Optimum Image Fusion via Sparse Representation

Guang Yang, Xingzhong Xu and Hong Man

ECE Department, Stevens Institute of Technology
Castle Point on Hudson, Hoboken, NJ 07030
{*gyang1,xxu7,hman*}*@stevens.edu*

***Abstract*—The fusion of images captured from multi-modality sensors has been studied for many years. It is aiming at combining multiple sources together to maximize the meaningful information and reduce the redundancy. Meanwhile, sparse representation of images has been attracting more and more attentions. It has been effectively utilized on image reconstruction, image denoising, super-resolution and others. In this paper, we propose an optimum function based on sparse representation model to accomplish image fusion tasks. For any pair of input source images, we first obtain their sparse vectors respectively on a pretrained dictionary. Then we pursuit the sparse vector for the the fused image by optimizing the Euclidean distances between fused image and each input, weighted by their own gradients. Optimization penalties are discussed to induce numerical or analytical solutions. And the experimental results have shown that the proposed method can effectively combine meaningful information and outperform traditional wavelet methods.**

I. Introduction

Image fusion is known as multi-sensor or multi-modality data fusion, aiming at actively combining multiple sources together to maximize the meaningful information and reduce the redundancy. Multi-modality cameras systems are now adopted by many fields, as remote sensing, medical image processing, and military applications. Original images could be captured from multiple optical sensors at different resolutions, different capturing angles, or different spectrum bands. Most frequently used systems are, such as, visible light cameras together with Infra-red cameras, and panchromatic images together with multi-spectral data.

Image gradient is often used in metrics to evaluate the quality of a image, since human visions are more sensitive to sharpen images than blurry ones. Xydeas [1] introduced an objective metric $Q^{AB/F}$ to evaluate how much the fused image F keeping from original images gradient weighted A and B. Inspired by this work, we also adopt the gradients as the weights in our objective function.

Recently years, compressive sensing and sparse representation on image applications are very popular since the proof of ℓ_1 norm penalty equivalent to ℓ_0 norm under certain circumstance with overwhelmingly high possibility [2], [3]. In many literatures, sparse representation models are applying to signal and image processing areas, and are successfully solving many problems including in blind source separation [4], image denoising [5], image super-resolution [6] and etc. One crucial step in our proposed method, is using sparse representation to reconstruct the original images, hence to obtain their sparse coding vectors.

Inspired by the aforementioned techniques, we propose a multi-modality image fusion algorithm based on sparse representation and optimization theory. In order to fuse two images together, the new sparse vector should be very ”same” as the original sparse vectors which are weighted by the their gradients, and the similarity between fused and the original images are evaluated by their Euclidean distances.

This paper is organized as following: section II introduces some related works about image fusion and sparse representation. Section III gives the mathematic model to bridge image representations and optimization models. The contributions of our work are described in section IV, which show in detail about our proposed objective function as well as discussions on optimum penalties and solutions. Experiments and comparison in section V illustrate the advantages of this work besides the sliding window strategy in our algorithm. Conclusions are discussed in the last section.

II. Related Work

A. Image Fusion Methods

The task of image fusion is defined as ‘*the combination of two or more different images to form a new image by using a certain algorithm*’, by Pohl and Genderen reviewed in [7]. It was considered in three processing stages, and they are categorized in three levels, pixel-level, feature-level and decision-level. Among them, in [7], pixel-level fusion is most important stage to perform precise work for later processing.

As developed in decades, wavelet transform [8] has been used in many works, such as [9]. Wavelet based image fusion is to transform data into frequency domain and apply certain fusion rules to corresponding sub-bands. Unlike wavelet-based fusion algorithms, we are seeking a fusion method not only working on spatial domain directly and effectively, but also focus on more generic optimization method.

