Infrared Physics & Technology 56 (2013) 93–99

Contents lists available at SciVerse ScienceDirect

Infrared Physics & Technology

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

Fusion of visible and infrared images using saliency analysis and detail preserving based image decomposition

Jufeng Zhao, Qiang Zhou, Yueting Chen, Huajun Feng *, Zhihai Xu, Qi Li

State Key Lab of Modern Optical Instrumentation, Zhejiang University, Hangzhou 310027, China

HIGHLIGHTS

- ▶ Multi scale detail preserving decomposition is constructed using image filter.
- ▶ Saliency map is designed with pixel value difference.
- ▶ Images are fused with saliency weight map at different scale levels.
- ▶ The results prove the good performance of the proposed algorithm.

ARTICLE INFO

Article history:
Received 12 October 2012
Available online 23 November 2012

Keywords:
Image fusion
Visible and infrared images
Saliency map
Multi scale decomposition

ABSTRACT

Image fusion for visible and infrared images is a significant task in image analysis. The target regions in infrared image and abundant detail information in visible image should be both extracted into the fused result. Thus, one should preserve or even enhance the details from original images in fusion process. In this paper, an algorithm using pixel value based saliency detection and detail preserving based image decomposition is proposed. Firstly, the multi-scale decomposition is constructed using weighted least squares filter for original infrared and visible images. Secondly, the pixel value based saliency map is designed and utilized for image fusion in different decomposition level. Finally, the fusion result is reconstructed by synthesizing different scales with synthetic weights. Since the information of original signals can be well preserved and enhanced with saliency extraction and multi scale decomposition process, the fusion algorithm performs robustly and excellently. The proposed approach is compared with other state-of the-art methods on several image sets to verify the effectiveness and robustness.

1. Introduction

Image fusion is an important research topic in image analysis, pattern recognition and computer vision. The fusion process combines different information from two or more images of a scene, which may be acquired at different situations with one sensor, or from multi sensors [1–3]. And infrared (IR) image usually contains important and particular target information which one could not find from visible (VI) image. However, the VI image which is captured with abundant object details also should be considered for information extraction. So, the fusion process is necessary to maintain the both advantages of IR and VI images [4]. Lots of fusion algorithms have been proposed. Multi-resolution approaches are very common, especially the wavelet and curvelet transform based approaches [5–8], and various pyramid algorithms such as contrast pyramid [9], gradient pyramid [10], ratio pyramid [11], and morphological pyramid [12]. Those kinds of method usually utilize the multi-scale analysis for image details extraction, which is helpful for a good image fusion. The disadvantage is that they would smooth some image details. Wang et al. [13] proposed a method based on dual-channel pulse coupled neural networks (dual-channel PCNN) for image fusion, which is sometimes complicated and time-consuming and especially for multi-focus fusion. Principal component analysis (PCA) and independent component analysis (ICA) based algorithms try to fuse the image with extracting the main information from original images [14–16]. As some information is lost in PCA and ICA process, the fusion results would lose details, which is not suitable for IR and VI image fusion. Morphology theory is also used for image fusion. Recently, the approach through region extraction by using multi scale center-surround top-hat transform for IR and VI image fusion is proposed [4]. The center-surround top-hat transform could well extract image regions and preserving details of original IR and VI images [17]. However, the fusion result quite depends on the three selectable parameters, which have a wide range of values. And it is difficult for inexperienced users. The visual saliency detection is popular

* Corresponding author. Tel./fax: +86 571 87951182.
E-mail address: fenghj@zju.edu.cn (H. Feng).

1350-4495/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.infrared.2012.11.003

recently [18–21]. Saliency extraction is an important step in many computer vision tasks including object recognition, image segmentation and adaptive compression. Some researchers have applied saliency preserving in multi-focus image fusion [22–24], even though the saliency features map is coarsely obtained using local Gaussian difference. Actually, since the saliency maps extracted by most methods are not fine enough with a low resolution, the saliency method used in image fusion is infrequent. And we have tried to solve this problem, but not for fusion of IR and VI images [25].

To well utilize the ability of saliency extraction for image processing, an algorithm using pixel value based saliency detection and multi scale edge preserving decomposition is proposed. The saliency feature extraction would give different weights to each pixel, and extract the target regions which attract people's attention or interest, which will greatly improve the visual quality of fusion result. Meanwhile, an edge preserving based multi scale analysis approach is introduced to combine with the above saliency detection operation. This multi scale decomposition method could help preserve even enhance the details of original IR and VI images.

