Object Detection on Compressive Measurements using Correlation Filters and Sparse Representation

Hector Vargas
Dept. of Electrical Engineering
Universidad Industrial de Santander
Bucaramanga, Colombia

Yesid Fonseca
Dept. of Electrical Engineering
Universidad Industrial de Santander
Bucaramanga, Colombia

Henry Arguello
Dept. of Computer Science
Universidad Industrial de Santander
Bucaramanga, Colombia

***Abstract*—Compressive cameras acquire measurements of a scene using random projections instead of sampling at Nyquist rate. Several reconstruction algorithms have been proposed, taking advantage of previous knowledge about the scene. However, some inference tasks require to determine only certain information of the scene without incurring in the high computational reconstruction step. By reducing the computation load related to the reconstruction problem, this paper proposes a computationally efficient object detection approach based on correlation filters and sparse representation that operate over compressive measurements. We consider the problem of object detection in remote sensing scenes with multi-band images, where the pixels are expensive. The correlation filters are designed using explicit knowledge of the target appearance and the target shape to provide a way to recognize the objects from compressive measurements. Numerical experiments show the validity and efficiency of the proposed method in terms of peak-to-side lobe ratio using simulated data.**

***Index Terms*—Object detection, compressive measurements, correlation filters, sparse representation**

I. Introduction

Compressed sensing (CS) is a signal processing technique to effectively recover a sparse signal, by solving an underdetermined linear systems [1]. CS has implications in imaging systems to reduce the number of measurements, power consumption, and storage space. The single-pixel architecture illustrates a physical CS system for imaging where the sampling matrices are created using a Digital Micromirror Device (DMD) [2] and time-resolved systems [3]. When the measurements are acquired, an inverse problem is solved by using convex optimization in order to reconstruct the image. Several works have focused on optimization algorithms for signal reconstruction and less attention has been considered to perform inference directly on the compressive measurements. In many practical scenarios, making a decision about an image is most efficient rather than computing a reconstruction. Furthermore, high-quality reconstruction is difficult at high compression ratios due to several parameters such as signal sparsity, sparsifying basis, noise, etc.

In this work, the reconstruction problem is bypassed since only a detection task is required. Specifically, object location can be directly estimated from compressive measurements by using Correlation Filters (CFs) [4]. CFs provide a solid foundation to object recognition due to they can detect objects through rotations translations and other distractions. Also, this work consider the application of remote sensing which provides high-resolution aerial images on demand. In remote sensing, it is common to collect an enormous amount of data, followed by the transmission of data to a ground station over a low bandwidth communication channel.

In [5]an object compressive classification method is proposed by recovering parameters directly from the compressed measurements. [6] theoretically showed that learning directly in compressed domain is possible and that, with high probability, the linear kernel Super Vector Machine (SVM) classifier can be as accurate as the best linear threshold classifier in the data domain. In [7] a sampling strategy to obtain the optical flow in video application is proposed. Recently, [8] proposed a method for action recognition from compressive cameras without reconstruction at high compression ratios by using nonlinear features. In this work, a sparse-based framework for object detection and localization directly from compressed measurements was proposed, thus avoiding the costly reconstruction process. In this paper, a training procedure to learn CFs is proposed in order to reconstruct the object location directly in the compressed .

II. Proposed Object Detection

The proposed object detection scheme consists of three steps: learning CFs, compressive acquisition and sparse optimization, as sketched in Figure 1. In this section, we first introduce the background of a compressive camera in Subsection II-A, then the learning CFs scheme to multi-band images in Subsection II-B. The computational advantage of the proposed sparse optimization problem is analyzed in Subsection II-C.

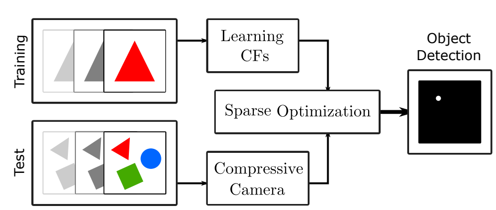

Fig. 1. Flow of the proposed approach. Three steps make up the proposed scheme, learning CFs, compressive acquisition and sparse optimization

A. *Compressive Camera*

Several camera designs have been proposed for compressive image acquisition. A common element among these is the use