B. Sparse Representation on Image

Sparse representation is a method to model image processing problems. Suppose there is a linear system:

$$\boldsymbol{D\alpha} = \boldsymbol{x} \qquad where \quad \mathbf{D} \in \mathbb{R}^{n\times k} \quad and \quad \boldsymbol{\alpha} \in \mathbb{R}^k \tag{1}$$

Where $\boldsymbol{D} \in \mathbb{R}^{n\times k}$ is an over-complete dictionary ($n \ll k$) trained by the given samples, and each column is named as atom. $\boldsymbol{\alpha} \in \mathbb{R}^k$ is the sparse coding vector for a given test sample obtained from the dictionary $\boldsymbol{D}$, with $L \ll k$ sufficiently small nonzero elements. Inducted from the traditional compressive sensing theory, the sparse representation model in signal processing can be written as following:

$$\min_{\boldsymbol{\alpha}} \|\boldsymbol{\alpha}\|_{\ell_0} \quad subject\ to \quad \boldsymbol{x} \approx \boldsymbol{D\alpha} \tag{2}$$

Here $\|\boldsymbol{\alpha}\|_{\ell_0}$ is a pseudo norm which counts the number of nonzero elements in $\boldsymbol{\alpha}$. Actually, this standard ℓ_0 optimization form is a NP-hard problem. However if $\boldsymbol{\alpha}$ is sufficiently sparse, ℓ_0-norm can be approximated by ℓ_1-norm [3]. Therefore, the problem (2) can be proved to equivalent to solving the following ℓ_1 minimization problem:

$$\min_{\boldsymbol{\alpha}} \|\boldsymbol{x} - \boldsymbol{D}\boldsymbol{\alpha}\|_{\ell_2}^2 + \lambda\|\boldsymbol{\alpha}\|_{\ell_1} \quad where, \quad \|\boldsymbol{\alpha}\|_{\ell_1} = \sum_{i=1}^{N} \mid \alpha_i \mid \tag{3}$$

In our task, the dictionary $\boldsymbol{D}$ is collected from both two input images, and also is the dictionary which the fused image is learned from. Dictionary learning is the process of updating atoms from raw image date to near orthonormal basis. (Since the dictionary is over-complete, it cannot be orthogonal. But with RIP property, each subset of basis works very like orthonormal [2], [3].)

The dictionary learning problem, at each updating iteration t, can be stated as following:

$$\boldsymbol{D}_t = \arg\min_{\boldsymbol{D}} \frac{1}{t}\sum_{i=1}^{t}\left(\frac{1}{2}\|\boldsymbol{x}_i - \boldsymbol{D}\boldsymbol{\alpha}_i\|_{\ell_2}^2 + \lambda\|\boldsymbol{\alpha}_i\|_{\ell_1}\right) \tag{4}$$

Dictionary learning algorithms have been introduced in many literatures, for instance, MOD [10], K-svd [11], and online learning [5]. In our experiments, we have adopted the online dictionary learning algorithm in SPAMS [12].

III. Mathematic Model

For any given image, one could always be decomposed into a data vector. Consider a data vector $\boldsymbol{x} \in \mathbb{R}^n$, there exists a span of a subspace $\boldsymbol{D} \in \mathbb{R}^{n\times k}$, such that,

$$\boldsymbol{x} = \sum_{i=1}^{p} \alpha_i D_i = \boldsymbol{D}\boldsymbol{\alpha}.$$

For this subspace $\boldsymbol{D}$, it can be either a typical orthonormal space or a nonorthogonal space such as an over-complete dictionary. In this paper, we adopt the later form. Under this particular $\boldsymbol{D}$, for any given $\boldsymbol{x}$, there exists a sparse vector $\boldsymbol{\alpha} \in \mathbb{R}^k$. Sparsity means that $L \ll k$, where

$$L \triangleq \|\boldsymbol{\alpha}\|_{\ell_0} \triangleq \#\{i \text{ s.t. } \alpha[i] \neq 0\}.$$

Assume $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are the sparse vectors of the original images. Suppose the fused data is also sparse under this dictionary which is trained by the original images. In another word, for the fusion result $\boldsymbol{F}$, it will exist a sparse $\boldsymbol{\gamma}$ such that $\boldsymbol{F} = \boldsymbol{D}\boldsymbol{\gamma}$. Therefore, the image fusion problem can be solved as an optimization problem which optimally select $\boldsymbol{\gamma}$ from $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ under the dictionary $\boldsymbol{D}$.

