mathbf{D}\mathbf{X} - \mathbf{E}\|_F^2 \\ \text{s.t. } &\|\mathbf{X}\|_0 \le LT_0 \\ &\|\mathbf{E}\|_{2,0} \le K_0, \end{aligned} \tag{5}$$

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:

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

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

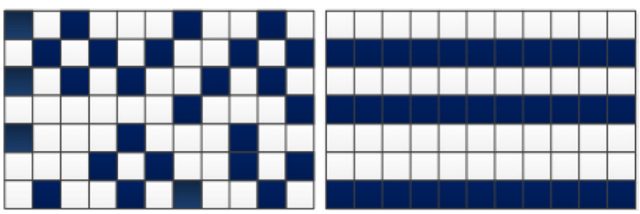

Figure 1 The non-zeros (dark squares) of the sparse representations matrix $\mathbf{X}$ for the SMV (left) and MMV (right) models.

4 The Proposed Approach

The solutions to problems[1] (5) and (6) can be approximated by solving the following unconstrained convex problem:[2]

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

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, 10] for robust non-negative matrix factorization. We propose to solve problem (7) with the Alternating Direction Method of Multipliers (ADMM) [11] due to the following reasons: (i) it is suitable for our problem format, (ii) it has proven convergence properties, and (iii) it leads to a simple, coordinate-descent structure. In the following we describe the ADMM 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:

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

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:

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

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 $\mathcal{L}_A(\mathbf{X},\mathbf{Z},\mu,\mathbf{M})$ is performed iteratively, while alternating between the minimizations of $\mathbf{X}$ and $\mathbf{Z}$:

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

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

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

$$\mu^{k+1} = \rho\mu^k;\ \rho>1. \tag{13}$$

ADMM can be extended to more than two variables, and its convergence[3] properties are analyzed in [11].

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:

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

Note that by converting Eq. (7) into a constrained problem, we have decoupled the first and second terms of Eq. (7), thus, avoiding the need for an iterated-shrinkage [17] solution for $\mathbf{X}$. The addition of the auxiliary variable $\mathbf{Z}$ results in a closed-form solution for $\mathbf{X}$ and a one-shot shrinkage solution for $\mathbf{Z}$. The Augmented-Lagrangian of problem (14) is given by:

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

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):

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

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

$$\begin{aligned}\mathbf{Z}^{k+1} &= \arg\min_{\mathbf{Z}} \alpha\|\mathbf{Z}\|_{1,1}+<\mathbf{M}^k,\mathbf{Z}-\mathbf{X}^{k+1}> \\ &\quad +\frac{\mu^k}{2}\left\|\mathbf{Z}-\mathbf{X}^{k+1}\right\|_F^2,\end{aligned} \tag{17}$$

which can be simplified to:

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

[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.

[2] Note that the related problem of a sparsely represented interference matrix $\mathbf{E} = \Psi\mathbf{C}$, where Ψ is a dictionary and $\mathbf{C}$ is sparse, can be formulated as follows:

$$\min_{\mathbf{X},\mathbf{C}} \frac{1}{2}\|\mathbf{Y}-\mathbf{D}\mathbf{X}-\Psi\mathbf{C}\|_F^2+\alpha\|\mathbf{X}\|_{1,p}+\beta\|\mathbf{C}\|_{1,1},$$

and its solution can be also obtained using the proposed approach in this section.

[3] Note that problem (7) is convex, therefore ADMM will approach the global minimum. Since the Frobenius-norm term in (7) is not separable, perfect convergence is not guaranteed, however, it was verified experimentally that the solution obtained by ADMM is sufficiently accurate after 50 iterations.

