Infrared Physics & Technology 57 (2013) 56–67

Contents lists available at SciVerse ScienceDirect

Infrared Physics & Technology

journal homepage: www.elsevier.com/locate/infrared

# Research on fusion method for infrared and visible images via compressive sensing

Meng Ding [a,*], Li Wei [b], Bangfeng Wang [a]

[a] College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China
[b] Jin Cheng College, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China

## HIGHLIGHTS

- We use compressive sensing to study fusion method of infrared and visible images.
- This paper firstly proposes the fusion rule of maximum absolute of entry of sparse vector.
- The method using OMP provides better results in the condition of the same parameter setting, dictionary and fusion rule.
- The method ***IRdictionary_maxabsolute_OMP*** takes almost all the largest objective evaluations.

## ARTICLE INFO

*Article history:*
Received 31 October 2012
Available online 21 December 2012

*Keywords:*
Infrared image
Visible image
Image fusion
Compressive sensing
K-SVD
Sparse vector approximation

## ABSTRACT

In order to obtain a more exact, reliable and better description than a single source image, we need to fuse source images taken from different sensors to a synthetic image. This paper employs infrared and visible images and uses the theory of compressive sensing to study image fusion method. The fusion method based on compressive sensing theory contains three parts: overcomplete dictionary, the algorithm of sparse vector approximation and fusion rule. This paper selects three trained overcomplete dictionaries by K-means Singular Value Decomposition (K-SVD) including the dictionary only using patches from the infrared images, the dictionary only using patches from the visible images and the dictionary using the combined patches, two sparse vector approximations containing orthogonal matching pursuit and polytope faces pursuit algorithms, and two fusion rules covering maximum $\ell^1$-norm and maximum absolute of entry of sparse vector which is firstly proposed in this paper to study twelve fusion approaches. The experimental results show that the method using orthogonal matching pursuit can provide better fusion results in the condition of the same parameter setting and the same dictionary and fusion rule, and the method using the dictionary only using patches from the infrared images, the fusion rule of maximum absolute of entry of sparse vector and orthogonal matching pursuit takes almost all the largest objective evaluations and the best fusion quality.



## 1. Introduction

In recent years, many images for one object or scene can be acquired by multiple sensors with the technology development of image sensors. In accordance with the natural properties of the sensors and the approach the images are obtained, all of these source images taken from the sensor directly contain unique information which is usually complementary. Taking infrared and visible images studied in this paper for example, infrared images have lower contrast and definition compared with visible images, but visible images cannot capture targets effectively in low visibility conditions [1,2]. The fusion of infrared and visible images can obtain a more exact, reliable and better description than a single source image. Therefore, we can think that image fusion is the foundation of the analysis of multisensor images [3]. In order to extend or enhance information about the scene, it is necessary to develop the technology of image fusion by combining the images captured by different sensors. Image fusion is the process of detecting salient features in the source images and fusing these details to a synthetic image. For the past few years, image fusion has been applied widely in diverse fields, such as target detection, intelligent surveillance, and nondestructive inspection [4–6].

All of image fusion methods developed or applied in the past two decades can be divided into three levels: pixel-level, feature-level, and decision-level in accordance with the stage where the information acquired by different image sensors is fused to a synthetic [7,8]. Pixel-level fusion method which combines the independent source images into a single image, reserves most of the information and is studied widely. Feature-level methods which typically use features of source images (such as edges or regions) to fuse them are usually robust to noise and misregistration. Since

* Corresponding author.
*E-mail address:* nuaa_dm@hotmail.com (M. Ding).

1350-4495/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.infrared.2012.12.014

decision-level methods combine image descriptions directly, the application field of them is limited greatly. The method proposed in this paper belongs to pixel-level fusion. The pixel-level approaches mainly contain two categories [9,10]: spatial domain-based methods and transformed domain-based methods. The simplest fusion method in spatial domain just takes the pixel-by-pixel average of the source images. Spatial domain-based methods often lead to undesirable side effects, such as lower contrast, blockness in the fused image [11]. In the past decades, one of the most successful image fusion methods is by using multiscale transform [12]. As an outstanding transformed domain-based method, the image fusion method based on multiscale transform has three basic steps [13]: (1) decompose source images into multiscale representations with different resolutions and orientations; (2) decompose coefficients denoting multiscale representations and linked with the salient features of the original images are integrated according to fusion rules; (3) reconstruct the fused image by the inverse transform of the integrated multiscale coefficients.

The earliest multiscale transform for image fusion is pyramid decomposition, for example, Laplacian pyramid [14], morphological pyramid [15], gradient pyramid [16]. Pyramid method firstly decomposes each source image into a series of images with different sizes (calls pyramid) in different resolutions, and then extracts the value in the pyramid with the highest saliency at each position in the decomposed images, finally reconstructs the fused image using the inverse transform of the composite images. Another multiscale transform for image fusion is wavelet transform-based methods which use a similar scheme to the pyramid decomposition, such as discrete wavelet transform [17,18], stationary wavelet transform [19], and dual-tree complex wavelet transform [20,21]. The principal shortcoming of these methods is that most of the multiscale transforms are not shift invariant, which is brought by the underlying down-sampling process. Recently, multiscale geometry analysis has been developed and used for image fusion to improve fusion results. Typical methods include ridgelet transform [22], curvelet transform [23]. However, although different wavelets can represent different image details, wavelets and related multiscale transform cannot extract all of the underlying information of the source images effectively. The reason is that the dictionary constructed by different basis functions is limited. Decomposed coefficients of the fused image obtained by a limited dictionary in the transformed domain may cause all the pixel values to change in the spatial domain. Consequently, in some cases multiscale transform-based fusion methods may produce undesirable artifacts.

