ARTICLE IN PRESS

JID: INS [m3Gsc;September 11, 2017;7:46]

Information Sciences 000 (2017) 1–14

Contents lists available at ScienceDirect

Information Sciences

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

A novel multi-modality image fusion method based on image decomposition and sparse representation

Zhiqin Zhu[a,b], Hongpeng Yin[a,b,*], Yi Chai[b], Yanxia Li[a,b], Guanqiu Qi[b,c]

[a] Key Laboratory of Dependable Service Computing in Cyber Physical Society of Ministry of Education, Chongqing University Chongqing, 400030, China

[b] College of Automation, Chongqing University, Chongqing 400044, China

[c] School of Computing, Informatics, and Decision Systems Engineering, Arizona State University, Tempe, Arizona 85287, USA

A R T I C L E I N F O

Article history:
Received 14 December 2016
Revised 1 September 2017
Accepted 3 September 2017
Available online xxx

Keywords:
Sparse representation
Dictionary construction
Multi-modality image fusion
Cartoon-texture decomposition

A B S T R A C T

Multi-modality image fusion is an effective technique to fuse the complementary information from multi-modality images into an integrated image. The additional information can not only enhance visibility to human eyes, but also mutually complement the limitations of each image. To preserve the structure information and perform the detailed information of source images, a novel image fusion scheme based on image cartoon-texture decomposition and sparse representation is proposed. In proposed image fusion method, source multi-modality images are decomposed into cartoon and texture components. For cartoon components a proper spatial-based method is presented for morphological structure preservation. An energy based fusion rule is used to preserve structure information of each source image. For texture components, a sparse-representation based method is proposed. A dictionary with strong representation ability is trained for the proposed sparse-representation based fusion method. Finally, according to the texture enhancement fusion rule, the fused cartoon and texture components are integrated. The experimentation results have clearly shown that the proposed method outperforms the state-of-art methods, in terms of visual and quantitative evaluations.

1. Introduction

Multi-modality image fusion combines the complementary information from multi-modality sensors to enhance visibility of human eyes or mutually complement the limitations of each image. Diverse modalities of images, such as infrared-visible images, multi-focus images, and medical images are utilized in visual surveillance systems for visibility enhancement and better situation awareness [12,17]. A large amount of research efforts were made to improve image fusion performance in the last two decades. Spatial-domain and transform-domain fusion are two typical branches of image fusion.

Spatial-domain based methods directly choose clear pixels, blocks, or regions of source images to compose a fused image without transformation [14]. Based on the measurement of image clarity, those pixels or regions with higher clarity are selected to construct the fused image by spatial-domain based methods. To decrease the constraints, averaging, max pixel schemes perform on single pixel to generate fused image. However, the contrast and edge intensity of fused image may

* Corresponding author at: Key Laboratory of Dependable Service Computing in Cyber Physical Society of Ministry of Education, Chongqing University Chongqing, 400030, China. Tel.: +8618623081817.

E-mail addresses: zhiqinzu@outlook.com (Z. Zhu), yinhongpeng@gmail.com (H. Yin), cqchaiyi@cqu.edu.cn (Y. Chai), liyanxia106@gmail.com (Y. Li), guanqiuq@asu.edu (G. Qi).

http://dx.doi.org/10.1016/j.ins.2017.09.010

decrease. In general, spatial-domain methods may lead to blurring edges, contrast decrease, and reduction of sharpness [25]. Some methods, as block-based and region-based algorithms [13], are proposed to improve the quality of fused image. Although lock-based algorithms improve the detailed expression of fused image, the sharpness of fused image may still be undesirable. Block-based algorithms may cause block effect, when they are applied to spatial-domain based methods.

Transform-domain based methods use a transform tool to decompose source images into coefficients and transform bases first. Then the coefficients are fused by diverse fusion rules in different applications. Finally, the fused image can be obtained by inversely transforming fused coefficients and transform bases. Multi-scale transform(MST) is one of the most popular fusion techniques in multi-modal image fusion. Starting with Discrete Wavelet Transform (DWT) [28], a variety of transforms including Dual-Tree Complex Wavelet Transform (DT-CWT) [11], Curvelet Transform (CVT) [27], Shearlet Transform [37] and Non-Subsampled Contourlet Transform (NSCT) [15] have been used in multi-modal image fusion. Although transform coefficients can reasonably represent important features of an image, each transform has its own merits and limitations corresponding to the context of input images [42]. Thus, selecting an optimal transform basis is not an obvious and trivial problem as it relies on scene contexts and applications [17].