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 $k=0$, $\mathbf{Z}^0$, $\mathbf{E}^0$, $\mathbf{M}^0$, μ^0, ρ, ϵ.
Repeat Until Convergence:

1. $\mathbf{X}^{k+1}=\arg\min_{\mathbf{X}}\mathcal{L}_p\left(\mathbf{X},\mathbf{Z}^k,\mathbf{E}^k,\mathbf{M}^k,\mu^k\right)$.
2. $\mathbf{Z}^{k+1}=\arg\min_{\mathbf{Z}}\mathcal{L}_p\left(\mathbf{X}^{k+1},\mathbf{Z},\mathbf{E}^k,\mathbf{M}^k,\mu^k\right)$.
3. $\mathbf{E}^{k+1}=\arg\min_{\mathbf{E}}\mathcal{L}_p\left(\mathbf{X}^{k+1},\mathbf{Z}^{k+1},\mathbf{E},\mathbf{M}^k,\mu^k\right)$.
4. $\mathbf{M}^{k+1}=\mathbf{M}^k+\mu^k\left(\mathbf{Z}^{k+1}-\mathbf{X}^{k+1}\right)$.
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: $\mathbf{X}^k$, $\mathbf{E}^k$.

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 [11]:

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

where:

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

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

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

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

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

where $\mathbf{P}(j,:)$ is the j-th row of $\mathbf{P}$.

The update equation of $\mathbf{E}^{k+1}$ is obtained from:

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

which results in a column-shrinkage operator (similar to the derivation of Eq. (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 $\mathbf{Q}$.

4.3 Computational Complexity

The complexity of the proposed approach depends linearly on the number of signals L and polynomially on the number of atoms M. The complexity of a single ADMM iteration is given by $O(M^3)+O(MNL)+O(ML)+O(M^2L)+O(ML)+O(NL)$, where the leading four terms comprise the complexity of the update step of $\mathbf{X}^{k+1}$, the fifth term is the complexity of the update equation of $\mathbf{Z}^{k+1}$, and the sixth term is the complexity of the update equation of $\mathbf{E}^{k+1}$. Figure 2 depicts the measured complexity of the proposed approach (SMV mode) for a varying number of signals between $L=50,000$ to $L=500,000$ (using $N=32$ and $M=128$), demonstrating the linear dependence on the number of signals .

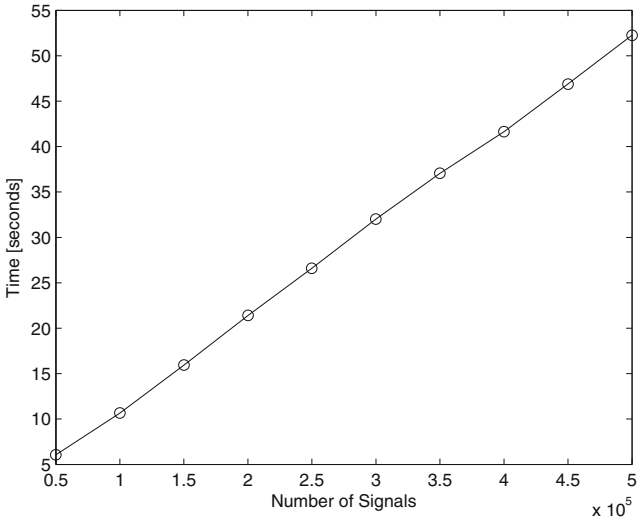

Figure 2 Computation time of the proposed approach (SMV mode) for a varying number of signals between 50,000 to 500,000.

5 Performance Evaluation

The purpose of this section is to show the usefulness of the proposed approach, by demonstrating[4] 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 and shadows removal from natural images. The simulations were performed on an i7 quad-core computer with 8GB of RAM memory.

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 h, which

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

promoted research efforts for automatic Arrhythmia detection algorithms, see for example [12, 13]. In this experiment we focused on the detection of one type of Arrhythmia event: a Premature Ventricular Contraction (PVC), which is demonstrated in Fig. 3. Sparse representations have been proposed by [14] 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 s 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 [15] that contains a collection of 30 min fully annotated ECG recordings, sampled at 360 Hz. We analyzed ECG recording #109, which includes [5] 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 5 min 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 [16] algorithm for each segment,[6] as demonstrated in Fig. 4. 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 min recording was 176 s). Figure 5 depicts $\|\mathbf{E}(:, i)\|_2$ for the first

