Neurocomputing 177 (2016) 1–8

Contents lists available at ScienceDirect

Neurocomputing

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

Image fusion with saliency map and interest points

Fanjie Meng*, Baolong Guo, Miao Song, Xu Zhang

School of Aerospace Science and Technology, Xidian University, Xi'an 710071, China

ARTICLE INFO

Article history:
Received 11 June 2015
Received in revised form
19 October 2015
Accepted 23 October 2015
Communicated by: Jungong Han
Available online 18 November 2015

Keywords:
Image fusion
Saliency map
Interest points
Convex hull

ABSTRACT

A novel region-based fusion algorithm to combine infrared (IR) and visible light (ViS) images using a saliency map and interest points is presented. First, the IR image is applied to a saliency detection process to generate a saliency map. Second, the IR image is further analyzed to detect the interest points. Once the freely distributed points of interest have been removed, the salient interest points are extracted. Third, the convex hull of the salient interest points is calculated to obtain the salient region determined by the salient interest points. Then, the initial saliency map is refined by combining the convex hull of the salient interest points. This strategy makes the saliency result more robust and better able to accurately locate the object. Finally, different fusion rules for the object region and the background are employed. Experimental results show that the proposed algorithm has promising performance.

1. Introduction

Image fusion aims to synthesize information from multiple source images to achieve a more accurate, complete and reliable description of the same scene. The fusion of infrared (IR) and visible light (ViS) images has been widely applied in many fields such as remote sensing, night vision, video surveillance, military defense, and so on [1,2]. This is attributed to essential differences between IR and ViS sensors. ViS images has the ability to represent the background more abundantly and legibly than IR images, while the target is often hard to detect in ViS images (that have low visual contrast) but can be easily discovered in IR images.

A number of fusion algorithms have been proposed in the literature over the years, and they can be generally divided into two categories: pixel-based and region-based methods. Pixel-based image fusion combines images at the pixel level while region-based image fusion considers pixels within the same object as an entity. Instead of pixel-based image fusion combing the pixel values without reference to the object they belong to, it seems more meaningful to combine objects or regions rather than pixels since meaningful objects are always more readily identifiable than incoherent individual pixels [3]. However, region-based image fusion is facing some challenge especially due to image segmentation which based the fusion. Although many state-of-the-art salient region detection methods have been proposed in recent years, there are still some limitations with them. The most prominent limitation is that those salient region detection methods may highlight not only the object region but also some of the background region. Even if some of those methods could improve the effect of image segmentation to a certain extent, it would be at the expense of being more time-consuming.

Taking the above issue into consideration, this paper proposes a novel region-based fusion algorithm for IR and ViS images. The core idea is to integrate the saliency map and interest points for object region detection followed by different fusion rules. The contributions of our approach include three aspects:

- It puts forward a novel object region detection method by taking both the saliency map and interest points into account, which has rarely been done before. Synthesizing the two aspects of advantages the combination result can locate the object more robustly and accurately, and achieve the state-of-the-art performance.
- It presents a fast region-based fusion algorithm for IR and ViS images. Through the high-quality object region detection and the contrapuntal fusion rules, our fusion method is superior rendered in both subjective and objective assessments and provides a new perspective for image fusion.
- The proposed object region detection algorithm can be further used in tasks dependent saliency detection and subsequent applications such as automatic image annotation and content-based image retrieval.

The remainder of this paper is organized as follows. A brief review of the related work is given in Section 2. The proposed framework is introduced in Section 3. The object region detection method is presented in Section 4. The fusion rule is described in

* Corresponding author.
E-mail addresses: fjmeng@xidian.edu.cn (F. Meng), blguo@xidian.edu.cn (B. Guo), songmiao69@sina.com (M. Song), 18088328@qq.com (X. Zhang).

http://dx.doi.org/10.1016/j.neucom.2015.10.080
0925-2312/

Section 5. Experiments are performed in Section 6 and the conclusion is given in Section 7.

2. Related work

Image fusion has been extensively studied in the past decades, and the reader is referred to [4–6] for an exhaustive overview. In the literature, the existing techniques can be classified into two categories: pixel-based fusion and region-based fusion.

Pixel-based image fusion is performed in a pixel-by-pixel basis either in the spatial or transform domain. The spatial pixel-value fusion combines the pixel values of the input images in a linear or non-linear way [7]. The simplest form is a weighted averaging of the input images, where the weight for each source image can be computed using prior knowledge. This kind of method is of low computational complexity for it is only applied to a single scale, but it may reduce the contrast of some fine-grained details consequently. Transform fusion methods first convert each input image from normal image space into a more convenient representation by applying a multiscale transform (MST), then apply different rules to different scales and resolution, finally employ the inverse transform to construct the fused image. The most typical MST tools that used in current image fusion approaches include the Laplacian pyramid [8], the wavelet transform (DWT) [9], the curvelet transform (CVT) [10] and the non-subsampled contourlet transform (NSCT) [11,12]. In general, due to the perfect properties of the DWT such as multiresolution, spatial and frequency localization, and direction, the DWT-based methods are superior to the pyramid-based methods [13]. However, the DWT also has some limitations involving limited directions and non-optimal-sparse representation of an image. Accordingly, the DWT-based methods are easily introduced some artifacts into the fused image, which will hence decline the quality of the resultant image. Recently developed multiscale geometric analysis (MGA) tools, such as the CVT and the NSCT, not only with multiscale, localization and multi-direction, but also with properties of shift-invariance and the same size between each subband image and the original image, they are more popular for image fusion.