IV. Objective function and optimum model

A. Objective function

Performance metrics are playing very important role in any image fusion system. The metrics evaluate how much information the fused image is keeping from each source image and how well it is combining many different source images together. In order to generate a good fusion result, we perform a two-side criteria to evaluate our result,

- The higher gradient should be kept in the fusion result from the source.
- The difference between the source and the fused image should be kept as small as possible.

Sharpened images will be more impressive than blurry ones. And inside the sharpness, it is the clarity of edges between color or shape changes. Moreover the gradients are the mathematical features of the edges in a image. Therefore, we are aiming to preserve those highest gradients information from the original ones. Our optimization function is the evaluation of the information preserved from the inputs in terms of the weighted local gradient for each image. Furthermore, the fused image should be as similar as their source images. Hence the Euclidean distances are used for describing the differences between the result and either one of the original images.

Based on these, we proposed our objective function:

$$\boldsymbol{y}^\star = \arg\min_{y} \|\nabla\boldsymbol{x_1}(\boldsymbol{y} - \boldsymbol{x_1})\|_{\ell_2}^2 + \|\nabla\boldsymbol{x_2}(\boldsymbol{y} - \boldsymbol{x_2})\|_{\ell_2}^2 \tag{5}$$

where $\boldsymbol{x_1}$ and $\boldsymbol{x_2}$ are the source data, and $\boldsymbol{y}$ is the fused result. The residues $\boldsymbol{y} - \boldsymbol{x_1}$ and $\boldsymbol{y} - \boldsymbol{x_2}$ characterize the difference between the original data and the result which needs to be minimized. And $\nabla\boldsymbol{x_1}, \nabla\boldsymbol{x_2} \in \mathbb{R}^{n\times n}$ here are the weights of the residues. Intuitively, in this minimization function, the one with higher gradient will lead $\boldsymbol{y}$ more similar to it.

Now consider the sparse representation of images, we define $\boldsymbol{x_1} = \boldsymbol{D}\boldsymbol{\alpha}$, $\boldsymbol{x_2} = \boldsymbol{D}\boldsymbol{\beta}$ and $\boldsymbol{y} = \boldsymbol{D}\boldsymbol{\gamma}$ in (5),

$$\boldsymbol{J} = \|\nabla\boldsymbol{x_1}\boldsymbol{D}(\boldsymbol{\gamma} - \boldsymbol{\alpha})\|_{\ell_2}^2 + \|\nabla\boldsymbol{x_2}\boldsymbol{D}(\boldsymbol{\gamma} - \boldsymbol{\beta})\|_{\ell_2}^2 \tag{6}$$

Let $\Phi_1 = \nabla\boldsymbol{x_1}\boldsymbol{D}$, and $\Phi_2 = \nabla\boldsymbol{x_2}\boldsymbol{D}$, then (6),:

$$\boldsymbol{J} = \|\Phi_1\boldsymbol{\gamma} - \Phi_1\boldsymbol{\alpha}\|_{\ell_2}^2 + \|\Phi_2\boldsymbol{\gamma} - \Phi_2\boldsymbol{\beta}\|_{\ell_2}^2 \tag{7}$$