Obviously, in order to make the fused image more accurate, it is necessary to explore a novel method to extract the underlying information of the source images more efficiently and completely. This paper proposes an image fusion method which is based on the recently developed theory of compressive sensing (CS). CS theory has been successfully applied in different fields of image process or computer vision [24–27], such as image denoising, image compression, feature extraction, and target classification. The core of CS is the sparse representation which describes natural signals including images by a sparse linear combination of columns of an overcomplete dictionary. Different from a limited dictionary within multiscale transformations, CS uses an overcomplete dictionary in which every column is also called a signal atom. Overcompleteness is the most prominent characteristic of the dictionary used in CS theory. Overcompleteness denotes that the number of signal atoms in the overcomplete dictionary is more than signal dimensions greatly and guarantees more meaningful and complete representation of source signals than the traditional multiscale transformations [28]. CS theory reveals the coefficients corresponding to the natural sign are sparse. Thus, Li and Yang employ the sparse coefficient vectors to give a framework of CS-based image fusion [29]. According to this framework, our method for image fusion proposed in this paper contains following steps: Firstly, the overcomplete dictionary is created and trained by K-means Singular Value Decomposition (K-SVD) [30]. Secondly, the source images are divided into patches by sliding window which is adopted to achieve better performance in capturing local salient features of source images and improve the image fusion quality. Thirdly, the patches are decomposed by the overcomplete dictionary into their corresponding sparse coefficients. Fourthly, tow fusion rules are employed to combine the coefficients of the source images. Finally, the fused image is reconstructed using the combined coefficients.

The rest of the paper is organized as follows: Section 2 presents the basic theory of CS and K-SVD for the overcomplete dictionary creation. In Section 3, the fusion scheme based on CS and the fusion rules based on sparse coefficients are discussed. Experiment and conclusion are demonstrated in Sections 4 and 5 respectively.

## 2. Basic theory of compressive sensing

In order to finish image fusion using CS-based method, a brief description of CS theory is necessary. CS theory has four essential factors. First, as mentioned above, overcompleteness is one of major characteristics of CS. Additional important conception is sparsity. Third, how to compute sparse coefficients denoting linear combinations of overcomplete dictionary is a problem. Lastly, creating an overcomplete dictionary is the key step of using CS theory in practices.

### 2.1. Overcomplete representation

Suppose a given signal vector $\mathbf{y} \in \mathcal{R}^{\boldsymbol{n}}$, and a collection of vectors $\boldsymbol{\varphi}_i \in \mathcal{R}^{\boldsymbol{n}}$, $i = 1, \ldots, m$, where $m > n$. Such collections are usually dictionaries and each vector $\boldsymbol{\varphi_i}$ is an atom. Given signal $y = \sum a_i \boldsymbol{\varphi_i}$ ($i = 1, \ldots, m$) can be represented as a linear combination of atoms in the dictionary. Different from traditional multiscale transform basis representation, such linear combination of dictionary offers a wider range of generating atoms [31]. Thus, this dictionary is overcomplete and called overcomplete dictionary [32]. Overcomplete representation based on such dictionary can allow more flexibility in signal representation and more effectiveness in signal process [33].

Considering the atoms as the columns of overcomplete dictionary $\boldsymbol{\Phi}$, overcomplete dictionary $\boldsymbol{\Phi} = [\boldsymbol{\varphi_1}, \boldsymbol{\varphi_2}, \ldots, \boldsymbol{\varphi_m}]$, so that the matrix $\boldsymbol{\Phi} \in \mathcal{R}^{n \times m}$. Linear algebra tells us that a representation of the given signal $\mathbf{y}$ can be described as a coefficient vector $\mathbf{a} = [a_1, a_2, \ldots, a_m]^{\mathbf{T}}$ satisfying $\mathbf{y} = \boldsymbol{\Phi}\mathbf{a}$. Since $m > n$, the problem of overcomplete representation is undetermined. That means there is no unique solution of the coefficients vector $\mathbf{a}$. In order to obtain unique solution, considering the impact of sparsity constraint on this situation, CS theory can in certain circumstances generate a sparse coefficient vector as the linear combination of overcomplete dictionary. This coefficient vector is called sparse representation (shown in Fig. 1).

### 2.2. Sparse representation and sparse vector approximation

Generally speaking, the purpose of CS theory is to solve the problem of finding the sparsest representation possible in an overcomplete dictionary. As a measure of sparsity of a vector $\mathbf{a}$, $\ell^0$-*norm* $\|\mathbf{a}\|_0$ denotes the number of non-zero entries in $\mathbf{a}$. The sparsest representation is the solution to the optimization problem [34]:

$$\min_{\mathrm{a}} \|\mathrm{a}\|_0 \qquad \text{s.t.} \qquad \mathrm{y} = \boldsymbol{\Phi}\mathrm{a} \tag{1}$$

In most practical situations, the above formula can be modified to include a noise allowance [33]: