Neurocomputing 171 (2016) 138–148

Contents lists available at ScienceDirect

Neurocomputing

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

Fractional differential and variational method for image fusion and super-resolution

Huafeng Li [a,b,*], Zhengtao Yu [a], Cunli Mao [a]

[a] Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Yunnan, Kunming 650500, PR China
[b] College of Automation, Chongqing University, Chongqing 400044, PR China

ARTICLE INFO

Article history:
Received 8 September 2013
Received in revised form
9 June 2015
Accepted 17 June 2015
Communicated by Hantao Liu
Available online 30 June 2015

Keywords:
Image fusion
Fractional differential
Fractional total variation
Image super-resolution
Edge enhancement

ABSTRACT

This paper introduces a novel fractional differential and variational model that includes the terms of fusion and super-resolution, edge enhancement and noise suppression. In image fusion and super-resolution term, the structure tensor is employed to describe the geometry of all the input images. According to the fact that the fused image and the source inputs should have the same or similar structure tensor, the energy functional of the image fusion and super-resolution is established combining with the down-sampling operator. For edge enhancement, the bidirectional diffusion term is incorporated into the image fusion and super-resolution model to enhance the visualization of the fused image. In the noise suppression term, a new variational model is developed based on the fractional differential and fractional total variation. Thanks to the above three terms, the proposed model can realize the image fusion, super-resolution, and the edge information enhancement simultaneously. To search for the optimal solution, a gradient descent iteration scheme derived from the Euler–Lagrange equation of the proposed model is employed. The numerical results indicate that the proposed method is feasible and effective.

1. Introduction

One image acquired by a single sensor usually does not contain enough information to solve a problem at hand. Several images concerning the same scene acquired by different modality sensors or different environmental conditions usually contain complementary and redundant information about the observed scene. To combine the complementary information from these images, image fusion is proposed. The purpose of this technique is commonly described as the task of combining substantial information from multiple images of the same scene. Then a single image that will be more suitable for human and machine perception or further image processing tasks is created. This technique has been successfully applied in many fields, such as digital imaging, clinical medicine, remote sensing, military surveillance, machine vision, and so on.

In recent years, image fusion has received increasing attention in the research community, and many fusion methods have been proposed to merge the multi-source images [1–6]. The most popular image fusion method is to use the multiscale transforms. The basic idea of this approach is to perform a multiscale transform (MST) on each source image, and then integrate all of these decomposition coefficients according to fusion principles to produce a composite representation. With the composite coefficients, the fused image is finally reconstructed by taking the inverse MST. In such fusion methods, the commonly used MSTs include the discrete wavelet transform (DWT) [1–4], lifting stationary wavelet transform (LSWT) [7], contourlet transform [5,8,9], nonsubsampled contourlet transform (NSCT) [10–12], and curvelet transform [13–15].

Recently, new image fusion methods, including those based on the sparse representations theory [16,17], estimation theory [18], neighbor distance [19], and variational theory [20–24], have been proposed. Note that most of these methods are proposed based on the assumption that the source images have high resolution. But in practice, many imaging systems, such as infrared imagers and CCD cameras cannot acquire an image with high resolution due to the limitations of the inherent array density of the sensors. For image fusion, if the resolutions of all the source images are very low, the fused results generated by the traditional methods would also have a low resolution. To address this problem, Yin et al. proposed a new method based on the Bicubic interpolation and sparse representation [6]. However, this method suffers from the need of three large sized dictionaries. In addition to this method, based on two steps, namely, image fusion and image super-resolution or image super-resolution and image fusion, one can obtain a fused result with high-resolution. However, this sequential approach is

* Corresponding author at: Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Yunnan, Kunming 650500, PR China.
E-mail address: hfchina99@163.com (H. Li).

http://dx.doi.org/10.1016/j.neucom.2015.06.035
0925-2312/

time consuming and may propagate the errors and the artifacts in each stage. Therefore, it is necessary to exploit a new method which can solve the image fusion and super-resolution problem simultaneously.