This objective function can be viewed as a quadratic form, and to be more analytically, it can be rewritten as:

$$\begin{aligned} \boldsymbol{J} &= \boldsymbol{\gamma}^T(\Phi_1^T\Phi_1 + \Phi_2^T\Phi_2)\boldsymbol{\gamma} - 2((\Phi_1\boldsymbol{\alpha})^T\Phi_1 + (\Phi_2\boldsymbol{\beta})^T\Phi_2)\boldsymbol{\gamma} \\ &\quad + (\Phi_1\boldsymbol{\alpha})^T\Phi_1\boldsymbol{\alpha} + (\Phi_2\boldsymbol{\beta})^T\Phi_2\boldsymbol{\beta} \\ &= \boldsymbol{\gamma}^T\boldsymbol{P}\boldsymbol{\gamma} - 2\boldsymbol{Q}^T\boldsymbol{\gamma} + \boldsymbol{C} \end{aligned} \tag{8}$$

Where we define $\boldsymbol{P}$, $\boldsymbol{Q}$, and $\boldsymbol{C}$ as:

$$\boldsymbol{P} = \Phi_1^T\Phi_1 + \Phi_2^T\Phi_2 \tag{9}$$

$$\boldsymbol{Q} = \Phi_1\boldsymbol{\alpha}\Phi_1^T + \Phi_2\boldsymbol{\beta}\Phi_2^T \tag{10}$$

$$\boldsymbol{C} = (\Phi_1\boldsymbol{\alpha})^T\Phi_1\boldsymbol{\alpha} + (\Phi_2\boldsymbol{\beta})^T\Phi_2\boldsymbol{\beta} \tag{11}$$

Now the optimum selector $\boldsymbol{\gamma}^\star$ is,

$$\begin{aligned} \boldsymbol{\gamma}^\star &= \arg\min_{\boldsymbol{\gamma}} \frac{1}{2}\boldsymbol{\gamma}^T\boldsymbol{P}\boldsymbol{\gamma} - \boldsymbol{Q}^T\boldsymbol{\gamma} + \frac{1}{2}\boldsymbol{C} \\ \text{subject to} &: \|\boldsymbol{\gamma}\|_{\ell_p} < L \end{aligned} \tag{12}$$

B. Sparse Regularity

In (12), a norm ℓ_p is applied to the inequality constraint. And here we will discuss the following three kinds of regularizer, when $p = 0, 1,$ or 2.

1) ℓ_0 norm: This represents the original concept of the sparsity, which is illustrating the restriction of non-zero element in given vector. Basis pursuit is usually the method of solving it, however the studies show that the ℓ_0 regularity is NP-Hard which means it is less practical to solve this problem.

2) ℓ_1 norm: This is the heuristic approximation of the sparsity, as we have discussed in section II-B. Although the ℓ_1 norm is not differentiable at every point, as a convex problem, it is practical to solve. In addition, the feasible set should not be empty, meanwhile L is required to be as small as possible. Thus, we use a bi-criterion function to take the place. Equation (12), can be written as following:

$$\boldsymbol{\gamma}^\star = \arg\min_{\boldsymbol{\gamma}} \frac{1}{2}\boldsymbol{\gamma}^T \boldsymbol{P}\boldsymbol{\gamma} - \boldsymbol{Q}^T\boldsymbol{\gamma} + \frac{1}{2}\boldsymbol{C} + \lambda\|\boldsymbol{\gamma}\|_{\ell_1} \tag{13}$$

where $\lambda > 0$ represents the trade-off between the sparsity of $\boldsymbol{\gamma}$ and fitting the data. This is also known as LASSO [13] problem, therefore it can be solved using some off-the-shelf convex optimization algorithms.

