Information Fusion 25 (2015) 72–84

Contents lists available at ScienceDirect

Information Fusion

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

Multi-focus image fusion using dictionary-based sparse representation

Mansour Nejati [a], Shadrokh Samavi [a,b,*], Shahram Shirani [b]

[a] Department of Electrical and Computer Engineering, Isfahan University of Technology, Iran
[b] Department of Electrical and Computer Engineering, McMaster University, Hamilton, Canada

ARTICLE INFO

Article history:
Received 31 January 2014
Received in revised form 7 September 2014
Accepted 19 October 2014
Available online 1 November 2014

Keywords:
Multi-focus image fusion
Dictionary learning
K-SVD
Sparse representation
Guided image filtering

ABSTRACT

Multi-focus image fusion has emerged as a major topic in image processing to generate all-focus images with increased depth-of-field from multi-focus photographs. Different approaches have been used in spatial or transform domain for this purpose. But most of them are subject to one or more of image fusion quality degradations such as blocking artifacts, ringing effects, artificial edges, halo artifacts, contrast decrease, sharpness reduction, and misalignment of decision map with object boundaries. In this paper we present a novel multi-focus image fusion method in spatial domain that utilizes a dictionary which is learned from local patches of source images. Sparse representation of relative sharpness measure over this trained dictionary are pooled together to get the corresponding pooled features. Correlation of the pooled features with sparse representations of input images produces a pixel level score for decision map of fusion. Final regularized decision map is obtained using Markov Random Field (MRF) optimization. We also gathered a new color multi-focus image dataset which has more variety than traditional multi-focus image sets. Experimental results demonstrate that our proposed method outperforms existing state-of-the-art methods, in terms of visual and quantitative evaluations.

1. Introduction

Depth of field in optical lenses of conventional cameras is limited. Thus, only the objects at a particular distance from the camera are in focus and captured sharply whereas objects at other distances in front of or behind the focus plane are defocused and blurred. However, for accurately interpreting and analyzing images, it is desired to obtain images with every object in focus [1]. Multi-focus image fusion is an effective technique to solve this problem by combining two or more images of the same scene taken with different focus settings into a single all-in-focus image with extended depth of field, which is very useful for human or machine perception. The multi-focus image fusion has been applied in various applications such as microscopic imaging, remote sensing, and computer vision [1].

During the past years, many multi-focus image fusion algorithms have been developed [3–16]. According to the fusion domain, these algorithms could be categorized into two main groups: transform domain and spatial domain fusion [2]. In the first group, transform coefficients are fused and the fused image is reconstructed from these composite coefficients. The transform domain fusion methods that are based on multi-scale transforms are the most commonly used methods in this group [3]. Many kinds of multi-scale transforms have been proposed and adopted for image fusion such as pyramid decomposition [4], discrete wavelet transform (DWT) [5,6], dual-tree complex wavelet transform (DTCWT) [7], and discrete cosine harmonic wavelet transform (DCHWT) [8]. Recently developed multiscale geometry analysis tools with higher directional sensitivity than wavelets, such as shearlet transform [9], curvelet transform (CVT) [10], and nonsubsampled contourlet transform (NSCT) [11] are employed too. Also, some novel signal decomposition methods like robust principal component analysis (RPCA) [1] and sparse representation (SR) [12–14], are also applied to image fusion. The transform domain fusion methods have three common basic steps. First, the source images are decomposed to get the transform coefficients. The transform coefficients are then integrated according to a certain fusion rule. The fused image is finally constructed by applying the inverse transform on the fused coefficients [12].

Unlike the transform domain fusion methods, in spatial domain methods, fusion rules are directly applied to image pixels or image regions. In general, spatial domain methods can be classified into two groups of pixel based [3,15,16], and region based methods [17,18]. The main principle in these methods is selecting the pixels or regions with more clarity according to some image clarity measure, namely focus measure, to construct the fused image. Energy of Laplacian and spatial frequency are two typical focus

* Corresponding author at: Department of Electrical and Computer Engineering, Isfahan University of Technology, Iran.

http://dx.doi.org/10.1016/j.inffus.2014.10.004
1566-2535/

measures used to make a decision about the clarity of pixels or regions. The main drawbacks of these spatial domain fusion methods are misalignment of decision map with boundary of focused objects and wrong decision in sub-regions of the focused or defocused regions which produce undesirable artifacts in the final fused image. To mitigate these artifacts, some spatial techniques use the weighted average of pixel values for fusing the source images, instead of using binary decision [3,16]. Depending on weight construction method, these methods lead to halo artifacts near some edges, contrast decrease, and/or reduction of sharpness.

In this paper we introduce a new multi-focus image fusion method in spatial domain which produces spatially smooth and edge aligned decision map to accurately merge the in-focus regions of multi-focus source images into a single fused image. Our method relies on dictionary learning and sparse representation of focus features. Sparse representation has proven to be an extremely powerful tool for analyzing a large class of signals [19]. In addition to successful application of sparse representation in many classical signal processing problems, such as compression and denoising [20,21], sparse-representation-based techniques are increasingly attracting attention in computer vision area due to its state-of-the-art performance in many applications, such as image classification [22,23], face recognition [24] action recognition [19], and object tracking [25]. Basic observation in these applications is that despite the images (or their features) are naturally very high dimensional, the images in the same class usually lie on a low-dimensional subspace [26]. Therefore, given a dictionary of representative samples for the distribution, it is expected that there exists a sparse representation with respect to such a dictionary for a typical sample.

