ISPRS Journal of Photogrammetry and Remote Sensing 105 (2015) 79–90

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

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

Human visual system consistent quality assessment for remote sensing image fusion

Jun Liu [a], Junyi Huang [c], Shuguang Liu [b,*], Huali Li [d], Qiming Zhou [a,c], Junchen Liu [e]

[a] Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Guangdong 518055, China
[b] Chongqing Institute of Green and Intelligent Technology, Chinese Academy of Sciences, Chongqing 400714, China
[c] Department of Geography, Hong Kong Baptist University, Hong Kong, China
[d] College of Electrical and Information Engineering, Hunan University, Hunan 410082, China
[e] Tianjin Institute of Surveying and Mapping, Tianjin 300381, China

ARTICLE INFO

Article history:
Received 24 June 2014
Received in revised form 13 November 2014
Accepted 22 December 2014
Available online 14 April 2015

Keywords:
Image fusion
Quality assessment
Human visual system
Gaussian scale space
Spatial quality index
Spectral quality index

ABSTRACT

Quality assessment for image fusion is essential for remote sensing application. Generally used indices require a high spatial resolution multispectral (MS) image for reference, which is not always readily available. Meanwhile, the fusion quality assessments using these indices may not be consistent with the Human Visual System (HVS). As an attempt to overcome this requirement and inconsistency, this paper proposes an HVS-consistent image fusion quality assessment index at the highest resolution without a reference MS image using Gaussian Scale Space (GSS) technology that could simulate the HVS. The spatial details and spectral information of original and fused images are first separated in GSS, and the qualities are evaluated using the proposed spatial and spectral quality index respectively. The overall quality is determined without a reference MS image by a combination of the proposed two indices. Experimental results on various remote sensing images indicate that the proposed index is more consistent with HVS evaluation compared with other widely used indices that may or may not require reference images.

1. Introduction

Both high-spatial and high-spectral resolution images are widely used in remote sensing applications, such as land use classification, urban planning, environmental monitoring, and resources management. However, the trade-off between spatial resolution and spectral resolution remains due to technical constraints (Lillesand et al., 2004). To overcome this problem, image fusion, which is defined as "the combination of two or more different images to form a new image by using a certain algorithm" (van Genderen and Pohl, 1994), is proposed and proven to be a feasible approach to produce images with high spatial and spectral resolution (Wald et al., 1997; Wald, 2002).

Many image fusion methods have been proposed and have been applied in diverse fields over the past decades (Zhang et al., 2013; Hartfield et al., 2011; Gumma et al., 2011; Witharana et al., 2013). Generally, these methods can be divided into three categories (Khan et al., 2009): (1) Component substitution (CS)-based methods (Laben et al., 2000; Tu et al., 2001, 2004), (2) Multiresolution-analysis (MRA)-based methods (Zhou et al., 2014; Aiazzi et al., 2006; Nunez et al., 1999; Otazu et al., 2005; González-Audícana et al., 2005; Nencini et al., 2007) and (3) the methods that make use of both CS and MRA (González-Audícana et al., 2004; Zhang et al., 2009).

Apart from image fusion methods, another important but difficult task is image fusion quality evaluation (Chavez et al., 1991; Eskicioglu et al., 1995). During image fusion quality assessment, theoretically, the fused image should be compared with the ideal multi-spectral (MS) reference image with the same spatial resolution as the panchromatic (PAN) image. However, such a reference image may not readily available. According to Wald's protocol (Wald et al., 1997), spectral quality should be measured on low-frequency components, whereas spatial quality should be measured on high-frequency components. Many indices have been proposed to evaluate the fusion performance, such as Correlation Coefficients (CC) (Ribeiro et al., 2012), root-mean-square error (RMSE) (Zhu et al., 2013), relative average spectral error (RASE) (Chen et al., 2009), Erreur Relative Golbale Adimensionnelle de Synthese (ERGAS, or relative global dimensional synthesis error) (Wald, 2002; Li et al., 2011), Spectral Angle Mapper (SAM), Universal Image Quality Index (UIQI) (Wang et al., 2002), and UIQI-based metrics (Piella et al., 2003; Xydeas et al., 2000), as well

* Corresponding author.
E-mail address: liushuguang@cigit.ac.cn (S. Liu).

http://dx.doi.org/10.1016/j.isprsjprs.2014.12.018

as Q4 (Wang et al., 2002) and mean structural similarity index (MSSIM) (Wang et al., 2004).

However, Alparone et al. (2008) suggested that evaluating the fusion quality by degrading the original PAN and MS images is inappropriate for high-resolution image fusion. Therefore, a quality evaluation method known as "Quality with no reference (QNR)" was proposed (Alparone et al., 2008) using spatial distortion D_s and spectral distortion D_λ. The reference image and degrading of the spatial resolution of the original PAN and MS images are no longer required. The interband spectral quality of fused data is assumed as the similarity relationship among unchanged bands after fusion. Analogous to QNR, Shah et al. (2008) and Khan et al. (2009) proposed quality assessment indices at full resolution.