3) ℓ_2 norm: The Euclidean norm constrain can not hold the sparsity well, however it have the good convexity and differentiable property. Following previous ℓ_1 discussion, the bi-criterion can be applied to trade-off the two-fold minimization. Consider the ℓ_2 norm,

$$\begin{aligned} \boldsymbol{\gamma}^\star &= \arg\min_{\boldsymbol{\gamma}} \frac{1}{2}\boldsymbol{\gamma}^T \boldsymbol{P}\boldsymbol{\gamma} - \boldsymbol{Q}^T\boldsymbol{\gamma} + \frac{1}{2}\boldsymbol{C} + \lambda\|\boldsymbol{\gamma}\|_{\ell_2}^2 && (14)\\ &= \arg\min_{\boldsymbol{\gamma}} \frac{1}{2}\boldsymbol{\gamma}^T (\boldsymbol{P}+\lambda\boldsymbol{I})\boldsymbol{\gamma} - \boldsymbol{Q}^T\boldsymbol{\gamma} + \frac{1}{2}\boldsymbol{C} && (15)\end{aligned}$$

This can be formulated to Tikhonov Regularization problem, therefore have analytic solution,

$$\begin{aligned} \frac{d}{d\boldsymbol{\gamma}}(\frac{1}{2}\boldsymbol{\gamma}^T (\boldsymbol{P}+\lambda\boldsymbol{I})\boldsymbol{\gamma} - \boldsymbol{Q}^T\boldsymbol{\gamma} + \frac{1}{2}\boldsymbol{C}) &= 0 && (16)\\ \boldsymbol{\gamma}^\star &= (\boldsymbol{P}+\lambda\boldsymbol{I})^{-1}\boldsymbol{Q} \end{aligned}$$

Each of these three regularizer has the different performance to hold the sparsity. The ℓ_0 norm have the greatest sparsity holding property, but need to solve a NP-Hard problem. The ℓ_2 norm have the lowest holding, however have the analytic solution. And for the eclectic ℓ_1 approach, it not only have heuristic ability to hold the sparsity, but also can be solved by numerical method.

V. Experiments and Comparisons

A. Image Gradient

As the data here is image data, therefore the 2-D gradients are including magnitudes and angles. Here Sobel operator is applied to take the edges from the image. And in this algorithm, the magnitudes will be take consider in as the weight, where we get from the horizontal and vertical gradients.

B. Sliding Window Strategy

Our experiments are using thermal images and visible images from TNO dataset [14]. In order to collect an overcomplete dictionary, we are using a sliding window to scanning through the whole image. Each small patch is used for an atom in the initial dictionary. The window is in the size of 8×8 pixels, with half window length sliding step.

Sliding window strategy is designed to obtain more meaningful information (specifically, image gradient is as the weight,) from both modalities. A pair of original images should have different performance at same place, but one can not always win at everywhere. Therefore evaluations on small patches are much more objective than to perform on the entire images. Biases will be highly eliminated by averaging the overlapping parts while reconstructing the results.

C. Fusion Results

In our experiments, we have tested ℓ_1 and ℓ_2 regularizer optimization for our problem, due to ℓ_0 is a less practical NP-hard problem. To solve the ℓ_1 regularizer which is a LASSO problem, we use well developed numerical optimization tool CVX [15] to find the optimum fusion strategy. For the ℓ_2 optimization, it is already been deduced to a Tikhonov Regularization problem. Hence we could have analytical close form solution from (16) for any of the given data.

As shown in Fig. 1 and Fig. 2, in infra-red images, Fig. 1a and 2a, a person can be seen clearly in a bright spot, but not in the visible ones. On the other hand, the visible images, Fig. 1b and 2b show some details of the fences, leaves and the building, which are very blurry in IR image. The fusion results Fig. 1c, 1d, 2c, 2d, not only keep the important person information , but also add those details in the scenes.

