Optik - International Journal for Light and Electron Optics 172 (2018) 1–11

Contents lists available at ScienceDirect

Optik

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

Original research article

Structure-aware image fusion

Wen Li[a], Yuange Xie[a], Haole Zhou[a], Ying Han[b], Kun Zhan[a,*]

[a] School of Information Science and Engineering, Lanzhou University, Lanzhou, Gansu 73000, China
[b] Xunhua People's Hospital, Xunhua, Qinghai 811100, China

ARTICLE INFO

Keywords:
Multi-modal image fusion
Iterative joint filter
Joint filter
Scale-aware
Structure-preserving

ABSTRACT

Most existing multi-modal image fusion methods require multi-scale transforms. However, this requirement does not necessarily lead to the fusion result containing the original intensity of source images, and multi-scale transforms need a high computational complexity. In this paper, we tackle the problem of multi-modal image fusion in the spatial domain with a low computational complexity. A salient structure extraction method and a structure-preserving filter are developed to fuse medical images. The developed structure-preserving filter has a property that it recovers small-scale details of the guidance image in the neighborhood of large-scale structures of the input image. Based on the property of the structure-preserving filter, the fusion result is constructed by combining the output of the structure-preserving filter and the source images. Experiments are conducted to demonstrate the effectiveness of the proposed method in comparison with the state-of-the-art approaches in terms of three performance metrics.

1. Introduction

The core devices of cameras are widely mounted optical lenses. A lens with long focal lengths suffers from the problem of limited depth of field, which may result in only few objects with the depth of field are clear while others are blurred. The most commonly used sensor is visible-light camera and visible images have more small details because of illumination variation. With an infrared sensor, the infrared image provides object-level information due to thermal radiation. In multi-sensor imaging, an infrared image is detection of thermal radiation and visible part of the spectrum is captured to be a visible image. So it is necessary to exploit an effective structure-aware fusion method for fusing details of different scales while preserving salient structures. Since details and salient structures in different modalities need to be combined into one fused output, *Image fusion* is crucial to form a single summary output.

Multi-modal images are generally merged in the spatial domain or in the transform domain [1,2]. In both domains, an implicit assumption is that detecting salient regions is equated with finding the principal pixels in different source images. In the spatial domain, a simplest way of image fusion is to calculate the average intensities in different modalities but the average rule cannot effectively combine details into output. This approach preserves the basis structures but suffers from metamerism, *i.e.*, different input intensities are assigned the same output intensity. Image gradient is a simple way to represent image structure in the spatial domain [3], and is adopted as a basis technique in image fusion methods [4–6]. Image structures in different scales can be easily extracted by gradient, and the pixel with the larger gradient are usually selected from multi-modality images in spatial domain. Salient structure of different scales can be captured by using multi-scale decomposition (MSD) transforms. In transform-domain fusion methods, relatively large coefficients which carry salient structures are selected from transform-domain subbands, and the fused image is obtained by corresponding inverse transforms with these selected coefficients. With the development of different MSD transforms, many MSD-

* Corresponding author.
E-mail address: ice.echo@gmail.com (K. Zhan).

https://doi.org/10.1016/j.ijleo.2018.06.123
Received 2 December 2017; Received in revised form 18 April 2018; Accepted 27 June 2018

based methods are applied to image fusion [7–9], such as algorithms based on Laplacian pyramid [10–13], wavelet [14–16], multiscale geometric analysis [17–19], *etc*. However, transform domain methods cannot preserve structures well in spite of the development of multiscale geometric analysis tools [20–23], they have a relatively high-computational complexity [24–26].

Scale-aware tools are widely applied to image processing and computer vision for extracting semantic information [27–30]. It is straightforward to check that the image gradient contains all image scales but MSD only captures few image scales determined by the decomposition levels. Fusing images in the spatial domain is different from fusing in the transform domain because the fused image of the former contains all the scales of the original pixel-value from the source images. In the spatial domain, we can directly process pixels rather than processing the transform domain coefficients. Thus, we propose a fast and effective spatial domain image fusion framework, and the main techniques of the framework are Salient Structure Extraction (SSE) and Iterative Joint Filter (IJF). We propose a simple and efficient SSE method, in which we utilize image gradient to obtain large-scale structures of the source image. Then, we use IJF to transfer the small-scale details from the source image to the fused output. The output of IJF is used as a weighted map and the fusion result is obtained by a weighed-sum rule. Both SSE and IJF are developed in a fast and effective way.