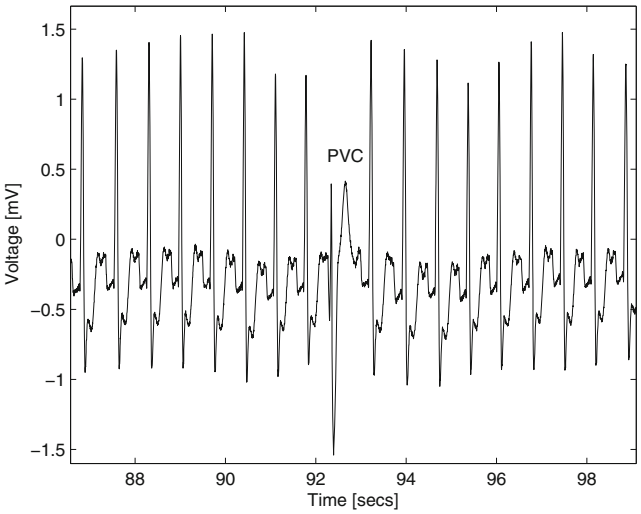

Figure 3 A Premature Ventricular Contraction (PVC) event interrupts a series of normal heartbeats in ECG recording #109 from the MIT-BIH Arrhythmia Database [15].

15 min (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. In order to demonstrate the effectiveness of the proposed approach, we repeated the same experiment using independent sparse coding per each ECG window: we employed the Orthogonal Matching Pursuit (OMP) algorithm to reconstruct each dimensionality-reduced window using the trained dictionary (with a fixed number of atoms). A window was marked as an outlier if its reconstruction error exceeded a threshold τ_{OMP}. We set the threshold value to achieve (approximately) the same true positive rate of the proposed approach, and obtained the following result: an average (over 10 experiments) of 98.46 % true positive detections with standard deviation 1.79 %, and 55.2 non-PVC events. Namely, the OMP-based approach resulted in a significantly higher false positive rate, compared to the proposed approach.

5.2 Specular Reflectance and Shadows Removal from Natural Images

The reflection of light from surfaces is associated with two main components [18]: diffuse and specular. The diffuse component scatters light uniformly in all directions, whereas the specular component scatters light in a direction

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

[6] A total of six dictionaries were trained—each using all of the dimensionality-reduced samples from each segment.