D. Fusion Evaluation and Comparisons

We use the fusion metric [1] to evaluate and compared our method to the others. In (17), the $Q^{AF}_{n,m}$ and $Q^{BF}_{n,m}$ denote the edge preservation from the original data A and B to the fusion result at the specific (n, m) pixel. And the $g^A_{i,j}$ denote the gradient of the data A at the (i, j) pixel.

$$Q^{AB/F} = \frac{\sum_{n=1}^{N}\sum_{m=1}^{M} Q^{AF}_{n,m} g^A_{n,m} + Q^{BF}_{n,m} g^B_{n,m}}{\sum_{i=1}^{N}\sum_{j=1}^{M}(g^A_{i,j} + g^B_{i,j})} \tag{17}$$

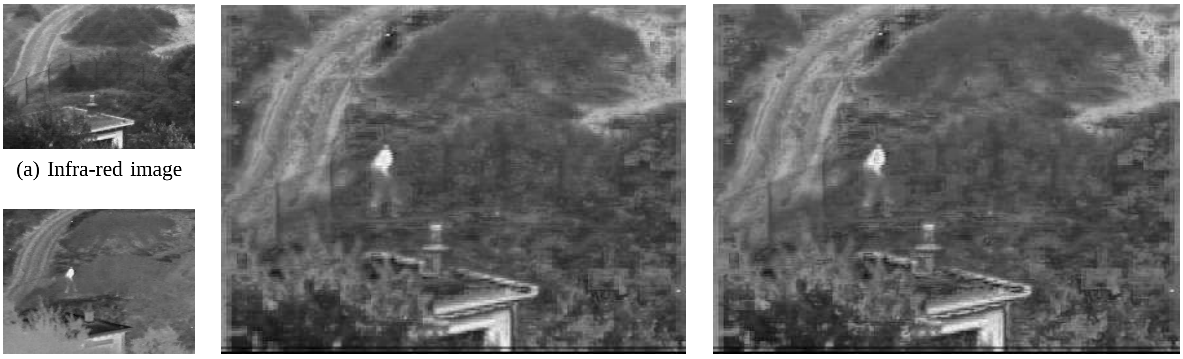
In TABLE I, we have tested our results as well as a typical wavelet fusion methods [8] through (17). The higher $Q^{AB/F}$ value is, the better preservation the fused image obtained from the source images. In this comparison, our results are clearly better than traditional fusion method. Moreover, the result of ℓ_1 norm optimization outperform the ℓ_2 one. However, ℓ_2 penalty has an analytical solution instead of the numerical approximation which make this problem computationally efficient.

TABLE I
Fusion Comparisons

$Q^{AB/F}$	wavelet	ℓ_1 norm	ℓ_2 norm
Experiment 1	0.3469	0.4296	0.4296
Experiment 2	0.3391	0.4218	0.4218

VI. Conclusions

In this paper, we have proposed an optimum image fusion method. Based on the sparse coding technique, the data has been translated into a sparse representations and corresponding