Variational method is very popular in image super-resolution and fusion in recent years. In aspect of image fusion, Gemma Piella presented a variational model to perform the fusion of an arbitrary number of images, and employed the perceptual contrast enhancement to enhance the visualization [21]. In [20], a variational approach for image fusion and denoising was proposed by Kumar and Dass. For image super-resolution, Babacan proposed a novel super-resolution method by employing a variational Bayesian analysis [25]. Based on a constrained variational model, a new time dependent convolutional model for super-resolution method is proposed in [26].

In this paper, we combine the variational super-resolution and the image fusion scheme to produce a novel variational model for image fusion and super-resolution in pixel level. To better preserve the edges of the super-resolution fused image, a novel bidirectional filter method is employed to enhance the sharpness of edges and suppress contour jaggies. In order to overcome the ill-posedness of the super-resolution fusion model and reduce the sensitivity to the noise, a novel fractional order total variation (TV) regularization term is exploited based on the fractional differential because the fractional differential can furthest preserve the low-frequency contour feature in the smooth areas, and nonlinearly keep high-frequency marginal feature in the areas whose gray-level changes greatly, and also enhance texture details in the areas where gray-level does not change evidently [27].

The rest of this paper is organized as follows. Section 2 describes the geometry of a multi-valued image. Section 3.1 presents the variational method employed for image fusion and super-resolution. Bidirectional filter method is introduced in Section 3.2. In Section 3.3, the noise suppression term based on the fractional order total variation is constructed in order to reduce the sensitivity to the noise. Experimental results and discussion, including the comparisons of the fused results and performance analysis, are presented in Section 4. Finally, some conclusions are given in Section 5.

2. Geometric structures of multi-channel image

2.1. Gradient operator of multi-channel image

As stated in [21], the local geometric structures are important parts of an image, because they usually correspond to perceptually important features. Therefore, the fused image should contain the basic geometry of the source images. Usually, the geometric structures can be described by the first-order or second-order derivatives, or scale-space decompositions [21]. Thus, the gradient operator associated with the first-order derivative can be used to describe the geometric information of image. In a similar way to [21,22], the structure tensor, which is related to the first derivative, is employed to describe the geometry of the source images in this paper.

Let Ω denote the image domain (usually a rectangle), and $u_1(x,y), u_2(x,y), \ldots, u_S(x,y)$ denote the S source images of the same scene. Each image which has been acquired using different instrument modalities or capture techniques exhibits diverse characteristics, such as salient features, degradation, and texture properties. In this study, the input images are assumed to have been registered to each other, otherwise, fusion would be meaningless. Now, we consider the source images as a multi-channel image, and express it as $u(x,y) = (u_1(x,y), u_2(x,y), \ldots, u_S(x,y))$. The square of the variation $u(x,y)$ at position (x,y) in direction θ can be presented [28] by

$$\begin{aligned}(du)^2 &= \|u(x+\varepsilon\cos\theta, y+\varepsilon\sin\theta) - u(x,y)\|_2^2 \\ &\approx \sum_{i=1}^{S}\left(\frac{\partial u_i}{\partial x}\varepsilon\cos\theta + \frac{\partial u_i}{\partial y}\varepsilon\sin\theta\right)^2 \\ &= \varepsilon^2\sum_{i=1}^{S}\left(\frac{\partial u_i}{\partial x}\cos\theta + \frac{\partial u_i}{\partial y}\sin\theta\right)^2. \end{aligned} \tag{1}$$