These metrics are widely used in literatures, however, in some cases, the assessment results of these metrics for the same fused image may contradict with each other, resulting in that an image with incorrect color or poor details may have the best objective quality evaluation. Thus, it is important to incorporate the human visual system as a paradigm of fusion quality assessment (Chen et al., 2009). Therefore, in this paper, a novel image fusion quality assessment is presented using the Gaussian Scale Space (GSS) technology while accounting for the human visual mechanism. The spatial details and spectral information are firstly separated in the GSS, and evaluated using the proposed spatial and spectral quality indices respectively. The overall evaluation is calculated based on these two indices. The highlight of the proposed method is that the evaluation results are consistent with human visual system, and the assessment processing does not require a reference.

Section 2 reviews the Human Visual System (HVS) and GSS technology. Section 3 presents the quality assessment method. Section 4 illustrates the fusion results obtained by popular fusion methods and compares the quality assessment capacity between the proposed index and the QNR, as well as other quality assessments that require a reference. The final section draws a conclusion based on the obtained results.

2. Human Visual System (HVS) and GSS technology

The multi-scale theory has recently gained considerable attention in the field of image processing. Humans have different perceptions of an image at different distances, which refers to the scale effect in the HVS. The spatial scale technique is applied to the original image to form a series of smooth and simplified images. As the scale increases, details in the image become increasingly blurred. The concept of "scale space" is regarded as a supplement to the well-known image pyramid, which was first adopted by Kelly (1971) in image processing and was later extended to various forms. In the modern scale space expression, each layer is a blurred result of the former layer. Each layer is usually generated from the original image through the variance of the Gaussian convolution function. The variance of each layer represents the "scale parameter". A greater scale parameter indicates that more details on the image are discarded. However, these details might contain key element for human visual perception, therefore could be effectively used for object recognition, feature extraction, and so on.

According to Lindeberg's theory, scale space describes the original image in different levels presented by a scale parameter (Tony, 1996). Scale space could be achieved by convolving the original image using the Gaussian kernel function. It is an ideal mathematical model that simulates the human visual mechanism. With a series of reasonable assumptions, the Gaussian kernel function is a unique linear transform kernel in scale space (Tony, 1994).

The two-dimensional Gaussian function is defined as

$$G(x,y;\sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (1)$$

where σ is the scale parameter of the Gaussian function, and (x, y) indicates the coordinates of a certain point in the convolution template. The Gaussian scale space $L(\sigma)$ generated from an image can be defined as the convolution between the Gaussian kernel function and the image

$$L(\sigma) = G(\sigma) * I \quad (2)$$

where I is the original image, σ is the scale-space factor, $G(\sigma)$ is the Gaussian function, and $*$ is the convolving operation.

To establish the Gaussian scale space, the scale space is divided into segments called octaves, and the scale parameter of the bottom layer is assumed to be σ. All Gaussian convolution image layers of the scale parameter from σ to 2σ form an octave. Suppose that s layers are in a given octave, and the proportional relationship of the scales of adjacent layers is $k = 2^{1/(s-1)}$. Each image layer can then be generated by the convolution of the Gaussian kernel with the original image, or with the former image layer. In this paper, we select the latter, that is,

$$I_p = G(k^{p-1}\sigma) * I_{p-1}, \quad p = 1, 2, \ldots, s, \quad \text{and} \quad k^{s-1} = 2 \quad (3)$$

If the scale-increasing relationship between layers is k, then the absolute scale of the pth layer is $k^{p-1}\sigma$. Thus, if the scale parameter σ and the number of layers s are determined, an octave and all of its images can be acquired. The difference of the Gaussian (DoG) image is used to acquire the spatial detail information related to the scale parameter. Section 3 will further illustrate the importance of DoG.

3. Proposed quality assessment

The central principle of image fusion can be considered as injecting the spatial detail information of the PAN image into the MS image. Therefore, the quality of the fused image can be assessed in terms of the preservation of the spatial details and spectral information. In this paper, this preservation is measured using spatial and spectral quality indices.

3.1. Spatial quality index

According to the GSS technology theory in Section 2, as the scale increases, the image layer becomes blurred, and detail information is gradually lost. Thus, we can use different Gaussian image layers to restore the lost details. For example, the differences between two adjacent layers can be regarded as the detail information of these two layers. If all spatial details of the PAN image are injected into the MS image, the difference between adjacent layers of the fused image in the GSS would be the same as the difference between the corresponding adjacent layers of the PAN image. In fact, these two differences can never be exactly the same. Thus, by comparing the similarity between these two differences, the preservation of the spatial details (i.e., spatial quality) can be assessed.

The quality index $G_s(i)$ of the ith band of fused image can be expressed as follows:

$$G_s(i) = \left(\frac{1}{s} \sum_{k=0}^{s-1} |Q_s[F_i(\sigma_k) - F_i(\sigma_{k+1}), P(\sigma_k) - P(\sigma_{k+1})]|^p \right)^{\frac{1}{p}} \quad (4)$$

where F_i indicates the ith band of the fused image; σ_k and σ_{k+1} stand for the scale factor of the kth and $k+1$th layer, respectively; and s denotes the number of scales. $F_i(\sigma_k)$ and $P(\sigma_k)$ are the Gaussian image layers of the fusion and PAN images with a scale factor σ_k, respectively. $F_i(\sigma_k) - F_i(\sigma_{k+1})$ and $P(\sigma_k) - P(\sigma_{k+1})$ are the differences between the adjacent image layers of the ith band of fusion and PAN images, respectively, i.e., the spatial detail information between