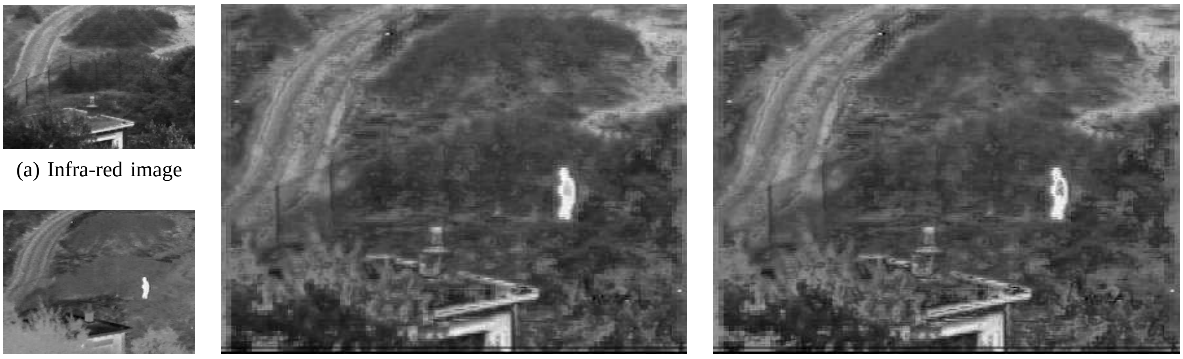
(a) Infra-red image

(b) Visible image

(c) ℓ_1 norm

(d) ℓ_2 norm

Fig. 1. Experiment 1



(a) Infra-red image

(b) Visible image

(c) ℓ_1 norm

(d) ℓ_2 norm

Fig. 2. Experiment 2

basis in the dictionary. An optimization model is designed in a two-fold objective function, fitting original data and preserving gradient. Furthermore, in this paper, we discussed three types of optimization regularities, ℓ_0, ℓ_1, and ℓ_2 norm, in this model. By introducing a trade-off factor λ to a bi-criterion form, it is not only in consistency with types of penalties, but also keeping a form of quadratic convex function. Then, with the help of numerical techniques, we conducted experiments with outperform results by solving the objective function with ℓ_1 norm, and deduced the analytical solution for the ℓ_2 norm. Finally, the experiments and comparisons illustrated improvements and advantages of the performance by our proposed optimization model.

REFERENCES

[1] C. Xydeas and V. Petrovic, “Objective image fusion performance measure,” *Electronics Letters*, vol. 36, no. 4, pp. 308 –309, feb. 2000.

[2] E. Candes, “Compressive sampling,” in *Proceedings of the International Congress of Mathematicians*, 2006.

[3] E. J. Candes and T. Tao, “Near-optimal signal recovery from random projections: Universal encoding strategies?” *Information Theory, IEEE Transactions on*, vol. 52, no. 12, pp. 5406 –5425, 2006.

[4] J. Bobin, Y. Moudden, J. Fadili, and J.-L. Starck, “Morphological diversity and sparsity in blind source separation,” in *Proceedings of the 7th international conference on Independent component analysis and signal separation*, ser. ICA’07, 2007, pp. 349–356.

[5] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, “Online learning for matrix factorization and sparse coding.” *Journal of Machine Learning Research*, pp. 19–60, 2010.

[6] J. Yang, J. Wright, T. Huang, and Y. Ma, “Image super-resolution as sparse representation of raw image patches,” in *IEEE Conference on CVPR*, 2008, pp. 1–8.

[7] C. Pohl and J. L. Van Genderen, “Review article multisensor image fusion in remote sensing: concepts, methods and applications,” *International Journal of Remote Sensing*, vol. 19, no. 5, pp. 823–854, March 1998.

[8] S. Mallat, “A theory for multiresolution signal decomposition: the wavelet representation,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 11, pp. 674 –693, jul 1989.

[9] S. Mitra, B. Manjunath, and H. Li, “Multisensor image fusion using the wavelet transform,” in *ICIP94*, 1994, pp. I: 51–55.

[10] K. Engan, S. O. Aase, and J. H. Husøy, “Multi-frame compression: theory and design,” *Signal Process.*, vol. 80, pp. 2121–2140, October 2000.

[11] M. Aharon, M. Elad, and A. Bruckstein, “k -svd: An algorithm for designing overcomplete dictionaries for sparse representation,” *Signal Processing, IEEE Transactions on*, vol. 54, no. 11, pp. 4311 –4322, 2006.

[12] J. Mairal, “SPAMS: Sparse modeling software, version 2.0,” http://www.di.ens.fr/willow/SPAMS/, Oct. 2010.

[13] R. Tibshirani, “Regression shrinkage and selection via the lasso,” *Journal of the Royal Statistical Society, Series B*, vol. 58, pp. 267–288, 1994.

[14] A. Toet, J. Ijspeert, A. Waxman, and M. Aguilar, “Fusion of visible and thermal imagery improves situational awareness,” *Displays*, vol. 18, no. 2, pp. 85–95, December 1997.

[15] M. Grant and S. Boyd, “CVX: Matlab software for disciplined convex programming, version 1.21,” http://cvxr.com/cvx, Oct. 2010.