The change rate $U(\theta)$ of the multi-channel image $u(x,y)$ in direction θ at position (x,y) can be descried as

$$\begin{aligned}U(\theta) &= \sum_{i=1}^{S}\left(\frac{\partial u_i}{\partial x}\cos\theta + \frac{\partial u_i}{\partial y}\sin\theta\right)^2 \\ &\triangleq E\cos^2\theta + 2F\cos\theta\sin\theta + G\sin^2\theta, \end{aligned} \tag{2}$$

where

$$E = \sum_{i=1}^{S}\left|\frac{\partial u_i}{\partial x}\right|^2, \quad F = \sum_{i=1}^{S}\frac{\partial u_i}{\partial x}\frac{\partial u_i}{\partial y}, \quad G = \sum_{i=1}^{S}\left|\frac{\partial u_i}{\partial y}\right|^2.$$

Let $d(\theta)$ be the unit vector $(\cos\theta, \sin\theta)$ in any direction θ. For a gray level image $(S=1)$, ∇u and $\nabla u^{\perp}$ are defined as $\nabla u = (u_x, u_y)^T$, $\nabla u^{\perp} = (-u_y, u_x)^T$. The function

$$U(\theta) = E\cos^2\theta + 2F\cos\theta\sin\theta + G\sin^2\theta = (d(\theta)\cdot\nabla u)^2 \tag{3}$$

is maximal if d is parallel to ∇u. Clearly, maximizing $U(\theta)$ is equivalent to maximizing the quadratic form $d^T\nabla u\nabla u^T d$. The eigenvalues of the positive semidefinite matrix $\nabla u\nabla u^T$ are $\lambda_1 = |\nabla u|^2$ and $\lambda_2 = 0$. The corresponding eigenvectors $v_1 = \nabla u/|\nabla u|$ and $v_2 = \nabla u^{\perp}/|\nabla u|$ respectively indicate the orientations maximizing and minimizing gray-value changes, while the corresponding eigenvalues give the rate of the change. Since $\lambda_1 = |\nabla u|^2$, $\lambda_2 = 0$, $v_1 = \nabla u/|\nabla u|$, and $v_2 = \nabla u^{\perp}/|\nabla u|$, we have $\nabla u = \sqrt{\lambda_1}v_1$, $\nabla u\nabla u^T = \lambda_1 v_1 v_1^T$.

For multi-channel images, the following matrix is known as the structure tensor:

$$Q = \begin{bmatrix} E & F \\ F & G \end{bmatrix} = \sum_{i=1}^{S}\nabla u_i\nabla u_i^T. \tag{4}$$

If we use $\lambda_{1,2}$ to express the eigenvalues of the matrix Q, then $\lambda_{1,2}$ can be described as

$$\lambda_{1,2} = \tfrac{1}{2}\left((E+G) \pm \sqrt{(E-G)^2 + 4F^2}\right). \tag{5}$$

The corresponding eigenvectors $v_{1,2}$ are given by

$$v_{1,2} = (\cos\theta_{\pm}, \sin\theta_{\pm}), \tag{6}$$

where $\theta_+ = \theta_- + \pi/2$. According to the related theory of linear algebra, we obtain

$$\theta_- = \frac{1}{2}\left(\arctan\frac{2F}{E-G}\right). \tag{7}$$

In the case $S>1$, the matrix Q is positive semi-definite, therefore, the structure tensor defined in (4) can be described as

$$Q = \lambda_1 v_1 v_1^T + \lambda_2 v_2 v_2^T. \tag{8}$$

It gives the direction information and local geometry description of all input images. We know that the fused image $\tilde{u}$ should contain the basic geometry of the multi-channel image u, therefore, the structure tensor $\nabla\tilde{u}\nabla\tilde{u}^T$ of the fused image should approximate Q. Based on this conclusion, we have $\nabla\tilde{u}\nabla\tilde{u}^T = \lambda_1 v_1 v_1^T$ and $\nabla\tilde{u} = \sqrt{\lambda_1}v_1$ for gray-valued images, because the minimal change, namely, λ_2 is zero [21].

2.2. Geometry representation of input images

In Eq. (4), each of the source images contributes equally to the geometry description of all the inputs. Clearly, it is not reasonable