ISBN 978-90-827970-1-5 © EURASIP 2018

of a random patterns to modulate the spatial information to the measurement collection. As an example, the Single Pixel Camera (SPC) [2] uses few optical blocks and number of random pixels that are projected into a single intensity detector. This setup has been extended to multi-band images, by combining the single intensity detector with spectral filtering. This system can be modeled as a linear mapping, where all pixels i, j of the image $x_{(i,j,k)}$ in each spectral band k are mapped to an lth single point $y_{(l,k)}$. This is expressed mathematically as

$$y_{(l,k)} = \sum_{i,j} \phi_{(l,i,j,k)} x_{(i,j,k)} + \omega_{(l,k)}, \tag{1}$$

where $\phi(l, i, j, k) \in \{-1, 1\}$ with $1 \leq i \leq I$, $1 \leq j \leq J$ and $1 \leq k \leq K$ is the discretization of modulation operator with I, J spatial dimension and K spectral dimension, $1 \leq l \leq L$ indexes the measurements and $\omega_{(l,k)}$ is additive noise in the sensor. Equation (1) can be expressed as the linear matrix-vector system

$$\begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_K \end{bmatrix} = \begin{bmatrix} \mathbf{\Phi}_1 & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{\Phi}_K \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_K \end{bmatrix} + \begin{bmatrix} \boldsymbol{\omega}_1 \\ \vdots \\ \boldsymbol{\omega}_K \end{bmatrix}, \tag{2}$$

where $\mathbf{x}_k \in \mathbb{R}^{IJ \times 1}$ is the vectorization of the kth band, $\mathbf{\Phi}_k \in \{-1, 1\}^{L \times IJ}$ represents the compressive system in the kth band, $\mathbf{y}_k \in \mathbb{R}^{L \times 1}$ denotes the L compressive measurements and $\boldsymbol{\omega}_k \in \mathbb{R}^{L \times 1}$ is the noise matrix assumed to be Gaussian with $\mathbb{E}[\boldsymbol{\omega}_k] = 0$ and $\mathbb{E}[\boldsymbol{\omega}_k \boldsymbol{\omega}_k^T] = \sigma^2 \mathbf{I}_L$. The acquisition procedure can be succinctly written as

$$\mathbf{y} = \overline{\mathbf{\Phi}} \mathbf{x} + \boldsymbol{\omega}, \tag{3}$$

where $\overline{\mathbf{\Phi}} \in \mathbb{R}^{KL \times IJK}$, $\mathbf{y} \in \mathbb{R}^{LK \times 1}$ and $\mathbf{x} \in \mathbb{R}^{IJK \times 1}$. Note that, by considering hardware implementation, $\mathbf{\Phi}_k$ must be chosen properly. Based on the Restricted Isometry Property (RIP) and Coherence [9], couples of existing structured sensing matrices can be used, which have special structures, high recovery performance, and many advantages such as simple construction, fast calculation and easy hardware implementation [10], [11]. This work uses Random Convolution (RC) sensing matrix due to its important property of Fourier domain operation [12]. In terms of RC, define $\mathbf{\Phi}_k = \mathbf{R}_k \mathbf{F}_{2D}^T \text{diag}(\hat{\mathbf{p}}_k) \mathbf{F}_{2D}$, where $\mathbf{F}_{2D} \in \mathbb{C}^{IJ \times IJ}$ is the 2D discrete Fourier matrix, $\hat{\mathbf{p}}_k = \mathbf{F}_{2D} \mathbf{p}_k$ is the 2D Fourier transform of the random vector $\mathbf{p}_k \in \mathbb{R}^{IJ \times 1}$ and $\mathbf{R}_k \in \{0, 1\}^{L \times IJ}$ is a subset of L rows from the identity matrix. The sensing matrix in Equation (2) can be modeled as

$$\overline{\mathbf{\Phi}} = \overline{\mathbf{R}} \overline{\mathbf{F}}_{2D}^T \begin{bmatrix} \text{diag}(\hat{\mathbf{p}}_1) & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \text{diag}(\hat{\mathbf{p}}_K) \end{bmatrix} \overline{\mathbf{F}}_{2D}, \tag{4}$$

where $\overline{\mathbf{R}} \in \{0, 1\}^{LK \times IJK}$ is a block diagonal matrix where each block has the sub-matrix $\mathbf{R}_k$, $\overline{\mathbf{F}}_{2D} = \mathbf{I}_K \otimes \mathbf{F}_{2D}$ with $\otimes$ being the Kronecker product and $\mathbf{I}_K \in \{0, 1\}^{K \times K}$ the identity matrix.

