Signal Processing 102 (2014) 64–76

Contents lists available at ScienceDirect

Signal Processing

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

Multi-focus image fusion using image-partition-based focus detection

Xiaoli Zhang [a,b], Xiongfei Li [a,b], Zhaojun Liu [a,b,*], Yuncong Feng [a,b]

[a] Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education, Jilin University, Changchun 130012, China
[b] College of Computer Science and Technology, Jilin University, Changchun 130012, China

ARTICLE INFO

Article history:
Received 4 January 2014
Received in revised form
26 February 2014
Accepted 27 February 2014
Available online 12 March 2014

Keywords:
Multi-focus image fusion
Focus detection
Image partition
Distance transform

ABSTRACT

Focus detection based fusion algorithm is a vital alternative in multi-focus image fusion applications. In this kind of fusion algorithms, focus detection measure is a key factor. However, nearly all of them tend to make incorrect predictions in the smooth regions which are close to edges and textures, because these regions are affected by edges and textures and intensities become quite different if they are blurred. In this paper, we propose a new focus detection based multi-focus image fusion algorithm. First of all, the source images are partitioned into three parts: edges, textures, and smooth regions. Pixels in smooth regions are further classified into two catalogues according to their distances from edges or textures. Then, we formulate a new focus detection rule in which pixels in smooth parts are treated differently according to their classification. Finally, the fused image is achieved with the assistance of fusing map. The interests of algorithm lie in its ability of improving the accuracy of focus detection and eliminating blockiness in fused images. Experimental results have shown that the proposed fusion algorithm retains good ratings by Human Visual System (HVS) and objective measures compared to other multi-focus fusion algorithms.

1. Introduction

Multi-focus image fusion is a procedure in which some images from one scene with focuses on objects at different distances are integrated in such a way that all objects appear to be in focus in the final image [1]. The reason why we need to combine these images with focus on different locations is that existing imaging cameras usually have only a finite depth of field, which makes it possible for us to obtain an image in which all the objects are focused [2]. Until now, the technology of multi-focus image fusion has been proven valuable in surveillance and microscopic imaging [3].

Image fusion methodologies can be classified into two catalogues: single-scale based methods and multi-scale/multi-resolution based methods. The former ones combine information in spatial space according to some certain rules such as maximum/minimum selection, weighted addition, PCA and the like. In contrast, the later ones conduct multi-scale/multi-resolution image decompositions before combining information, and the fusion rules are designed based on frequency context. Generally, multi-scale/multi-resolution based methods include three stages [4]: (1) Decomposition, (2) Coefficients fusion, and (3) Reconstruction. Pyramid based image decompositions are the first one introduced into this field, and they include Filter Subtract Decimate Pyramid [5], Gradient Pyramid [6], Laplacian

* Corresponding author at: Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education, Jilin University, Changchun 130012, China.
E-mail addresses: xiaolizhang1987@gmail.com (X. Zhang), lxf@jlu.edu.cn (X. Li), zhaojun@jlu.edu.cn (Z. Liu), fengyc11@mails.jlu.edu.cn (Y. Feng).

http://dx.doi.org/10.1016/j.sigpro.2014.02.024
0165-1684

Pyramid [7], Ratio Pyramid [9], Morphological Pyramid [8] and Contrast Pyramid [10]. Discrete Wavelet Transform (DWT) is more advanced than pyramid decompositions in many ways. For instance, it can capture directional information which is vital in Human Visual system (HVS) [11]; in addition, fused images by DWT based schemes can achieve higher Peak-to-peak Signal-to-Noise Ratio [12]. Later, Shift-Invariant Wavelet Transform is adopted to improve temporal stability and consistency of fused images [13], Dual-tree complex wavelet transform (DTCWT) is proposed to capture more edge information [14]. More recently, some sophisticated multi-scale/multi-resolution image decomposition methods have been designed, such as Ridgelet transform, Shearlet Transform [15], Curvelet transform [16] and (nonsubsampled) Contourlet transform [17,18].