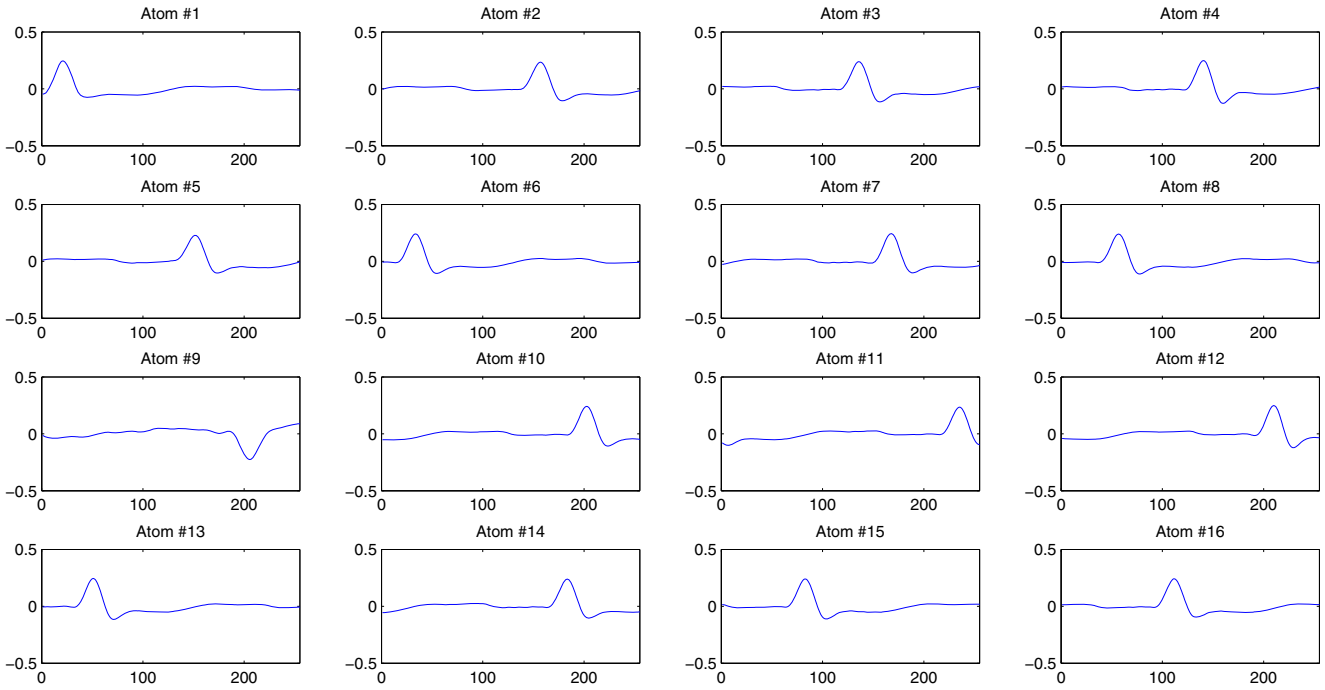

Figure 4 ECG dictionary:16 atoms of one ECG segment, displayed after reconstruction using the 32 leading PCA basis vectors of the segment.

that depends on the angles of incident light and the surface normal. Light energy due to specular reflections is often concentrated, causing strong bright regions (highlights) to appear in the image, as demonstrated in Fig. 7 (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 [19] for a review. According to *Phong* shading model [18], the intensity of the diffuse component at image pixel k is given by:

$$i_k(\lambda) = L(\lambda)\rho_k(\lambda)\max(0, \hat{\mathbf{n}}_k \cdot \hat{\mathbf{v}}), \tag{23}$$

where λ is the wavelength (color), $L(\lambda)$ is the intensity profile of incident light, $\rho(\lambda)$ is the *albedo*, $\hat{\mathbf{n}} =$

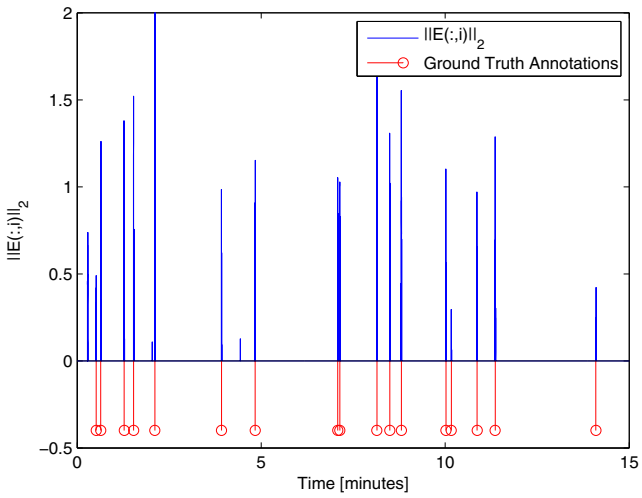

Figure 5 Arrhythmia detection: columns of $\mathbf{E}$ with an l_2-norm above $\tau = 0.1$ indicate an ECG anomaly.

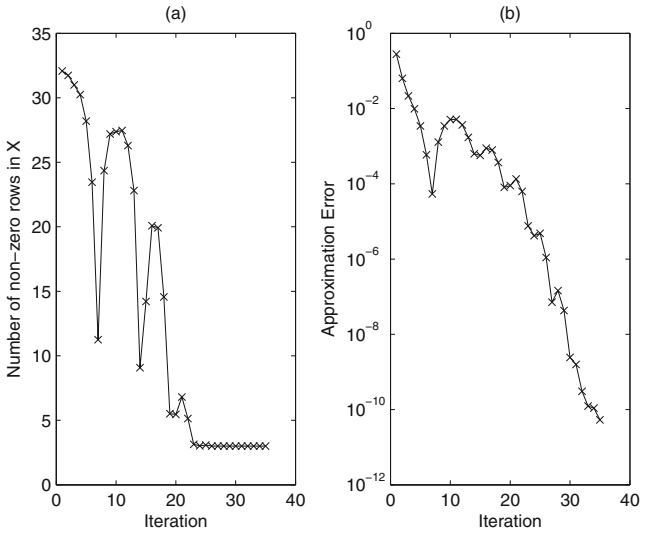

Figure 6 Specular reflectance removal convergence: **a** number of non-zero rows in $\mathbf{X}$, **b** approximation error $\frac{\|\mathbf{Z}^k - \mathbf{X}^k\|_F^2}{\|\mathbf{X}^k\|_F^2}$.