B. Learning Correlation Filters (CFs)

CFs are commonly used for object recognition due to that they can detect objects through rotations, translations and other distractions [4]. The multi-band CFs are extensions of CFs, which aggregates the responses of all feature bands to produce a correlation output which is constrained to have a sharp peak only at the target location. Then, the goal is to learn a multi-band CF $\mathbf{h} = [\mathbf{h}_1^T \dots \mathbf{h}_K^T]^T$ where $\mathbf{h}_k \in \mathbb{R}^{IJ \times 1}$ (k indexes the band of the CF) that optimally maps a set of S training samples $\{(\mathbf{z}_k)_{(s)}\}_{s=1,\dots,S}$ with $(\mathbf{z}_k)_{(s)} \in \mathbb{R}^{IJ \times 1}$ to the desired output $\{(\mathbf{r})_{(s)}\}_{s=1,\dots,S}$ with $(\mathbf{r}_k)_{(s)} \in \mathbb{R}^{IJ \times 1}$. Then, learning CF can be interpreted as a regression problem optimizing the localization loss defined as the Minimum Square Error (MSE) between an ideal desired correlation output $(\mathbf{r})_{(s)}$ and the training sample $(\mathbf{z}_k)_{(s)}$ as

$$\mathbf{h} = \min_{\mathbf{h}} \frac{1}{S} \sum_{s=1}^{S} \| \sum_{k=1}^{K} (\mathbf{z}_k)_{(s)} * \mathbf{h}_k - (\mathbf{r})_{(s)} \|_2^2, \tag{5}$$

where $*$ denotes the convolution operation. The Parseval's theorem can be used to express the MSE in the Fourier domain and solve (5) efficiently with respect to $\mathbf{h}$ [4]. In this work, we concentrate on the use of the Maximum a Posteriori (MAP) probability estimator to learn CFs, which regularizes the estimation process using an assumed prior distribution on the filter [13].

1) MAP-Analysis Learning Approach: Under appropriate statistical assumptions and choice of priors and models, the MAP between the desired correlation output and the training sample in 2D Fourier domain is

$$\hat{\mathbf{h}} = \min_{\hat{\mathbf{h}}} \frac{1}{S} \sum_{s=1}^{S} \| (\overline{\mathbf{Z}})_{(s)} \hat{\mathbf{h}} - (\hat{\mathbf{r}})_{(s)} \|_2^2 + \lambda \| \hat{\mathbf{h}} \|_2^2, \tag{6}$$

where $\overline{\mathbf{Z}} = [\text{diag}(\hat{\mathbf{z}}_1) \dots \text{diag}(\hat{\mathbf{z}}_K)]$, $\hat{\mathbf{z}}_k, \hat{\mathbf{h}}_k$ and $\hat{\mathbf{r}}$ are 2D Fourier representation of $\mathbf{z}_k, \mathbf{h}_k$ and $\mathbf{r}$ respectively and λ is a parameter of the prior model [13]. Note that, the optimization problem (6) is an analysis model to convolutional dictionary learning with fixed sparse vector and a regularization term to dictionary [14]. When $\mathbf{h}$ is estimated, the object location can be determined as $\sum_{k=1}^{K} \mathbf{h}_k * \mathbf{x}_k = \mathbf{u}$.

2) MAP-Synthesis Learning Approach: The proposed approach to learn CFs in this work has several similarities with the formulation used in (6). However, it is based on a synthesis model to learn a correlation filter which is used as a sparse representation basis. Let $\mathbf{H} = [\mathbf{h}_1 \dots \mathbf{h}_K]$ with $\mathbf{H} \in \mathbb{R}^{IJ \times K}$, the MAP-Synthesis option to learn the correlation filter in the 2D Fourier domain is

$$\hat{\mathbf{H}} = \min_{\hat{\mathbf{H}}} \frac{1}{S} \sum_{s=1}^{S} \| (\hat{\mathbf{R}})_{(s)} \hat{\mathbf{H}} - (\hat{\mathbf{Z}})_{(s)} \|_F^2 + \lambda \| \hat{\mathbf{H}} \|_F^2, \tag{7}$$

where $\hat{\mathbf{H}} = [\hat{\mathbf{h}}_1 \dots \hat{\mathbf{h}}_K]$, $\hat{\mathbf{R}} = \text{diag}(\hat{\mathbf{r}})$ and $\hat{\mathbf{Z}} = [\hat{\mathbf{z}}_1 \dots \hat{\mathbf{z}}_K]$. The solution to (7) becomes,