In recent years, sparse representation(SR) has been successfully implemented to image classification [2,19], image super-resolution [44],image recognition [21], image feature extraction [20], image deblurring [30], image object recognition [18,20] and multi-modality information fusion [47]. As a transform-based method, SR was first applied to image fusion by Li and Yang [41]. They used DCT transform to build the dictionary for SR, and an SR-based fusion framework was proposed. A medical image fusion and de-noising method by group sparse-representation was introduced in [17]. However, this method is not tested on color medical images. Yang and Liu [42] proposed several kinds of mathematical models, that were utilized to construct hybrid dictionaries for image fusion. The hybrid dictionaries can well reflect several specific structures, but are still lack of the adaptability to represent different types of images. In that case, learning-based adaptive dictionaries were implemented for SR-based image fusion [36]. KSVD-based method is the most widely used adaptive dictionary construction method for SR-based image fusion [26,45]. Multi-focus image fusion methods based on KSVD were proposed by Yin [45] and Nejati [26], and showed good state-of-art performances. Yin [47] also proposed a multi-modality medical image fusion method by KSVD, which can enhance the performance of image details. Nonparametric Bayesian adaptive dictionary learning was proposed in [40] for remote-sensing image fusion.

Although SR gets great performance in image fusion, it still has two limitations in multi-modality image fusion. The first limitation is the most advanced Max-L1 sparse coefficients fusion rule may cause spatial inconsistency in the integrated multi-modality images [22]. The second limitation is only one generally trained dictionary cannot accurately reflect the complex structures of input image[17].

For the first limitation, decomposing source images into both high and low frequency components is the most common solution. Gaussian filter as a common used filter is applied to image decomposition of SR-based image fusion method [10]. They used SR and Max-L1 rule for image high-frequency components fusion, which can remain more fusion details. Multi-scale transformation(MST) filter [22] is also used for SR-based image fusion method. However, MST-based filter [22] shows limitations on decomposing specific kinds of images. Gaussian filter [10] also has limited performance on decomposing the detailed and structure information of input images.

For the second limitation, in [10], Kim first clustered training samples into a few structural groups by k-means method. Then Kim trained a specific sub-dictionary for each group. In this way, each sub-dictionary can fit a particular structure and the whole dictionary has strong representation ability. Similarly, Wang et al. separately constructed spectral and spatial dictionary for the fusion of multi-spectral and panchromatic images [39]. However, Kim's method needs to set the cluster numbers before clustering and Wang's method can only be used in remote-sensing image fusion.

In this paper, a novel image fusion method based on image decomposition and sparse representation is proposed, and is specific to the mentioned two limitations. For the first limitation, a compact and informative dictionary learning method for texture component fusion is proposed. In the dictionary learning processes, input images are clustered into a few pixel groups for sub-dictionary learning. Sub-dictionary of each pixel group can fit particular structures of different image features. To cluster image pixels with different features, the local regression weight of steering kernel (SKR) [33] is implemented. As a sophisticated and effective feature, SKR feature can reflect local image structures effectively even with the presence of noises [10]. In this case, SKR feature is implemented as the pixel feature for clustering. In image pixel clustering, local density peaks (LDP) clustering method is implemented [1]. Local density peaks clustering method can group image pixels without setting the cluster number in advance. A few compact sub-dictionaries can be trained by extracting the underlying information of each pixel cluster.

For the second limitation, a cartoon-texture based image fusion framework is proposed. According to the properties of cartoon and texture contents, proper fusion methods are constructed to fuse cartoon and texture components respectively. For texture components, SR-based method is conducted. For cartoon components, energy-based fusion rule is implemented to preserve the geometric structure information of all the source images. A gradient information based method is proposed to integrate cartoon and texture components.

The main contributions can be summarized as follows:

1. We propose a novel dictionary training method to enhance the sparse-representation ability. In our proposed, image pixels are clustered and sub-dictionaries are trained for each cluster. Sub-dictionaries are combined to construct the final dictionary.