$[n_x, n_y, n_z]^T$ is the surface normal, and $\hat{\mathbf{v}} = [v_x, v_y, v_z]^T$ is a unit vector pointing to the direction of incident light. Equation (23) is interpreted as follows: The measured intensity is given by the product of the source intensity, the albedo and the cosine of the angle θ_i between the surface normal and direction of incident light. In the case that $|\theta_i| > \pi/2$ the intensity equals zero, which results in a self-shadow[7] effect. By column-stacking m pixels, and neglecting the self shadowing effect (i.e. allowing $|\theta_i| > \pi/2$), the following matrix formulation is obtained:

$$\mathbf{i}(\lambda) = \mathbf{N}(\lambda)\hat{\mathbf{v}} \in \mathbb{R}^{m\times 1}, \tag{24}$$

where

$$\mathbf{i}(\lambda) = \begin{bmatrix} i_1(\lambda) \\ i_2(\lambda) \\ \vdots \end{bmatrix} \in \mathbb{R}^{m\times 1}, \mathbf{N}(\lambda) = \begin{bmatrix} L(\lambda)\rho_1(\lambda)\hat{\mathbf{n}}_1^T \\ L(\lambda)\rho_2(\lambda)\hat{\mathbf{n}}_2^T \\ \vdots \end{bmatrix} \in \mathbb{R}^{m\times 3}. \tag{25}$$

Given a collection of K images of a diffuse object, photographed from the same view-point and under varying light source directions, the following rank-3 model is obtained:

$$\mathbf{I}(\lambda) = \mathbf{N}(\lambda)\mathbf{V}, \tag{26}$$

where

$$\mathbf{I}(\lambda) = \begin{bmatrix} \mathbf{i}_1(\lambda) \\ \mathbf{i}_2(\lambda) \\ \vdots \end{bmatrix} \in \mathbb{R}^{m\times K}, \mathbf{V} = [\hat{\mathbf{v}}_1, \cdots, \hat{\mathbf{v}}_K] \in \mathbb{R}^{3\times K}. \tag{27}$$

Therefore, the diffuse component can be modeled by a low-dimensional subspace, and the works [20, 21] proved that the dimension of this subspace is upper bounded by 9. The basis for this subspace can be computed from the PCA basis of the images. However, specular components and shadows are not represented by this subspace. 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 and shadows $\mathbf{E}$ 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 and shadows 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 view-point 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 $p = 2, \alpha = 4.5, \beta = 0.5, \mu = 0.05, \rho = 1.15, \epsilon = 10^{-10}$. Figure 6 presents convergence of the algorithm within 35 iterations (processing time was 22 s) to a joint-sparsity pattern with 3 non-zero rows (a 3-dimensional subspace). Figure 7 presents specular reflectance removal results (best viewed in the electronic version of this paper) for five images: the obtained diffuse components (equal to $\mathbf{DX}(:, l)$, where l is the corresponding index of each input image) are free of specular reflections and the shadows are significantly removed.

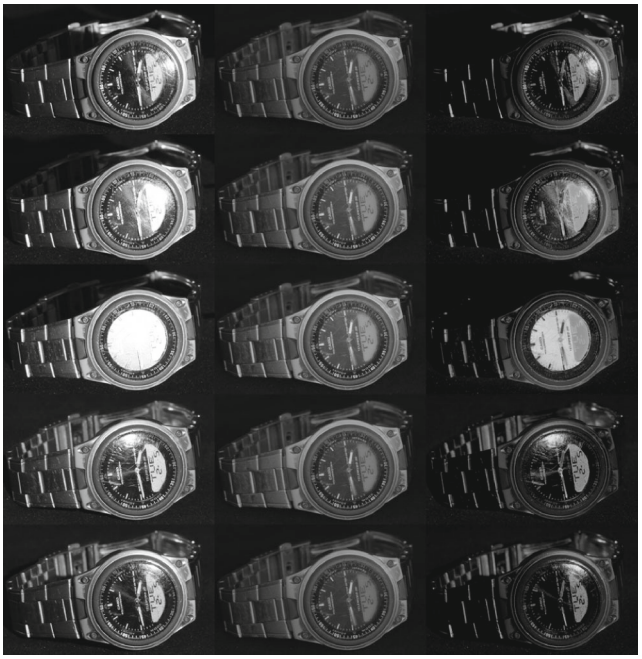

Figure 7 Specular reflectance removal: input images (*left*), diffuse components (*center*) and specular components (*right*).