$$\hat{\mathbf{H}} = (\sum_{s=1}^{S} (\hat{\mathbf{R}})_{(s)}^H (\hat{\mathbf{R}})_{(s)} + \lambda \mathbf{I}_{IJ})^{-1} (\sum_{s=1}^{S} (\hat{\mathbf{R}})_{(s)}^H (\hat{\mathbf{Z}})_{(s)}), \tag{8}$$

where H denotes the conjugate transpose. Then, the synthesis model assumes that the signal $\mathbf{x}$ is given by

$$\mathbf{x} = \overline{\mathbf{F}}_{2D}^T \begin{bmatrix} \text{diag}(\hat{\mathbf{h}}_1) \\ \vdots \\ \text{diag}(\hat{\mathbf{h}}_K) \end{bmatrix} \mathbf{F}_{2D}\mathbf{u} = \overline{\mathbf{H}}\mathbf{u}, \quad (9)$$

where $\overline{\mathbf{H}} \in \mathbb{R}^{IJK \times IJ}$ is usually called a synthesis dictionary and $\mathbf{u}$ is the coefficients vector of $\mathbf{x}$ in $\overline{\mathbf{H}}$. Note that when $\mathbf{h}$ is correlated with the input $\mathbf{x}$, the expected output presents high sparsity structure. With this knowledge, we can assume that the vector $\mathbf{x}$ has a sparse representation in the filter matrix $\overline{\mathbf{H}}$. Then, we propose a sparse formulation for object detection, which can be used even in an uncompressed setup.

C. Sparse Representation

In this section, we formulate object detection based on sparse representation that uses a sparse basis (multi-channel CFs) to exploit the high sparsity structure of the filter response. The goal of sparse representation is to reconstruct a vector $\mathbf{u} \in \mathbb{R}^{IJ \times 1}$ from measurements $\mathbf{y} = \overline{\mathbf{\Phi}}\mathbf{x}$, where $L \ll IJ$. This idea can be generalized to the case in which $\mathbf{x}$ is sparse under a given basis $\overline{\mathbf{H}}$, so that there is a sparse vector $\mathbf{u}$ such that $\mathbf{x} = \overline{\mathbf{H}}\mathbf{u}$. Then, the acquisition model can be expressed in terms of the vector $\mathbf{u}$ by combining Equation (3) and (9) as

$$\mathbf{y} = \overline{\mathbf{\Phi}}\mathbf{x} + \boldsymbol{\omega} = \overline{\mathbf{\Phi}\mathbf{H}}\mathbf{u} + \boldsymbol{\omega} = \mathbf{A}\mathbf{u} + \boldsymbol{\omega}, \quad (10)$$

where

$$\mathbf{A} = \overline{\mathbf{R}\mathbf{F}}_{2D}^T \begin{bmatrix} \text{diag}(\hat{\boldsymbol{a}}_1) \\ \vdots \\ \text{diag}(\hat{\boldsymbol{a}}_K) \end{bmatrix} \mathbf{F}_{2D} \in \mathbb{R}^{LK \times IJ},$$

$\hat{\boldsymbol{a}}_k = \hat{\mathbf{p}}_k \circ \hat{\mathbf{h}}_k$ with $\circ$ being the Hadamard product. The sparse representation is adopted due to two reasons: It is computationally less costly, since it only involves simple operations of matrix-vector product and the object location can be easily derived in terms of sparse coefficients.

1) Sparse Optimization: The vector $\mathbf{u}$ can be estimated by introducing an auxiliary variable $\mathbf{v} \in \mathbb{R}^{IJ \times 1}$ and the Alternating Direction Method of Multiplier (ADMM) [15] framework

$$\begin{aligned} \min_{\mathbf{u},\mathbf{v}} \quad & (1/2)\|\mathbf{y} - \mathbf{A}\mathbf{u}\|_2^2 + \lambda\|\mathbf{v}\|_1 \\ \text{s.t} \quad & \mathbf{u} = \mathbf{v}, \mathbf{v} \geq 0 \end{aligned}, \quad (11)$$

where $\|\cdot\|_2^2$ and $\|\cdot\|_1$ denotes the l_2 and l_1 norms respectively and λ is a regularization parameter. The procedure is summarized in Algorithm 1. The augmented Lagrangian associated to the optimization problem (11) can be written as

$$\begin{aligned} \mathcal{L}(\mathbf{u}, \mathbf{v}, \mathbf{d}) =& (1/2)\|\mathbf{y} - \mathbf{A}\mathbf{u}\|_2^2 + \lambda\|\mathbf{v}\|_1 + \\ & \mathcal{I}_+(\mathbf{v}) + (\rho/2)\|\mathbf{u} - \mathbf{v} + \mathbf{d}\|_2^2 \end{aligned} \quad (12)$$

where $\mathbf{d} \in \mathbb{R}^{IJ \times 1}$ is the scaled dual variable, $\rho > 0$ is the weighting of the augmented Lagrangian term and $\mathcal{I}_+(\mathbf{v})$ is the indicator function that is zero if $\mathbf{v}$ belongs to the nonnegative orthant and $+\infty$ otherwise.

