Information Fusion 18 (2014) 107–118

Contents lists available at SciVerse ScienceDirect

Information Fusion

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

Image fusion with morphological component analysis

Yong Jiang [a,b,*], Minghui Wang [a]

[a] College of Computer Science, Sichuan University, Chengdu 610064, China
[b] College of Computer Science and Technology, Southwest University of Science and Technology, Mianyang 621010, China

ARTICLE INFO

Article history:
Received 11 February 2012
Received in revised form 27 May 2013
Accepted 5 June 2013
Available online 19 June 2013

Keywords:
Multiscale transform
Morphological component analysis
Sparse representation
Image fusion
Multi-component fusion

ABSTRACT

Image fusion can produce a single image that describes the scene better than the individual source image. One of the keys to image fusion algorithm is how to effectively and completely represent the source images. Morphological component analysis (MCA) believes that an image contains structures with different spatial morphologies and can be accordingly modeled as a superposition of cartoon and texture components, and that the sparse representations of these components can be obtained by some specific decomposition algorithms which exploit the structured dictionary. Compared with the traditional multiscale decomposition, which has been successfully applied to pixel-level image fusion, MCA employs the morphological diversity of an image and provides more complete representation for an image. Taking advantage of this property, we propose a multi-component fusion method for multi-source images in this paper. In our method, source images are separated into cartoon and texture components, and essential fusion takes place on the representation coefficients of these two components. Our fusion scheme is verified on three kinds of images and compared with six single-component fusion methods. According to the visual perceptions and objective evaluations on the fused results, our method can produce better fused images in our experiments, compared with other single-component fusion methods.

1. Introduction

The abilities of the imaging devices to capture the information are different from one to another; even the capture emphasis of the same image device also varies with the imaging environments. Therefore, multiple images of one scene may be acquired by different image sensors, under different optic conditions or at different times. Many applications, such as clinical medicine, military surveillance, molecular biology and remote sensing, need a single composite image that can provide more comprehensive descriptions of the scene, compared with the individual source image; this goal can be achieved by image fusion, which fuses the complementary or salient information of the multiple source images [1]. Image fusion can be performed at pixel-level, feature-level and decision-level [2,3]. Pixel-level fusion, pixel by pixel or region by region, selects or combines the information of the source images, according to some fusion criteria, to construct the fused image. Feature-level fusion performs its fusion by using some extracted features. Decision-level fusion forms the fused image by considering the image descriptions such as relational graphs [4,5]. Currently, most of the fusion algorithms are pixel-level. Pixel-level fusion can be performed in either spatial domain or transformed domain. In the spatial domain, pixels or regions are directly selected, according to some salience measures, and combined in either a linear or non-linear way to form the fused image. The most successful spatial-domain-fusion algorithms include intensity-hue saturation transform method [6,7], weighted average method, principal component analysis method [8], independent component analysis method [9] and Brovey transform method [10]. In the transformed domain, a certain frequency or time–frequency transform is used to fuse images. Of all transformed-domain-fusion methods, multiscale transform method is the most frequently-used one. Classic multiscale transforms include pyramid decomposition such as Laplacian pyramid (LP) [11], morphological pyramid (MP) [12] and gradient pyramid (GP) [13], wavelet transform methods such as discrete wavelet transform (DWT) [14–16], dual-tree complex wavelet transform (DTCWT) [17], and stationary wavelet transform (SWT) [18], and multiscale geometry analysis such as curvelet transform (CVT) [19,20], ridgelet transform [21], and nonsubsampled contourlet transform (NSCT) [22–25]. All these multiscale transform based fusion methods need to perform the following steps [16,26,27].

1. Perform the forward transform on the source images to obtain their multiscale representations (transform coefficients with different scales and directions).
2. Combine these multiscale representations to obtain the fused multiscale coefficients, according to the fusion rules designed for certain purposes.

* Corresponding author at: College of Computer Science, Sichuan University, Chengdu 610064, China. Tel.: +86 13628081535.
E-mail address: randomjy@126.com (Y. Jiang).

1566-2535/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.inffus.2013.06.001

3. Perform the inverse-transform over the combined multiscale coefficients to obtain the fused result.

The completeness and effectiveness of the transformed representations for the underlying information of the source images are crucial to the fusion quality [3,13]. All methods mentioned above view the whole image as a single component. In this way, there may be a limit with respect to representing the inherent information of images completely. If an image is decomposed into multiple components, the presentations of the image content would be more effective and more complete. Morphological component analysis (MCA) [28] can decompose an image into two components. Next, we discuss the advantages of MCA, from the perspective of the completeness and effectiveness.

(1) Completeness: In the multiscale transform, the transform bases are constructed to reveal the salient features of an image; once the transform bases are given, the transform coefficients of an image are accordingly determined and represent salient features of this image [26]. However, each multiscale transform has its own merits and demerits because it uses a limited dictionary (explanation in Section 2.1) to perform its operation. For example, wavelets, ridgelet bases and discrete cosine transform (DCT) bases are appropriate to reveal image details, lines and periodic textures, respectively, but none of them can simultaneously capture two or more features [3]. Therefore, there is no single transform which is optimal to completely represent all features because the content of an image is often complex and contains structures with different spatial morphologies. Morphological diversity assumes that an image can be described as the sum of multiple components. In [29], one real image is decomposed into two components: a cartoon image and a texture map. The cartoon image describes the piecewise smooth changes in the illumination or the salient parts, as well as edges. The texture map delivers the texture information in the regions enclosed by edges. MCA [28] can also separate an image into cartoon and texture components and thus provides more complete representation for the content of an image. Some applications of MCA are image inpainting [30], image super-resolution [31] and image denoising [32].