For the case of color images, the low-dimensional subspace model still holds since for a certain image pixel, only the albedo and incident light intensity are color (wavelength) dependent. Using an RGB color representation we obtain the following model:

$$\mathbf{i}_{RGB} = \begin{bmatrix} \mathbf{i}(R) \\ \mathbf{i}(G) \\ \mathbf{i}(B) \end{bmatrix} = \begin{bmatrix} \mathbf{N}(R)\hat{\mathbf{v}} \\ \mathbf{N}(G)\hat{\mathbf{v}} \\ \mathbf{N}(B)\hat{\mathbf{v}} \end{bmatrix} = \begin{bmatrix} \mathbf{N}(R) \\ \mathbf{N}(G) \\ \mathbf{N}(B) \end{bmatrix} \hat{\mathbf{v}} \in \mathbb{R}^{3m\times 1}. \tag{28}$$

[7] This is in contrast to cast-shadows, where one part of an object is shadowed by another part.

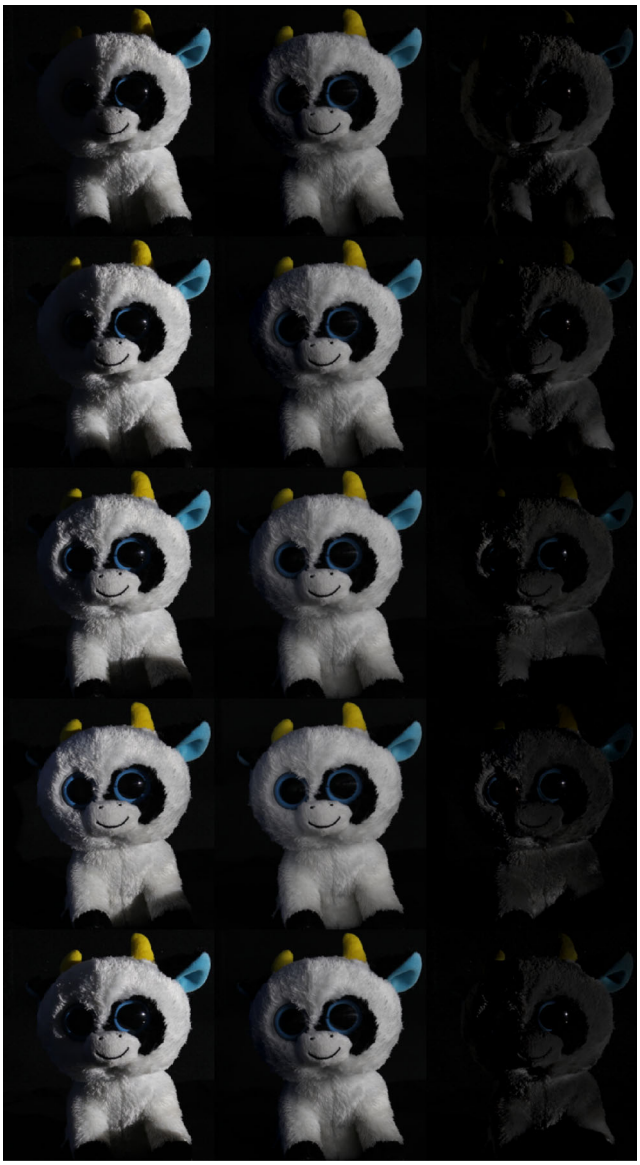

Figure 8 Specular reflectance and shadows removal from color images: input images (*left*), diffuse components (*center*) and specular components (*right*).

Given a collection of K color images of a diffuse object, photographed from the same view-point and under varying light source directions, a rank-3 model is obtained, similar to Eq. (26). Figure 8 demonstrates specular reflectance and shadows removal results for color images: the obtained diffuse components are free of specular reflections and the shadows are significantly removed.

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.

