Infrared Physics & Technology 71 (2015) 363–369

Contents lists available at ScienceDirect

Infrared Physics & Technology

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

Fusion for visible and infrared images using visual weight analysis and bilateral filter-based multi scale decomposition

Zaifeng Shi [a], Jiangtao Xu [a], Yu zhang [b], Jufeng Zhao [b,*], Qing Xin [b]

[a] School of Electronic and Information Engineering, Tianjin University, Tianjin, PR China
[b] Institute of Electronic and Information, Hangzhou Dianzi University, Hangzhou, PR China

HIGHLIGHTS

- Multi scale decomposition is constructed using simple bilateral filter.
- The authors extract image detail information by visual weight analysis.
- Visual weight map is designed with local frequency-tuned method.
- Images are fused with visual weight map at each scale level.
- The results demonstrate the good performance of the proposed method.

ARTICLE INFO

Article history:
Received 12 February 2015
Available online 1 June 2015

Keywords:
Image fusion
Visual weight map
Multi scale decomposition

ABSTRACT

Fusion for visible and infrared images has been an important and challenging work in image analysis. Both the feature information in infrared image and abundant detail information in visible image should be preserved and enhanced in fused result. In this paper, a detail enhanced fusion algorithm through visual weight analysis based on smooth-inspired multi scale decomposition is proposed. With variable parameter, bilateral filter-based idea successfully decomposes the two source image into several scales. At each scale level, visual weight map is calculated and used for fusion. Finally, those levels are synthetized with proper weights. Using this idea, the detail information could be enhanced easily. The experimental results demonstrate the proposed approach performs better than other methods, especially in visual effect and keeping details.

1. Introduction

Image fusion plays an important role in multi-source image characteristic extraction and expression. People make effort in this research to combine feature information from two or more source images, which sometimes are captured by different sensors [1,2]. Visible image (VI) usually contains abundant object details. And infrared (IR) image mirrors particular target information which we cannot find in VI ones. Fusion for visible and infrared images aims to maintain the both advantages of VI and IR images [3].

Many fusion algorithms have been proposed. The most common methods are multi resolution-based ideas, such as wavelet and curvelet transform based approaches [4,5]. Another kind of multi resolution approaches are various pyramid-based algorithms, including contrast pyramid [6], ratio pyramid [7], gradient pyramid [8] and morphological pyramid [9]. But these multi scale usually do down sampling and up sampling, the details would be smoothed. The theory of shearlets has been studied [10], and this novel transform can be well applied in image fusion [11], though it is sometimes time-consuming. Researchers consider the PCA(principal component analysis) analysis to extract the main information in image fusion [12], however, the result will lose details due to the information lose when using PCA. Making use of region extraction by using multi scale center-surround top-hat transform, Bai proposed an excellent method for IR and VI image fusion [3]. However, the parameters are difficult to select if user is not familiar with their method sometimes. Moreover, saliency extraction and visual weight is quite popular, and this research is important in object recognition and adaptive compression. People have tried to use saliency preserving in multi-focus image fusion [13]. Especially, Zhao and his partners develop several visual saliency-based image enhancement [14,15] and image fusion [16,17] method. With saliency extraction and visual weight designing for image feature extraction, the fused result not only looks good, but also the objective evaluation is very good.

* Corresponding author. Tel./fax: +86 571 86919029.
E-mail address: dabaozjf@hdu.edu.cn (J. Zhao).

http://dx.doi.org/10.1016/j.infrared.2015.05.015
1350-4495/

In order to well extract the feature information in IR and VI images, an algorithm using local frequency-tuned-based visual weight detection and non-band-limited multi scale decomposition is proposed. The visual weight map is obtain, to give different weights to each pixel and region. Those potential target area which attract people's attention or interest, will be given large value in visual weight map, which will greatly improve the visual effect of result. Moreover, making use of bilateral filter, an edge preserving multi scale decomposition is constructed, without any up sampling or down sampling. This kind of multi scale decomposition method could help enhance and highlight the detail and characteristic information in IR and VI images.

In this paper, the fusion method using visual weight analysis and multi scale decomposition is structured as follows. In Section 2, the theory for bilateral filter is introduced, and we also describe how to extract visual weight map. Using multi scale decomposition and visual weight map, image fusion approach is designed in Section 3. Section 4 shows the experimental results and comparisons. And we do the conclusion in Section 5.

2. Mathematical theory

2.1. Edge-preserving smoothing via bilateral filter

How to preserve edges is a key problem in image smoothing. In traditional idea, multi-scale decompositions are usually constructed with linear filters for images analysis, such as Laplacian pyramid [18,19]. However, the results usually produce halo artifacts near edges due to the linear filters. Using non-linear edge-preserving smoothing filters, those artifacts could be reduced.