Algorithm 1: Sparse Optimization.

Input : $\mathbf{y}, \mathbf{A}, \lambda > 0, \rho > 0$
1: $\mathbf{v}^{(0)} = \mathbf{0}, \mathbf{d}^{(0)} = \mathbf{0}$ `// ADMM Iterations`
2: **for** $t = 1$ to MAXITER **do**
3: $\mathbf{u}^{(t)} = \min_{\mathbf{u}} \mathcal{L}(\mathbf{u}, \mathbf{v}^{(t-1)}, \mathbf{d}^{(t-1)})$
4: $\mathbf{v}^{(t)} = \max(0, \text{soft}(\mathbf{u}^{(t)} + \mathbf{d}^{(t-1)}, \lambda/\rho))$
5: $\mathbf{d}^{(t)} = \mathbf{d}^{(t-1)} + \mathbf{u}^{(t)} - \mathbf{v}^{(t)}$
6: **end for**

a) Updating $\mathbf{u}$*:* Forcing the derivative of (12) with respect to $\mathbf{u}$ to be zero leads to the following linear system

$$\begin{aligned} \mathbf{u} =& \min_{\mathbf{u}} \quad (1/2)\|\mathbf{y} - \mathbf{A}\mathbf{u}\|_2^2 + (\rho/2)\|\mathbf{u} - \mathbf{v} + \mathbf{d}\|_2^2 \\ =& (\mathbf{A}^T\mathbf{A} + \rho\mathbf{I}_{IJ})^{-1}(\mathbf{A}^T\mathbf{y} + \rho(\mathbf{u} - \mathbf{d})) \end{aligned}, \quad (13)$$

where $\mathbf{I}_{IJ} \in \{0, 1\}^{IJ \times IJ}$ is the identity matrix. There may be scenarios where $\mathbf{y}$ is completely observed, in this case, the sensing matrix is simply $\mathbf{\Phi} = \mathbf{I}_{IJK}$, thus

$$\begin{aligned} \mathbf{A}^T\mathbf{A} + \rho\mathbf{I}_{IJ} =& \mathbf{F}_{2D}^T(\text{diag}(\hat{\mathbf{h}}_1^H \circ \hat{\mathbf{h}}_1) + \ldots \\ & + \text{diag}(\hat{\mathbf{h}}_K^H \circ \hat{\mathbf{h}}_K) + \rho\mathbf{I}_{IJ})\mathbf{F}_{2D} \end{aligned}, \quad (14)$$

can be easily inverted and precomputed. On the other hand, if the sensing matrix is modeled as (4), the solution of this system has a heavy computational cost. In order to get a closed solution and reduce the computational cost in (13), the linearized quadratic term $(1/2)\|\mathbf{y} - \mathbf{A}\mathbf{u}\|_2^2$ is used to update $\mathbf{u}$ as follow

$$\begin{aligned} (1/2)\|\mathbf{y} - \mathbf{A}\mathbf{u}\|_2^2 \approx& (1/2)\|\mathbf{y} - \mathbf{A}\mathbf{u}^{(t)}\|_2^2 + \nabla(\mathbf{u}^{(t)})^T \times \\ & (\mathbf{u} - \mathbf{u}^{(t)}) + (\eta/2)\|\mathbf{u} - \mathbf{u}^{(t)}\|_2^2 \end{aligned}, \quad (15)$$

where $\nabla(\mathbf{u}^{(t)}) = \mathbf{A}^T(\mathbf{A}\mathbf{u}^{(t)} - \mathbf{y})$ is the gradient of $(1/2)\|\mathbf{y} - \mathbf{A}\mathbf{u}\|_2^2$ at the current point $\mathbf{u}^{(t)}$ and η is a positive proximal parameter [16]. Then, the solution is given by

$$\mathbf{u} = (1/(\eta + \rho))(\eta\mathbf{u}^{(t)} - \nabla(\mathbf{u}^{(t)}) + \rho(\mathbf{v} - \mathbf{d})), \quad (16)$$

with step size $\eta > \rho\|\mathbf{A}\|_2^2$ [16].