Appendix

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

$$\min_{\mathbf{X}} \frac{1}{2} \left\| \mathbf{Y} - \mathbf{D}\mathbf{X} - \mathbf{E}^k \right\|_F^2 + < \mathbf{M}^k, \mathbf{Z}^k - \mathbf{X} > + \frac{\mu^k}{2} \left\| \mathbf{Z}^k - \mathbf{X} \right\|_F^2. \tag{29}$$

The solution of Eq. (29) is computed from:

$$\frac{\partial}{\partial \mathbf{X}} \left(\frac{1}{2} \mathbf{Tr} \left\{ \left(\mathbf{Y} - \mathbf{D}\mathbf{X} - \mathbf{E}^k \right) \left(\mathbf{Y} - \mathbf{D}\mathbf{X} - \mathbf{E}^k \right)^T \right\} + \mathbf{Tr} \left\{ \left(\mathbf{Z}^k - \mathbf{X} \right)^T \mathbf{M}^k \right\} + \frac{\mu^k}{2} \mathbf{Tr} \left\{ \left(\mathbf{Z}^k - \mathbf{X} \right) \left(\mathbf{Z}^k - \mathbf{X} \right)^T \right\} \right) = 0, \tag{30}$$

which results in:

$$\mathbf{D}^T \left(\mathbf{Y} - \mathbf{D}\mathbf{X} - \mathbf{E}^k \right) + \mathbf{M}^k + \mu^k \left(\mathbf{Z}^k - \mathbf{X} \right) = 0, \tag{31}$$

and the update equation is given by:

$$\mathbf{X}^{k+1} = \left(\mathbf{D}^T \mathbf{D} + \mu^k \mathbf{I} \right)^{-1} \left(\mathbf{D}^T \left(\mathbf{Y} - \mathbf{E}^k \right) + \mathbf{M}^k + \mu^k \mathbf{Z}^k \right). \tag{32}$$

References

1. Chandola, V., Banerjee, A., Kumar, V. (2009). Anomaly detection: a survey. *ACM Computing Surveys, 41*(3), 15:1–15:58.
2. Bruckstein, A.M., Donoho, D., Elad, M. (2009). From sparse solutions of systems of equations to sparse modeling of signals and images. *Society for Industrial and Applied Mathematics Review, 51*(1), 34–81.
3. Nguyen, N.H., Nasrabadi, N.M., Tran, T.D. (2011). Robust multi-sensor classification via joint sparse representation. In *International conference on information fusion (FUSION)*.
4. Shekhar, S., Patel, V.M., Nasrabadi, M., Chellappa, R. (2012). Joint sparsity-based robust multimodal biometrics recognition. In *The 12th European conference on computer vision (ECCV)*.

5. Cong, Y., Yuan, J., Liu, J. (2011). Sparse reconstruction cost for abnormal event detection. In *Computer vision and pattern recognition (CVPR)*.
6. Zhao, B., Fei-Fei, L., Xing, E.P. (2011). Online detection of unusual events in videos via dynamic sparse coding. In *Computer vision and pattern recognition (CVPR)*.
7. Levorato, M., & Mitra, U. (2012). Fast anomaly detection in smartgrids via sparse approximation theory. In *Sensor array and multichannel signal processing workshop (SAM)*.
8. Cotter, S.F., Rao, B.D., Engan, K., Kreutz-Delgado, K. (2005). Sparse solutions to linear inverse problems with multiple measurement vectors. *IEEE Transactions on Signal Processing, 53*(7), 2477–2488.
9. Kong, D., Ding, C., Huang, H. (2011). Robust nonnegative matrix factorization using l21-norm. In *The 20th ACM international conference on information and knowledge management (CIKM)*.
10. Dobigeon, N., & Févotte, C. (2013). Robust nonnegative matrix factorization for nonlinear unmixing of hyperspectral images. In *The 5th workshop on hyperspectral image and signal processing: evolution in remote sensing (WHISPERS)*.
11. Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J. (2010). Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning, 3*(1), 1–122.
12. Ince, T., Kiranyaz, S., Gabbouj, M. (2009). A generic and robust system for automated patient-specific classification of ecg signals. *IEEE Transactions on Biomedical Engineering, 56*(5), 1415–1426.
13. Ye, C., Kumar, B.V., Coimbra, M.T. (2012). Heartbeat classification using morphological and dynamic features of ecg signals. *IEEE Transactions on Biomedical Engineering, 59*(10), 2930–2941.
14. Mailhe, B., Gribonval, R., Bimbot, F., Lemay, M., Vandergheynst, P., Vesin, J.M. (2009). Dictionary learning for the sparse modelling of atrial fibrillation in ecg signals. In *Internatioanl conference on acoustics, speech and signal processing (ICASSP)*.
15. Moody, G.B., & Mark, R.G. (2001). The impact of the mit-bih arrhythmia database. *IEEE Engineering in Medicine and Biology Magazine, 20*(3), 45–50.
16. Aharon, M., Elad, M., Bruckstein, A.M. (2006). K-svd: an algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on Signal Processing, 54*(11), 4311–4322.
17. Beck, A., & Teboulle, M. (2009). A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Image Science*.
18. Szeliski, R. (2010). *Computer vision: algorithms and applications*. New York: Springer.
19. Artusi, A., Banterle, F., Chetverikov, D. (2011). A survey of specularity removal methods. *Computer Graphics Forum*, *30*(11).
20. Basri, R., & Jacobs, D.W. (2003). Lambertian reflectance and linear subspaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *25*(2).
21. Shashua, A. (1997). On photometric issues in 3d visual recognition from a single 2d image. *International Journal of Computer Vision*, *21*(1–2).

