Digital Signal Processing 58 (2016) 50–63

Contents lists available at ScienceDirect

Digital Signal Processing

www.elsevier.com/locate/dsp

Multi-focus image fusion based on sparse decomposition and background detection

Zhang Baohua [a,b], Lu Xiaoqi [a,b,*], Pei Haiquan [b], Liu Yanxian [b], Zhou Wentao [b], Jiao Doudou [b]

[a] School of Computer Engineering and Science, Shanghai University, Shanghai 200444, China
[b] School of Information Engineering, Inner Mongolia University of Science and Technology, Baotou 014010, China

ARTICLE INFO

Article history:
Available online 29 July 2016

Keywords:
Multi-focus image
RPCA
Guided filter
Difference image

ABSTRACT

The goal of image fusion is to accurately and comprehensively describe complementary information of multiple source images in a new scene. Traditional fusion methods are easy to produce side-effects which cause artifacts and blurred edges. To solve these problems, a novel fusion algorithm based on robust principal component analysis (RPCA) and guided filter is proposed. The guided filter can preserve the edges effectively, which is often used to enhance the images without distort the details. Considering edges and flat area are treated differently by the guided filter, in this paper, sparse component of the source image is filtered by the guided filter to generate the enhanced image which contains the preserved edges and the enhanced background. And then the focused regions of the source images are detected by spatial frequency map of the difference images between the enhanced image and the corresponding source image. Finally, morphological algorithm is used to obtain precise fusion decision map. Experimental results show that the proposed method improves the fusion performance obviously which outperforms the current fusion methods.

1. Introduction

An optical imaging system can only focus on objects with same depth of field in one scene because of the limited depth-of-focus of optical lenses, which is unable to clearly represent all the objects in complex scenes. Complementarity between source images is usually employed in multi-focus image fusion algorithms to produce a fused image with all objects clearly shown in the scene [1]. Fusion results can describe the scene of the image more accurately and comprehensively, which effectively enhances the utilization of multi-focus images [2,3].

Existing fusion techniques can be generally classified into spatial and transform domain methods. The spatial domain methods exhibit the advantage of directly fusing sharp regions from source images [3], which highly depends on the selection of clarity measures, such as the energy of image gradient, standard deviation or spatial frequency, etc. [4,5]. Region-based fusion methods can effectively extract the focused region from the source image because structural information cannot be represented by a single pixel [6].

Li and Yang [7] proposed a region segmentation based fusion method, which accurately located the focused region based on spatial frequency. However, the fused image suffered from serious blocking effects, which significantly compromised the performance of the method. Furthermore, the fusion scheme was sensitive to noise interference and artifacts because the selection criteria cannot reflect local features. In Ref. [8], the source image was optimally divided into sub-blocks by measuring morphological gradient energy of a sub-block, and then completely extracting the sharp regions. But the blocking effect was inevitable. Sun employed Markov Random fields to determine the focused region [9], which avoided reducing the contrast of the fused image. However, the location of the boundary was unreliable due to the unstable performance of the segmentation method. Taking into account the absence of direct correspondence between the contour and the gray value of the pixel, Tian applied gradient intensity and phase matching to determine the focusing characteristics of a sub-block [10], which accurately located the focused region. However similarly to Sun's method [9], the universality of this technique was unsatisfactory because the sliding window cannot be adaptively adjusted. In Ref. [11], each image was divided into a clear region, a blurred region and an intermediate transition region by adaptively computing the depth information, and thus overcoming the blocking effects. But the fusion result was unstable and the edge transition was unnatural.

The identification mechanism of human eye lies in characterizing the objects in a scene as specific geometric features. In re-

* Corresponding author.
E-mail address: zbh_wj2004@imust.cn (B. Zhang).

http://dx.doi.org/10.1016/j.dsp.2016.07.010
1051-2004/© 2016 Elsevier Inc. All rights reserved.

cent years, transform-based methods have been widely used in image fusion because of the requirement of multi-channel information classification [12]. However, none of the existing methods can characterize all the geometric features of an image. For example, although wavelet [13], Contourlet [14], Curvelet [15] and Surfacelet [16] can optimally represent isotropic structure information, rectangles, curves and surfaces respectively, only limited directional information can be captured, which cannot effectively extract the entire contour accurately. Side effects, such as pseudo Gibbs phenomenon and false contour seriously compromise the quality of the fused image. Although non-subsampled operation can avoid the pseudo Gibbs phenomenon [17], the fusion method modifies the original pixel values of the images; thus, the sharp regions of a source image cannot be directly merged into the fused image, which cannot take full advantage of the complementary information between the source images. Compared with the sharp regions of source images, the blurred corresponding part of a fused image cannot precisely represent the target [2].