Bilateral filter is a typical non-linear smooth filter, which does well in extracting details [20]. This filter is operated locally rather than globally. With this filter in local window, each pixel in image has a spatial support contributed from neighboring pixels. The filter result is determined by not only the relative distance between center pixel and its neighbor pixels, but also the gray different between them [21,22].

Within local area of input image I, the bilateral filter output f at coordinate (i, j) can be defined as:

$$f_{ij} = \frac{1}{M_{ij}} \sum_{p \in \Omega} \{G_s[D_s(p), \sigma_s] G_v[D_v(p), \sigma_v] I_p\} \tag{1}$$

where Ω denotes the spatial support which called local area, and (i, j) is center pixel of, Ω, p represents arbitrary pixel in Ω which means the neighboring pixel in the support. In Eq. (1), I_p is the intensity value at position p, and $G_s[D_s(p), \sigma_s]$ and $G_v[D_v(p), \sigma_v]$ are two gaussian functions, respectively. A normalization term M_{ij} is determined by:

$$M_{ij} = \sum_{p \in \Omega} \{G_s[D_s(p), \sigma_s] G_v[D_v(p), \sigma_v]\} \tag{2}$$

The gaussian function $G_s[D_s(p), \sigma_s]$ is designed for spatial application, $D_s(p)$ is the size, and σ_s is deviation. $D_s(p)$ denotes the spatial distance between center pixel (i, j) and p. While the gaussian function $G_v[D_v(p), \sigma_v]$ is designed for intensity value application, $D_v(p)$ is the size, and σ_v is deviation. $D_s(p)$ is the distance of intensity value between center pixel (i, j) and p:

$$D_s(p) = |I_{ij} - I_p| \tag{3}$$

From the above expression, we can learn that $G_s(\cdot)$ and $G_v(\cdot)$ determine the impact of neighbor pixels on center pixel (i, j).

With the change of deviation σ_s and σ_v, different smoothing results can be obtained. Fig. 1 shows the smoothing results with different σ_s and σ_v for Gaussian function. (a) is the original image, and (b)–(d) are filtered results with $\sigma_s = 3$ and $\sigma_v = 0.05$, $\sigma_s = 12$ and $\sigma_v = 0.05$, $\sigma_s = 12$and $\sigma_v = 0.15$, respectively. One can conclude that, larger σ_s or σ_v can make the result more smooth.

According to above description, we express the smooth result f as the function of input image I and parameter σ_s and σ_v for short, instead of complicated expression like Eq. (1)

$$f = S(I, \sigma_s, \sigma_v) \tag{4}$$

The larger σ_s or, σ_v the more smooth the result is, especially σ_v. Typically σ_s is within the range [3, 15], and $\sigma_v \in [0, 0.2]$.

2.2. Visual weight map design

Learning from theory of psychology, human visual system (HVS) could simply extract the salient area from scene. And HVS seems sensitive to image contrast, such as intensity or color. If we could detect the visual interest of images with algorithm, it is very helpful for our image processing. Visual weight map (V_{map}) is considered as a matrix that mirrors the weight distribution of HVS focusing upon image I.

With frequency-tuned idea [14,23], one could obtain full resolution V_{map} with well-defined boundaries of salient objects. From this theory, the largest salient object would be highlighted to form a V_{map}.

Making good use of Gaussian band-pass function $G_{\text{band}}(\cdot)$, a visual weight map M is obtained from original image I as following equation:

$$M = |I * G_{\text{band}}(r, \sigma_1, \sigma_2)| \tag{5}$$

where $*$ is a convolution operator. Since $G_{\text{band}}(\cdot)$ is a band-pass function, it has two cut-off frequency values, that is low frequency cut-off value υ_{low} and high frequency cut-off value υ_{high}. We define this band-pass function as follows:

$$G_{\text{band}}(r, \sigma_1, \sigma_2) = G(r, \sigma_1) - G(r, \sigma_2) \tag{6}$$

where $G(\cdot)$ is classical Gaussian function, with parameter r and σ. r is the size that $r^2 = i^2 + j^2$, i and j denote coordinates of pixel, σ is the standard deviation of the function. Since it is band-pass function, it is clear that $\sigma_1 > \sigma_2$ in Eq. (6). And σ_1 determines the low frequency

Fig. 1. Smoothing result with different σ_s and σ_v for Gaussian function. (a) Original image, (b) smoothing result with $\sigma_s = 3$ and $\sigma_v = 0.05$, (c) smoothing result with $\sigma_s = 12$ and $\sigma_v = 0.05$, (d) smoothing result with $\sigma_s = 12$ and $\sigma_v = 0.15$.