Information Fusion 30 (2016) 15–26

Contents lists available at ScienceDirect

Information Fusion

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

Perceptual fusion of infrared and visible images through a hybrid multi-scale decomposition with Gaussian and bilateral filters

Zhiqiang Zhou, Bo Wang*, Sun Li, Mingjie Dong

School of Automation, Beijing Institute of Technology, 5 South Zhongguancun Street, Haidian District, Beijing 100081, PR China

ARTICLE INFO

Article history:
Received 29 July 2014
Revised 8 September 2015
Accepted 5 November 2015
Available online 12 November 2015

Keywords:
Multi-scale decomposition
Image fusion
Gaussian filter
Bilateral filter
Human perception

ABSTRACT

In order to achieve perceptually better fusion of infrared (IR) and visible images than conventional pixel-level fusion algorithms based on multi-scale decomposition (MSD), we present a novel multi-scale fusion method based on a hybrid multi-scale decomposition (hybrid-MSD). The proposed hybrid-MSD transform decomposes the source images into multi-scale texture details and edge features by jointly using multi-scale Gaussian and bilateral filters. This transform enables to better capture important multi-scale IR spectral features and separate fine-scale texture details from large-scale edge features. As a result, we can use it to achieve better fusion result for human visual perception than those obtained from conventional multi-scale fusion methods, by injecting the multi-scale IR spectral features into the visible image, while preserving (or properly enhancing) important perceptual cues of the background scenery and details from the visible image. In the decomposed information fusion process, three different combination algorithms are adaptively used in accordance to different scale levels (i.e., the small-scale levels, the large-scale levels and the base level). A regularization parameter is introduced to control the relative amount of IR spectral information injected into the visible image in a soft manner, which can be adjusted further depending on user preferences. Moreover, by testing different settings of the parameter, we demonstrate that injecting a moderate amount of IR spectral information with this parameter can actually make the fused images visually better for some infrared and visible source images. Experimental results of both objective assessment and subjective evaluation by human observers also prove the superiority of the proposed method compared with conventional MSD-based fusion methods.

1. Introduction

Infrared (IR) and visible image fusion is an important technique in multi-sensor information fusion applications. Since IR sensors are able to capture thermal information in a scene that is not directly seen by human eyes, they can more clearly detect some objects in low-light, occlusion and adverse weather conditions. Visible imagery normally provides more details of the scene in the visible spectrum, and also presents more natural intensities and contrasts that are consistent with human visual perception. Integrating IR and visible information into a fused image allows us to construct a more complete and accurate description of the scene. However, due to heat emissions and differing spectral sensitivities, the relative luminance response in the IR spectrum is quite inconsistent with that in the visible spectrum, which makes the IR imagery hard to be interpreted. As a result, the fusion result of IR and visible imageries may also be visually unpleasing for human observers. Therefore, besides determining the best way to take full advantage of all information of the two source images in the fusion process, a more significant task should be to make the fused image easy to be interpreted, and thus can lead to better situation awareness.

To fuse thermal IR and visible imageries, many algorithms have been proposed by considering principles of biological color vision. Waxman et al. [1] introduce opponent-color neural networks to produce the fused image with false colors. Since a proper color representation of the fused image can allow for better scene interpretation, color remapping is usually used additionally in these methods to generate more natural colors. However, a highly natural coloring is difficult since the IR relative luminance response is quite different with that in the visible spectrum. With the assistance of a natural daytime color reference image, Toet et al. [2,3] managed to use a color mapping method to obtain natural color representation of the night-vision fusion results.

A more common approach for pixel-level image fusion is conducted in grayscale space based on various multi-scale decomposition (MSD) transforms, e.g., Laplacian pyramid (LAP) [4], gradient

* Corresponding author. Tel.: +86 010 68914781.
E-mail address: wangbo.bitauto@gmail.com (B. Wang).

http://dx.doi.org/10.1016/j.inffus.2015.11.003
1566-2535/

Fig. 1. The results of our fusion method and the conventional DT-CWT fusion method on the "UN Camp" source images.

pyramid (GP) [5], wavelet transform [6] and support vector transform [7]. For these MSD-based fusion methods, MSD transforms are performed first on each source image to obtain different subbands containing the decomposed information of different frequencies or orientations. The corresponding subbands of all source images are then combined together based on certain fusion rules. Finally, the fused images can be produced by inverse MSD transforms. To merge thermal and visual images, Toet et al. [8] propose one of the earliest MSD transform named contrast pyramid or ratio of low pass pyramid (ROLP). The ROLP is formed as a ratio of each level of Gaussian pyramid to its next coarser-level one. Fusion with ROLP enables to identify and preserve important high-contrast details that are more relevant to visual perception.