Sparse decomposition can represent the salient information of an image by building the relationship between features and sparse coefficients. Hence, sparse decomposition-based methods, such as sparse representation based, NMF-based and RPCA-based methods are proposed to fuse the multi-focus images, which are proven to be more comprehensive and effective to extract the structure information of the source image [18,19]. The learned dictionary preserves the hierarchical structure and intrinsic characteristics of images [20]. However, the over-complete dictionary constructed by the randomly selected sub-block is unreliable because of its uncertainty, which compromises the performance of the fusion method. Moreover, larger dimensions of a dictionary lead to longer calculating times.

Unlike sparse representation-based fusion methods, which should learn the redundant dictionaries from images, NMF-based methods construct feature space to capture the local features of source images and can more easily extract salient features [18]. Furthermore, the edges can be better preserved by sparse feature matrix decomposition. Although a fused image obtained by NMF based method achieves better fusion results than that produced by traditional multi-scale decomposition fusion methods [21,22], sections of the details are lost in the fused image due to local and global features are equally treated. Moreover, the boundary location strongly depends on the performance of the segmentation method, which increases the instability of the fusion method. In [23], RPCA has been utilized to detect the infrared target adaptively, which proves that a target region can be easily separated from the sparse matrix. Wan employed RPCA to extract the salient information of multi-focus images from a sparse matrix and construct a fused image by integrating the sharp regions [24]. A fusion decision map was generated by calculating the standard deviation of the sub-block; hence, the blocking effects still seriously compromised the performance of the fusion method. Zhang introduced a pulse coupled neural network to construct the fusion decision map, which avoided segmenting a sparse matrix into blocks [25]. Although only a few artificial side effects were found in the fused image, the computational complexity was unsatisfactory.

To address the aforementioned problems, a novel multi-focus image fusion method is proposed in this paper. There are two main contributions of this paper for constructing a robust fusion method based on sparse decomposition and the detection for the background. The first lies in the introduction of a novel background enhancement method, which incorporates the extraction of the sparse component and the enhancement of the background. The sparse component obtained by using RPCA contained structure information of source images, which can separate the focused regions from the background coarsely. Considering boundary and flat area are treated differently, the guided filter is used to filter the sparse component to generate the enhanced image which contains the preserved edges and the enhanced background. The second main contribution of this work is the accurate detection of the background. The differenced image between the enhanced image and the source image provides robust contour information to achieve precise positioning of the background which significantly differs from traditional detection methods. Thus, the proposed method can adequately resolve artificial side effects, such as blocking effects, ringing effects, false contours and the glitches. Experimental results show that compared with other existing fusion methods, this method can fuse the focused region exactly and is insensitive to the registration of the source images.

The remainder of the paper is organized as follows. In Section 2, detection method for background is presented. Section 3 describes the proposed algorithm in details. Experimental results and performance analysis on several groups of multi-focus images are presented and discussed in Section 4. Finally, Section 5 concludes the paper.

2. Related work

2.1. Robust principal component analysis

Principal component analysis (PCA) cannot capture the contour information completely due to it is insensitive to the blurred edges. To effectively extract structure information, Wright et al. [26] proposed robust principal component analysis (RPCA) to represent image as linear superposition of low-rank structure and sparse matrix by optimizing nuclear norm.

Similar to other sparse representation method, RPCA employs nuclear norm l_1 as approximate sparsity constraint. Assume input data matrix $D \in R^{m\times n}$ (m and n are matrix dimensions) that can be decomposed as:

$$\begin{aligned} &\min_{A,E}\left(rank(A)+\lambda\|E\|_0\right) \\ &s.t. \qquad A+E=D \end{aligned} \tag{1}$$

where A is low rank matrix, E is sparse matrix, as shown in Fig. 1, $\|\cdot\|_0$ is 0-norm matrix, which represents number of non-zero elements. The optimization of eq. (1) is a NP-hard problem, so eq. (2) is commonly used as the objective function.

$$\begin{aligned} &\min_{A,E}\|A\|_*+\lambda\|E\|_1 \\ &s.t. \qquad A+E=D \end{aligned} \tag{2}$$

where $\lambda > 0$, $\|\cdot\|_*$ is nuclear norm, that is the sum of singular value, $\|\cdot\|_1$ represents 1-norm matrix, that is the sum of the absolute value of all elements of the matrix.

Sparse decomposition of source image is shown in Fig. 1, we can see that only focused region of the source image is able to observe in grade map of sparse component. Compared with typical salient analysis algorithms [27–29], RPCA is more suitable for representing feature vector of the image [30], which is useful to extract the object in complex background.

2.2. Guided filter

Although RPCA has superior ability to represent structural information, background is easily misclassified as focused region due to contour of the focused region are easily distorted by noise, so various noise filters are employed to avoid the interference. Although bilateral filter can preserve the edges, the feature vector of Gaussian kernel function cannot accurately capture details in complex background. As a result, stitching traces are obvious in the transition region between the object and the background, and the details of the background are lost seriously. Moreover, running time of the bilateral filter is proportional to r^2 (r is radius of the filter window), which is unable to satisfy real time needs.