In this paper, the fusion algorithm using saliency extraction and edge preserving based multi scale decomposition is structured as follows. Firstly, the multi-scale decomposition is applied on the original IR and VI image by using weighted least squares (WLS) filter with different parameters. Secondly, the pixel value based SWM (saliency weight map) is utilized for image fusion in different decomposition level. Finally, the fusion result is reconstructed by doing image synthesis with synthetic weights. Comparison experiments are used to verify the validness of proposed fusion approach.

2. Mathematical theory

2.1. Edge-preserving filters

Traditionally methods, such as Laplacian pyramid [26], utilize multi-scale decompositions for images analysis. However, since those pyramids are constructed using linear filters, the results usually produce halo artifacts near edges. And the artifacts may be suppressed by using non-linear edge-preserving smoothing filters, such as the bilateral filter [27] and WLS [28,29].

Farbman et al. proposed an alternative edge-preserving operator based on the WLS framework [30]. It can well construct multi-scale decompositions with detail preserving at a variety of scales. For an original image f, this method tries to seek an image g that is as close as possible to f, meanwhile is as smooth as possible everywhere except those significant gradients in f. The image g can be defined as:

$$g = \arg\min_g \{\|g - f\|_2^2 + \lambda(w_x\|D_x g\|_2^2 + w_y\|D_y g\|_2^2)\} \quad (1)$$

where $\| \ \|_2$ denotes the second order norm operator, D_x and D_y are difference operators. The regularization (the second term) is to obtain smoothness by minimizing the partial derivatives of g. The smoothness weights w_x and w_y are determined by g. λ is the regularized factor to make balance between two terms.

According to Eq. (1), one can obtain the image g.

$$g = G_\lambda(f) = (I + \lambda H)^{-1} f \quad (2)$$

where $H = D_x^T w_x D_x + D_y^T w_y D_y$.

As λ increases, the output image g becomes smoother.

2.2. Pixel value based saliency map definition

The human vision system (HVS) is sensitive to contrast in images. For every image, the regions that have large center-surround contrast usually attract people's attention. According to this, the saliency value of image pixels is calculated upon the gray value contrast between image pixels. The saliency value of pixel p of image f is defined as:

$$W_p = \sum_{\forall q \in f} F(p,q) \quad (3)$$

where q is an arbitrary pixel of image f. $F(p, q)$ represents the pixel value distance between pixel p and q, which is similar to the color distance metric [31]. $F(p, q)$ can be expressed as the difference of pixels value in image f.

$$\mathrm{F}(p,q) = |f_p - f_q| \quad (4)$$

Calculating all the pixels in f with Eq. (3), we can finally get a saliency image. M_f is normalized version of this saliency image. Finally, for arbitrary pixel p in M_f, $W_p \in (0, 1)$. Since this image reflects the weight distribution of people's attention upon original image f, we call the M_f saliency weight map (SWM). Fig. 1 shows an example of SWM of IR and VI images. Figure (a) and (b) are the original IR and VI images, the SWM of which are shown in (c) and (d), respectively.

3. Method

3.1. Multi-scale analysis

With the edge-preserving operator we mentioned in Section 2.1, one can easily construct a multi-scale decomposition. Take IR image for example, we transform the IR image into multi-scale representation as shown in Fig. 2. Let f_{IR} and f_{VI} denote input IR and VI images. And we want to construct a $(n + 1)$-level decomposition. The kth ($k = 1, 2 \ldots n$) scale approximation coarse images are obtained by using Eq. (2).

$$g_{\mathrm{IR}}^k = G_{\lambda_k}(f_{\mathrm{IR}}) \quad (5)$$

$$g_{\mathrm{VI}}^k = G_{\lambda_k}(f_{\mathrm{VI}}) \quad (6)$$

where λ_k represents the regularized factor at kth level WLS, $\lambda_k < \lambda_{k-1}$.Treating the g_{IR}^n and g_{VI}^n as base layers, the detail layers can be defined as:

(a) IR image (b) VI image (c) SWM of (a) (d) SWM of (b)

Fig. 1. An example of SWM extraction.