(2) Effectiveness: The sparse representation uses an overcomplete dictionary, whose number of columns is greater than that of rows, and models an image of 1-D representation as the linear combination of columns (atoms) [33]; if the columns are selected appropriately, the combination coefficients would be sparse. Regardless of whether such an overcomplete dictionary is implicit (this kind of dictionary consists of several transform bases) or explicit (this kind of dictionary is comprised of some non-parametric trained dictionaries), it should be rich. A dictionary, via its richness, may reveal the salient features of an image more effectively than the traditional transform base. The overcompleteness enables sparser representation to reflect the common image features. Some classic applications of the sparse representation are face recognition [34], image super-resolution [35] and image denoising [36,37]. MCA also exploits the sparse representation to decompose an image into the cartoon and texture parts.

Exploiting the property of morphological diversity of images and the advantages of MCA, we propose a novel multi-component fusion method in this paper. The core idea of the proposed method is that fusing multiple components (cartoon and texture) could outperform fusing a single component. We structure the following content of this paper as follows. Next section briefly introduces the sparse representation technique and reviews how to decompose an image with MCA. Section 3 presents the image fusion scheme based on MCA, including the fusion algorithm flow and fusion rule. Section 4 discusses the detailed experimental settings and analyses experimental results. Section 5 draws conclusions and puts forward our future work.

2. Decomposing an image with MCA

2.1. Sparse representation

In the sparse representation framework, a dictionary $\Phi = [\varphi_1, \ldots, \varphi_T]$ is viewed as an $N \times T$ matrix. When $T > N$ or even $T \gg N$, the dictionary is overcomplete and is constructed by merging several dictionaries. An image $x \in R^N$ (an image with N pixels can be expressed as a lexicographically ordered 1-D vector) is modeled as the linear combination of $M (M < T)$ elementary atoms of the dictionary, according to Eq. (1).

$$x = \Phi\alpha = \sum_{i \in I_M} \alpha[i]\varphi_i \tag{1}$$

where $\alpha[i]$ is the representation coefficients of x; I_M is the subset of $\{1, \cdots, T\}$ and $Card(I_M) = M$; φ_i represents the atoms of Φ. Obviously, from Eq. (1), x has many candidate representations. The sparsest one is the objective of the representation. Thus, the sparse representation problem need to solve the following minimization problem:

$$\min_{\alpha \in R^M} ||\alpha||_0 \quad s.t. \quad x = \Phi\alpha. \tag{2}$$

The complexity of the problem formulated in Eq. (2) exponentially grows with the number of columns of the dictionary because the problem is nonconvex. For reducing the complexity, the nonconvex l_0 sparsity measure is substituted by the l_1-norm [38]. Thus, Eq. (2) evolves into a tractable convex optimization problem Eq. (3), which can be solved with basis pursuit (BP) [39] algorithm.

$$\min_{\alpha \in R^M} \|\alpha\|_1 \quad s.t. \quad x = \Phi\alpha. \tag{3}$$

2.2. Morphological component analysis

In [28], the authors expatiate the morphological component analysis (MCA) that can acquire the sparse multi-component representations of an image. MCA seeks for these presentations of an image x based on two assumptions. One assumes that x can be modeled as the sum of K components $\left(x = \sum_{k=1}^{K} x_k\right)$, and that each component x_k looks morphologically different. The other assumes that the representation coefficients of x_k are sparse in a given dictionary Φ_k, but nonsparse in other dictionaries $\Phi_{k' \neq k}$ (called mutual incoherence). Generally, $K = 2$. x_1 contains the texture only, while x_2 contains the cartoon only. x_1 and x_2 are sparsely represented in Φ_1 and Φ_2, respectively, but nonsparsely represented in Φ_2 and Φ_1, respectively. For the clarity of reading, the subscript '1' is rewritten as 'T' abbreviation for 'texture', and '2' as 'C' abbreviation for 'cartoon'. For an arbitrary image x, to seek for the sparse representations of both texture and cartoon components over the combined dictionary containing both Φ_T and Φ_C, one needs to solve

$$\{\alpha_T^{opt}, \alpha_C^{opt}\} = \arg \min_{\{\alpha_T, \alpha_C\}} \|\alpha_T\|_1 + \|\alpha_C\|_1 \quad s.t.\ x = \Phi_T\alpha_T + \Phi_C\alpha_C \tag{4}$$

where α_C and α_T are the representation coefficients for cartoon and texture components, respectively. In [40], considering that the first assumption may not be well satisfied, for instance, image containing additive noise, Eq. (4) evolves into Eq. (5) by relaxing the constraint in Eq. (4).

$$\left\{\alpha_T^{opt}, \alpha_C^{opt}\right\} = \arg \min_{\{\alpha_T, \alpha_C\}} \|\alpha_T\|_1 + \|\alpha_C\|_1 + \lambda\|x - \Phi_T\alpha_T - \Phi_C\alpha_C\|_2^2 \tag{5}$$

where λ is the scalar factor. Eq. (5) can be efficiently solved by using the iterative block-coordinate relaxation (BCR) [41] method. Fig. 1