Region-based image fusion uses some segmentation algorithm to separate an original image into different regions, and then calculate various properties of these regions to determine which features from which images are included in the fused image. This sort of methods is usually robust to noise and misregistration. Furthermore, because pixel-based methods fail to take into account the relationship between points and points, the fused image might lose some feature information. On the contrary, region-based fusion can obtain the best fusion results by considering the nature of points in each region altogether. Therefore, region-based fusion has much advantages over the counterpart. Many region-based methods have been proposed in recent years. In [6], Lewis et al. generate a joint-segmentation map out of the input images. Then, the characteristics of each region are calculated in the wavelet domain during the fusion procedure. In [14], instead of using the unimodal segmentation, authors adopt joint segmentations to deal with sets of multimodal images. This sort of joint segmentation can improve the segmentation quality and the fused image quality. Guo et al. [15] present a region-based fusion algorithm for IR and ViS images. The IR image is first segmented according to the physical features of the target, and then the NSCT coefficients of target regions and background regions from source images are extracted and merged separately. Wan et al. [16] integrate multiscale image segmentation and a statistical fusion scheme. In the segmentation component, the local statistical characteristics of images have been employed in order to drive the initial texture segmentation process. In the fusion component, the salient information contained in the regions was modeled using bivariate alpha-stable distributions. In [17], Zaveri and Zaveri propose an accurate segmentation method for region based fusion by using graph based normalized cut algorithm. In addition, two different fusion rules—the mean max standard deviation and spatial frequency are introduced for different fusion images. In a recent publication, Han et al. [5] propose a saliency-aware image fusion algorithm which takes the saliency of the object into account. In order to generate a consistent saliency map from an IR image, the co-occurrence of hot spots and motion are input into a Markov Random Field (MRF) model. Moreover, Luo et al. [18] propose an adaptive multi-strategy fusion rule in different regions for region-based image fusion.

The above region-based fusion techniques extract object regions by means of a region partition algorithm, and then adopt different fusion rules separately for object regions and background regions. Detection of visually salient image regions has been an active research topic for a long time, and many state-of-the-art methods have emerged. A milestone work is presented by Itti et al. [19]. It develops a biologically plausible architecture that determines the center-surround contrast using a difference of Gaussians (DoG) across multiple scales. The final saliency map is derived by the linear summation of color, intensity, and orientation contrast. In [20], Ma and Zhang adopt the difference of windows (DoW) to calculate the color distribution distance between a location and its surrounding location within a window to measure the contrast. Han et al. [21] extract viewer's attention objects in a two-stage process. The first stage is to generate the saliency map by Itti's model, which encodes the attention value at every location in the image. In the second stage, only a few attention seeds are first selected according to the saliency map. Then, a Markov random field (MRF) model integrating the attention value and the low-level features is employed to sequentially grow the attention objects starting from those selected attention seeds. In [22], Harel et al. first form activation maps on certain feature channels, and then normalize them in a way which highlights conspicuity and admits combination with other maps. Hou and Zhang [23] analyze the log spectrum of each image and obtain the spectral residual and transform the spectral residual to spatial domain to obtain the saliency map, which suggests the positions of proto-objects. Achanta et al. [24] introduce a frequency-tuned approach to estimate the center-surround contrast using color and luminance features. Goferman et al. [25] propose the saliency detection based on four principles observed in the psychological literature: local low-level considerations, global considerations, visual organizational rules, and high-level factors. Shen et al. [26] represent an image as a low-rank matrix plus sparse noises in a learned feature space, where the low-rank matrix explains the non-salient regions (or background), and the sparse noises indicate the salient regions. Those methods are attempted to improve the effect of image segmentation. However, there are still some problems with them. One obvious problem is that they may highlight not only the object region but also some of the background region. Even if some of those methods could improve the quality of image segmentation, it would be at the expense of being more time-consuming.

3. The approach overview

The proposed approach architecture with its main functional units and data flows is shown in Fig. 1. First, saliency detection is applied to the IR image to generate a saliency map. Second, the IR image is further conducted to interest points detection and free points removal, followed by the convex hull of salient interest points computation. Third, the initial saliency map is refined by combining the convex hull of the salient interest points to extract