Amir Adler received the B.Sc. (Cum Laude) degree in 1994 and the M.Eng. (Cum Laude) degree in 2001, both in Electrical Engineering, from the Technion, Israel. He is currently pursuing a PhD degree at the Computer Science Department, Technion, Israel. From 2003 to 2006 he was Director of Signal Processing at GO Networks and from 2006 to 2008 he was Chief Technology Officer of the Wi-Fi Division at NextWave Wireless. His research interests include applications of sparse representations and convex optimization for image and image processing. Amir Adler is the recipient of the 2011 Google Europe Doctoral Fellowship in Multimedia.

Michael Elad received his B.Sc. (1986), M.Sc. (1988) and D.Sc. (1997) from the department of Electrical engineering at the Technion, Israel. Since 2003 he is a faculty member at the Computer-Science department at the Technion, and since 2010 he holds a full-professorship position. Michael Elad works in the field of signal and image processing, specializing in particular on inverse problems, sparse representations and super-resolution. Michael received the Technion's best lecturer award six times, he is the recipient of the 2007 Solomon Simon Mani award for excellence in teaching, the 2008 Henri Taub Prize for academic excellence, and the 2010 Hershel-Rich prize for innovation. Michael is an IEEE Fellow since 2012. He is serving as an associate editor for SIAM SIIMS, and ACHA. Michael is also serving as a senior editor for IEEE SPL.

Yacov Hel-Or Ph.D. in Computer Science, the Hebrew University in Jerusalem. He has been a faculty member at the school of computer science, the Interdisciplinary Center (IDC), Herzliya, since 1998. Before joining IDC, served as a researcher at the Hewlett-Packard Israel Science Center, at the Technion, Haifa. Prior to this he held post-doctoral positions at the Weizmann Institute of Science and at the NASA Ames research center, Moffet-Field, California. His main areas of interest include computer vision, image processing, computer graphics, and robotics.

Ehud Rivlin is the director of the Center for Intelligent Systems and faculty member at the Computer Science Department at the Technion Israel Institute of Technology. Professor Rivlin has a BSc and MSc Summa Cum Laude in Computer Science from the Hebrew University in Jerusalem. He got his PhD from the University of Maryland specializing in computer vision. Prof. Rivlin has about 200 papers published in professional journals and conferences. He served on various editorial boards and conference committees. Prof. Rivlin has been working on image understanding and vision based navigation (VBN) for the last 20 years. Within his work he set a new framework for object classification and developed a family of sensory-based navigation algorithms for robots. Some of Prof Rivlins research interests are within biologically motivated vision, where one tries to get inspiration from biological systems to build better vision systems and algorithms while trying, in some cases, to provide insight and suggest models for the computational processes underlying those abilities of biological systems. Working in this domain Prof Rivlin developed mathematical models of various animal visual behaviors.