Other more complex MSD transforms including dual-tree complex wavelet (DT-CWT), curvelet and contourlet have also been successfully applied in image fusion [9–13]. These MSD transforms achieve better properties of shift invariance and directional selectivity, which are essential for multispectral image fusion since they enable to correctly transfer more directional details into the fused image. However, since the IR imagery usually presents large differences in relative luminance response (and thus may generate coarser-scale information, rather inconsistent contrasts or different edges and contours, etc.) against that of the corresponding visible imagery, directly combining the two image salient features based on these conventional MSDs may clutter the image content in the fusion result and make it visually unpleasing (e.g., blurring of details, introducing unnatural contrasts or visual artifacts).

In this paper, we present a novel multi-scale fusion method based on a hybrid MSD transform (hybrid-MSD) to achieve better fusion results for human visual perception. Unlike the previous MSD transforms that attempt to capture more directional information with comparatively more complex filters, the hybrid-MSD decomposes the source image into texture details and edge features at multiple scales by jointly using multi-scale Gaussian and bilateral filters. In a perceptual evaluation of different image fusion schemes, Toet et al. [14] indicate that the IR imagery serves best for target detection and recognition, whereas the visible imagery contributes most to global scene awareness. Our method manages to employ the hybrid-MSD as well as a novel asymmetrical multi-scale fusion scheme to inject the important IR spectral features into the visible image, while preserving (or properly enhancing) important perceptual cues of the background scenery and details captured from the visible spectrum. Thus, it would lead to perceptually better fusion results for human interpretation.

A fusion example can be seen in Fig. 1. While the conventional DT-CWT method tends to transfer all the irrelevant details from the infrared image into the fused image and make the visible details of the background scenery (e.g., the fence and trees labeled in the rectangles of Fig. 1(d)) less clear for visual perception, our fusion method better preserves the background scenery and details from the visible image overall (see Fig. 1(c)). The rest of this paper is organized as follows. In Section 2, the concept and construction of the hybrid-MSD based on Gaussian and bilateral filters are presented. In Section 3, we can see the hybrid-MSD can be used to select important multi-scale IR spectral features from the infrared image. Hence, an asymmetrical fusion scheme is then proposed to inject the selected IR spectral features into the visible image. By cooperating with three different combination algorithms adapted to different decomposed-information scale levels, we can simultaneously preserve or properly enhance the background scenery and details from the visible image in the fusion process. Experimental results and comparisons are given in Section 4. Conclusions are presented in Section 5.

2. The hybrid-MSD based on Gaussian and bilateral filters

Gaussian and bilateral filters are both known as important filtering methods extensively used in image processing applications. The Gaussian filter is one of the basic tools for noise reduction and image smoothing. The linear scale-space representation of an image is constructed via the convolutions with a series of Gaussian kernels [15]. Computing with repeated Gaussian filtering and downsampling, we can also obtain the Gaussian pyramid. Based on this pyramid structure, several MSD transforms including the LAP, ROLP and GP have been proposed and employed in image fusion.

The bilateral filter is a non-linear edge-preserving filter that can smooth small-scale details while preserving strong edges [16]. It filters images in the spatial and range domains simultaneously through the combination of spatial and range Gaussian kernels. With a Gaussian function denoted as $g_\sigma(x) = \exp(-x^2/\sigma^2)$, the bilateral filtering of image I at pixel $\mathbf{p}$ is performed as:

$$I_b(\mathbf{p}) = \frac{1}{W} \sum_{\mathbf{q}\in\Omega} g_{\sigma_s}(\|\mathbf{p}-\mathbf{q}\|) g_{\sigma_r}(|I(\mathbf{p}) - I(\mathbf{q})|) I(\mathbf{q}), \tag{1}$$

with

$$W = \sum_{\mathbf{q}\in\Omega} g_{\sigma_s}(\|\mathbf{p}-\mathbf{q}\|) g_{\sigma_r}(|I(\mathbf{p}) - I(\mathbf{q})|), \tag{2}$$

where σ_s and σ_r are the standard deviations of the spatial and range Gaussians, which control the influences of neighboring pixel $\mathbf{q}$ in terms of spatial and intensity differences, respectively. $\Omega \in R^2$ represents the image domain.

By setting a relatively small value of σ_r, we can obtain a bilateral filtered image I_b that retains the "large-scale features" separated by strong edges [17], and removes the fine-scale details that are dominated by texture features. Furthermore, assuming I_g is the corresponding Gaussian filtered image that is computed by solely using the spatial Gaussian g_{σ_s}, we can see that I_b contains certain additional edge information compared with I_g. Consequently, we can obtain the removed fine-scale texture details D^0 and the additional edge features D^1 retained in I_b by the following subtractions respectively:

$$D^0 = I - I_b, \tag{3}$$