b) Updating $\mathbf{v}$*:* To compute $\mathbf{v}$, the optimization problem to be solved is written as

$$\begin{aligned} \mathbf{v} =& \min_{\mathbf{v}} \quad \lambda\|\mathbf{v}\|_1 + (\rho/2)\|\mathbf{u} - \mathbf{v} + \mathbf{d}\|_2^2 \\ =& \max(0, \text{ soft}(\mathbf{u} + \mathbf{d}, \lambda/\rho)) \end{aligned}, \quad (17)$$

where soft$(\cdot; \tau)$ denotes the component-wise application of the soft-threshold function $y =$sign(y)max$(|y| - \tau; 0)$.

III. Experimental Results

This section presents numerical results on object detection for simulated data. The proposed scheme was implemented in Matlab and all numerical experiments were conducted on a computer with an Intel(R) Core(TM) i7-4790 CPU@3.60GHz and 32 GB RAM.

1) The UC Merced dataset (UCM): The UC Merced dataset (UCM) [17], contains 21 distinctive scene categories. Each class consists of 100 images with a size of 256×256 pixels. Each image has a pixel resolution of one foot. Figure 2 shows two examples of each category included in this dataset.

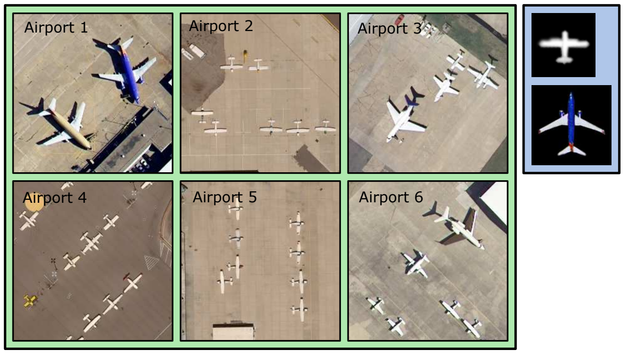

Fig. 2. Some samples from UC Merced Land data sets. The green box are the scene and blue box are the training samples.

2) Quality Metrics: Peak-to-Sidelobe Ratio (PSR) is a common metric used in correlation filter literature for detection/verification tasks [18]. PSR is calculated using the formula PSR $= (\text{ peak } - \mu_u)/\sigma_u$, where μ_u is the mean and σ_u is the standard deviation of the correlation output in a bigger region around a mask centered at the peak. Figure 3 illustrates how the PSR is estimated. Firstly, the peak is located (shown as the bright pixel in the center of Figure 3). The mean and the standard deviation of the 20×20 sidelobe region (excluding a 5×5 central mask) centered at the peak are computed.

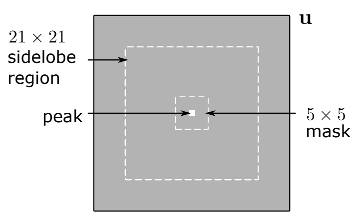

Fig. 3. This figure shows how the peak-to-sidelobe ratio (PSR) is estimated.

3) Parameters of the Algorithm: The ADMM stopping criterion is satisfied when $t = 100$ or when the dual residual is $\|\rho(\mathbf{v}^{(t)} - \mathbf{v}^{(t-1)})\|_2^2 \leq 1 \times 10^{-4}$ and the primal residual is $\|\mathbf{u}^{(t)} - \mathbf{v}^{(t)}\|_2^2 \leq 1 \times 10^{-4}$ with $\rho = 0.1$ and $\eta > \rho(\|\mathbf{A}\|_2^2)$. The regularization parameter is set to $\lambda = C\|\mathbf{A}^T\mathbf{y}\|_\infty$ with $C \in (0, 1)$ [15].

A. Parameter Analysis

This section, explores the effect of selecting the parameter λ and the Compression Ratio (CR) on the performance of the proposed method in terms of PSR obtained by applying the proposed method. The results are averaged over ten simulations, for each λ, CR and generated $\overline{\mathbf{\Phi}}$. Figure 4 depicts the behavior of PSR for Airport 1 in Figure 2 when $0.01 \leq \lambda \leq 0.1$ and compression ratio is $0.05 \leq$ CR ≤ 0.4. It can be seen that, for $\lambda \geq 0.05$, the difference between PSR is not large and the selection of parameters λ offers the potential to improve the performance of PSR obtained by the proposed method.