The effectiveness of the algorithm is evaluated on many multi-modal images for the image fusion problem. The proposed structure-aware image fusion method is compared with the state-of-the-art algorithm: directive contrast-based method in contourlet domain (DCCD) [18], neuron-fuzzy-based method in contourlet domain (NFCD) [24], two-scale decomposition (TSD) method [13], multi-scale transform using sparse representation (MSSR) [31], gradient transfer fusion (GTF) [32], respectively. The experimental results indicate the performance of the proposed structure-aware fusion scheme is better than the state-of-the-art methods in term of three performance measures and it is the fastest algorithm.

The reset of the paper is organized as follows. In Section 2, we review the bilateral filter and domain transform filter, and analyse the relation between bilateral filter and domain transform filter. In Section 3, we propose a novel framework to fuse multi-modal images. In Section 4, we presents numerical experiments and comparison results. We use many images and compare with four state-of-the-art methods. Section 5 concludes with some discussion.

2. Preliminaries

The pioneer work of anisotropic diffusion is a structure-aware smoothing technique which can preserve image structures [33]. Anisotropic diffusion tends to over-sharpen structures and has a high computational complexity [34,35], so the improved schemes for structure-preserving smoothing filters, the bilateral filter [36], the guided filter [37], and domain transform filter [38] are proposed recently. The bilateral filter is generalized to joint bilateral filter [39], then the three filters have two inputs, i.e., an *input image* and a *guidance image*. The two inputs are jointly filtered to obtain one output. They are three popular structure-preserving joint image filters. They can preserve large-scale salient structures of the input image, while small-scale details of the guidance image is transferred to the output image [37–40].

Image filtering is performed on the image pixels encompassed by a window and updating the center pixel value. Given an input image I and a guidance image G, the output of the bilateral filter for a pixel p is,

$$J_p = \frac{1}{k_p}\sum_{q\in\Omega_p} I_q\left(\exp\left(-\left(\frac{\|p-q\|_2^2}{2\sigma_s^2}+\frac{\|G_p-G_q\|_2^2}{2\sigma_r^2}\right)\right)\right) \tag{1}$$

where Ω_p is a sliding window centered at the pixel p, σ_s is a space domain scaling parameter, σ_r is a range domain scaling parameter, and k_p is a normalizing term:

$$k_p = \sum_{q\in\Omega_p}\left(\exp\left(-\left(\frac{\|p-q\|_2^2}{2\sigma_s^2}+\frac{\|G_p-G_q\|_2^2}{2\sigma_r^2}\right)\right)\right). \tag{2}$$

The mechanism of the bilateral filter is easier to be understood than the guided filter, *i.e.*, support that one intensity $G_q\in\Omega_p$ is similar to the intensity G_p at the centre of Ω_p and the two pixels are located at the same side of an edge. In this case, $\exp\left(-\frac{\|G_p-G_q\|_2^2}{2\sigma_r^2}\right)$ tends to 1 and the weight of I_q is determined by $\exp\left(-\frac{\|p-q\|_2^2}{2\sigma_s^2}\right)$, so it is similar to a conventional Gaussian smoothing filter. On the other hand, if an intensity G_q is largely different from G_p and the two pixels are located at the different sides of an edge, $\exp\left(-\frac{\|G_p-G_q\|_2^2}{2\sigma_r^2}\right)$ tends to 0 and it renders the weight of I_q tend to 0. Thus, the relationship between I_q and I_p is very weak so that the intensity value of I_q almost has no contribution to the filter result J_p. The similar intensities in the local neighborhood Ω_p are smoothed while the intensities at the different sides of the edge almost have no influence on each other. Therefore, joint filters are structure-aware smoothing filters.

Gastal and Oliveira proved that the domain transform filter can obtain comparable results as same as the joint bilateral filter [38]. The domain transform filter is faster than the fastest bilateral filtering algorithm of equivalent quality [38].

The power term of (1) can be defined to a distance metric:

$$d_{\mathrm{pq}} = \frac{\|p-q\|_2^2}{2\sigma_s^2}+\frac{\|G_p-G_q\|_2^2}{2\sigma_r^2}. \tag{3}$$

(3) is transformed and rewritten by,