State-of-the-art performance of sparse representation in many computer vision tasks motivated us to use this powerful tool in a new multi-focus image fusion framework. In this framework, a dictionary is learned for sparse representation of focus features extracted from small patches of source images. Max pooling is then applied to the sparse representations from training samples to get the pooled features. Correlation of training pooled features with the sparse representations of input source images produces a pixel level score map which is employed for an initial decision about the clarity of each pixel. Final regularized decision map is obtained using Markov Random Field (MRF) optimization. Our approach is different from previous sparse-representation-based fusion methods [12–14] in two aspects. First, fused images in those mentioned references are constructed by using the inverse transform of the fused sparse coefficients while our method has the advantage of working in the spatial domain. Second, we use the sparse representation, which is obtained from a learned dictionary, for feature extraction and classification of pixels as focused or unfocused ones. To validate the effectiveness of our method, extensive experiments are conducted using two datasets under three objective quality metrics. Experimental results demonstrate that our proposed method outperforms existing state-of-the-art methods, in terms of visual and quantitative evaluations.

The rest of this paper is organized as follows. In Section 2, the signal sparse representation theory is a briefly reviewed. Section 3 describes the details of the proposed multi-focus image fusion based on dictionary learning and sparse representation of focus feature. Experimental results, comparison to state-of-the-arts and objective evaluations are demonstrated in Section 4. Finally, Section 5 concludes the paper.

2. Sparse representation theory

Sparse signal representations have recently drawn much interest in vision, signal and image processing [26], [27]. Sparse signal representation has proven to be an extremely powerful tool for analyzing a large class of signals. This is mainly due to the fact that signals and images of interest can be sparse or compressible in some dictionary of bases [3]. In sparse representation modeling of an input signal $y \in \mathbb{R}^n$, it is represented as a linear combination of a few atoms of an over-complete dictionary $\Phi \in \mathbb{R}^{n\times K}$ $(K > n)$ as

$$y = \Phi x \tag{1}$$

where the vector $x \in \mathbb{R}^K$ contains the representation coefficients of the signal y and the dictionary matrix Φ contains K prototype signals referred as atoms for columns, $\{\phi_j\}_{j=1}^K$. The linear system in (1) with a full-rank over-complete dictionary Φ, becomes an underdetermined system of linear equations having an infinite number of solutions, hence constraints on the solution must be set. Finding the sparsest solution with the fewest number of nonzero coefficients involves solving the optimization problem

$$\min_x \|x\|_0 \quad \text{subject to} \quad y = \Phi x \tag{2}$$

in exact representation, or

$$\min_x \|x\|_0 \quad \text{subject to} \quad \|y - \Phi x\|_2 \leqslant \epsilon \tag{3}$$

in approximate representation [20], where $||\cdot||_0$ is the ℓ_0 semi-norm that counts the number of nonzero entries in a vector. These are in general NP-hard problems and thus, approximation techniques such as pursuit algorithms are used to get an approximated solution [20]. Orthogonal matching pursuit (OMP) is a greedy pursuit algorithm which is widely used because of its simplicity and efficiency [21].

Construction of a proper dictionary is an important issue in sparse representation. It has been observed that learning a dictionary directly from training signals leads to better representation and hence provides superior performance when compared to predefined dictionaries (such as Fourier or wavelet) in many image and vision applications [28]. Particularly, in computer vision, we often have to learn a task-specific dictionary from given sample images [26].

Methods for learning a dictionary for sparse coding from the training data have been proposed recently [20,23,29]. Let $Y \in \mathbb{R}^{n\times N}$ be a set of N input signals of dimension n, *i.e.* $Y = \{y_i\}_{i=1}^N, y_i \in \mathbb{R}^n$. Learning a dictionary $\Phi \in \mathbb{R}^{n\times K}$ $(K > n)$ with K atoms for sparse representation of Y is formally written as the following optimization problem:

$$\widehat{\Phi}, \hat{X} = \underset{\Phi, X}{\operatorname{argmin}} \|Y - \Phi X\|_F^2 \quad \text{subject to} \quad \forall i, \|x_i\|_0 \leqslant T \tag{4}$$

where $||\cdot||_F$ is the Frobenius norm, $X = \{x_i\}_{i=1}^N, x_i \in \mathbb{R}^K$ are the sparse representation of input signals Y, and T is a sparsity constraint of sparse representations to be contained no more than T nonzero coefficients.

3. Proposed multi-focus image fusion

Motivated by the powerful ability of sparse coding in classification, we propose a new framework for multi-focus image fusion base on learning dictionary and its corresponding sparse representation using focus measure of local patches of source images. In this framework, every pixel from each source image is classified as an either in-focus pixel to be used in the fused image or not. According to the classification, a decision map for fusing of input images is obtained that label of each pixel in this map indicates intensity value of which input images must be used for that location in fused image.

The proposed algorithm consists of two phases: training and testing. Fig. 1 gives an overview of the proposed method for the case of two source images. In the training phase, an overcomplete