B. Comparison with Traditional Object Detection

In this section, the robustness of the proposed method is tested against the traditional object detection based CFs when

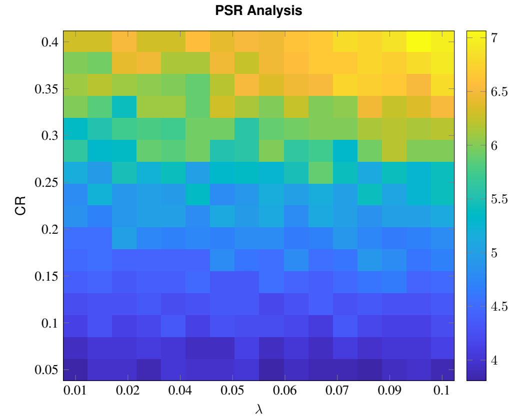

Fig. 4. Parameter sensitive analysis in terms of PSR for the proposed method on the Airport 1 data base. Different values of compression ratio (CR) are used while varying $0.01 \leq \lambda \leq 0.1$.

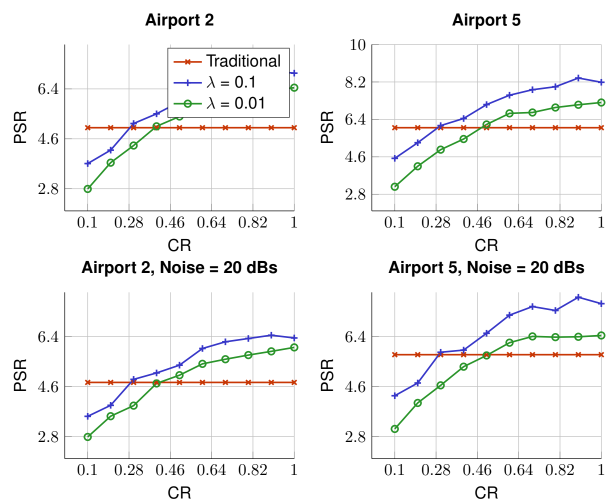

Fig. 5. Comparison of the proposed method applied and the traditional CFs, for different CR with and without additive noise.

both are corrupted by additive Gaussian noise (SNR = 20 dBs). Figure 5 compares the performance of the proposed method with respect to the traditional based CFs method. The parameter has been fixed to $\lambda = 0.01$ and different compression ratios, with or without additive noise. The results are averaged over ten simulations each time varying the matrix $\overline{\mathbf{\Phi}}$, applying the proposed method on the Airport 2 and Airport 5 images. It is possible to observe that the proposed scheme attains PSR less than 2 of difference in both noisy and noise-free cases. This test empirically validates the feasibility of the proposed scheme by showing that the object location can be directly extracted from the compressed data by solving the proposed model.

Figure 6 displays object location for both data sets. By visually comparing the reconstructed maps locations, one can observe that spatial variations of regions are preserved good enough to achieve a good detection, above CR = 0.1. Also, the proposed method captures the spatial variation and generates homogeneous regions by incorporating the regularization term

ISBN 978-90-827970-1-5 © EURASIP 2018

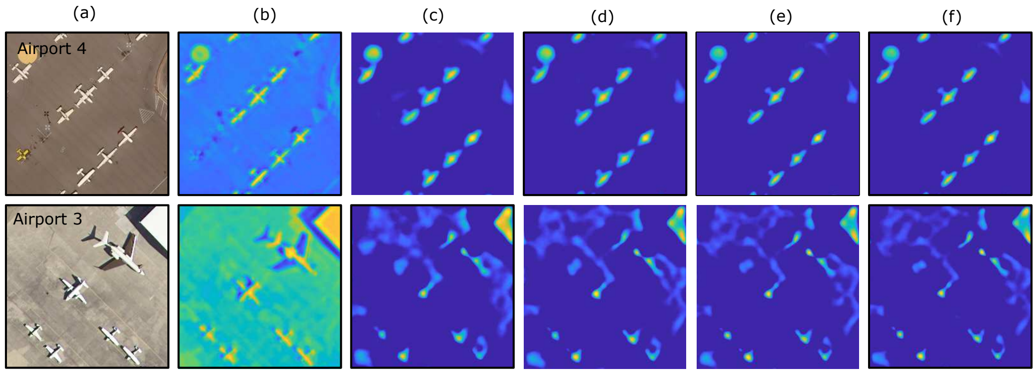

Fig. 6. Detection maps obtained by using the proposed method from the Airport 4 ((a) 1st row) and Airport 3 ((a) 2nd row) multi-band data sets. Response of filter for (b) traditional CFs, (c) proposed method for CR = 0.1, (d) CR = 0.2, (e) CR = 0.3 and (f) CR = 0.5.

λ. Notice that by using less threshold value λ, the extracted locations lose detail.

IV. Conclusion

This paper studies an object detection scheme for remote sensing data that does not require to reconstruct all data. The proposed scheme consists of three major steps: 1) Compressive multi-band image acquisition 2) Learning CFs by using the MAP synthesis model and 3) object detection by solving a sparse optimization problem with positive constraint. An efficient algorithm has been constructed for solving the sparse optimization scheme based on the alternating direction method of multipliers. In the experiments, it was shown that the proposed method considerably outperforms the traditional Correlation Filter method in terms of PSR. Numerical results clearly demonstrate that compressively acquired data of size ranging from 10% to 25% of the full size can produce satisfactory results.

References

[1] D. L. Donoho, "Compressed sensing," *IEEE Transactions on information theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[2] M. F. Duarte, M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk, "Single-pixel imaging via compressive sampling," *IEEE signal processing magazine*, vol. 25, no. 2, pp. 83–91, 2008.

[3] G. Satat, M. Tancik, and R. Raskar, "Lensless imaging with compressive ultrafast sensing," *IEEE Transactions on Computational Imaging*, vol. 3, no. 3, pp. 398–407, 2017.

[4] J. A. Fernandez, V. N. Boddeti, A. Rodriguez, and B. V. Kumar, "Zero-aliasing correlation filters for object recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 37, no. 8, pp. 1702–1715, 2015.

[5] M. A. Davenport, M. F. Duarte, M. B. Wakin, J. N. Laska, D. Takhar, K. F. Kelly, and R. G. Baraniuk, "The smashed filter for compressive classification and target recognition," in *Computational Imaging V*, vol. 6498. International Society for Optics and Photonics, 2007, p. 64980H.

[6] R. Calderbank, S. Jafarpour, and R. Schapire, "Compressed learning: Universal sparse dimensionality reduction and learning in the measurement domain," *preprint*, 2009.

[7] T. Goldstein, L. Xu, K. F. Kelly, and R. Baraniuk, "The stone transform: Multi-resolution image enhancement and compressive video," *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 5581–5593, 2015.

[8] K. Kulkarni and P. Turaga, "Reconstruction-free action inference from compressive imagers," *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 4, pp. 772–784, 2016.

[9] E. J. Candes, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on pure and applied mathematics*, vol. 59, no. 8, pp. 1207–1223, 2006.

[10] W. Yin, S. Morgan, J. Yang, and Y. Zhang, "Practical compressive sensing with toeplitz and circulant matrices," in *Visual Communications and Image Processing 2010*, vol. 7744. International Society for Optics and Photonics, 2010, p. 77440K.

[11] M. F. Duarte and Y. C. Eldar, "Structured compressed sensing: From theory to applications," *IEEE Transactions on Signal Processing*, vol. 59, no. 9, pp. 4053–4085, 2011.

[12] J. Romberg, "Compressive sensing by random convolution," *SIAM Journal on Imaging Sciences*, vol. 2, no. 4, pp. 1098–1128, 2009.

[13] H. K. Galoogahi, T. Sim, and S. Lucey, "Correlation filters with limited boundaries," in *Computer Vision and Pattern Recognition (CVPR), 2015 IEEE Conference on*. IEEE, 2015, pp. 4630–4638.

[14] C. Garcia-Cardona and B. Wohlberg, "Convolutional dictionary learning," *arXiv preprint arXiv:1709.02893*, 2017.

[15] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein *et al.*, "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.

[16] Y. Ouyang, Y. Chen, G. Lan, and E. Pasiliao Jr, "An accelerated linearized alternating direction method of multipliers," *SIAM Journal on Imaging Sciences*, vol. 8, no. 1, pp. 644–681, 2015.

[17] Y. Yang and S. Newsam, "Bag-of-visual-words and spatial extensions for land-use classification," in *Proceedings of the 18th SIGSPATIAL international conference on advances in geographic information systems*. ACM, 2010, pp. 270–279.

[18] B. V. Kumar, A. Mahalanobis, and R. D. Juday, *Correlation pattern recognition*. Cambridge University Press, 2005.