Huang and Jing noted that single-scale based methods are more suitable for multi-focus image fusions than multi-scale based methods [2], because the former ones can overcome the problem of shift-variance, and possess the merits of simple implementation and few occupation spaces. Different from other kinds of fusion (such as multi-modal and multi-spectral fusion), multi-focus fusion algorithms should select the in-focus regions and totally neglect out-of-focus regions. So the key issue in multi-focus image fusion is to detect the focuses of each source image correctly. The basic assumption of the multi-focus image fusion is that focused objects seems sharper than the unfocused objects, and the sharpness is linked to some easily computed measures. Until now, varieties of measures have been developed, such as Spatial Frequency (SF) [19,20], sum-modified-Laplacian (SML) operator [21], and Tenenbaum gradient (Tenengrad) [22].

Generally, focus detection results given by the existing measures can correlate well with human visual perception, namely the basic assumption of the multi-focus image fusion mentioned above is correct in most cases. However, if these measures are used to detect the smooth regions which are close to edges, they tend to give incorrect answers. In effect, for the smooth areas that are close to edge images, if out of focus, they are usually influenced by edge information, and consequently they will be given higher scores by focus detection measures than the ones which are not influenced by edge information. This phenomenon will force focus-detection-based image fusion algorithms to make incorrect decisions on which information to be selected. In [23], the authors found that smooth areas of an image together with the edge information have a great influence on our visual perception. Thus, an incorrect detection in the smooth regions can affect the visual perception of fused images. In order to tackle this problem, a more sophisticated focus detection algorithm is proposed. In this algorithm, source images are partitioned into three parts: edges, textures and smooth parts. Then, all the pixels in smooth regions will be checked whether they are influenced by edges or textures. And finally based on the judgments, focus detection rules are formulated in which pixels in smooth regions are treated differently.

The rest of paper is organized as follows: in Section 2, the problems that existing focus detection methods encounter are investigated; Section 3 details the multi-focus image fusion scheme based on a novel focus detection algorithm; in Section 4, the fusion scheme is tested on several groups of images; and the concluding remarks are presented in Section 5.

2. What's wrong with the existing focus detection measures?

As mentioned in Introduction part, the basic assumption of the multi-focus image fusion is that focused objects seem sharper than unfocused ones. So the key issue in the field is how to formulate information measures to detect which parts are in focus, and which parts not. However, the fusion schemes which follow the basic assumption cannot always get satisfactory fusion results, especially in smooth regions.

In this section, the underlying theories of existing focus detection measures are investigated. These focus measures include Standard Deviation (SD), Spatial Frequency (SF), Average Gradient (AG), Sum-Modified-Laplacian (SML) and Tenengrad measure.

Let $\mathbf{x}=\{x_{i,j}|i=1,2,\ldots,M;\ j=1,2,\ldots,N\}$ be an image, where $x_{i,j}$ is the intensity of the pixel at the intersection of the ith row and jth column in image $\mathbf{x}$, M is the number of rows, and N is the number of columns. Here, a new operation that is used later is first defined.

Definition 1. For two matrixes $\mathbf{X}$ and $\mathbf{Y}$, they share the same size of $m\times n$, the operation $\bullet$ is defined as

$$\mathbf{X}\bullet\mathbf{Y}=\begin{Bmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{m1} & \cdots & x_{mn} \end{Bmatrix}\bullet\begin{Bmatrix} y_{11} & \cdots & y_{1n} \\ \vdots & \ddots & \vdots \\ y_{m1} & \cdots & y_{mn} \end{Bmatrix} = \begin{Bmatrix} x_{11}y_{11} & \cdots & x_{1n}y_{1n} \\ \vdots & \ddots & \vdots \\ x_{m1}y_{m1} & \cdots & x_{mn}y_{mn} \end{Bmatrix} \tag{1}$$

These measures are listed in Table 1. From the table, it can be observed that the theories of the five measures are similar, they measure the active level of an image by examining the intensity differences among neighboring pixels. If the differences are large, these measures repute that the image contains high active level, and vice versa. The assumption is correct in most cases, because the regions with larger intensity differences seem to be sharper. From this point, it is proper that these measures are adopted to detect the focus of source images. However, we should not neglect such a case which is very prevalent in the practice. Assuming that there exists a region in which intensities are the same in one image; in another image, there also exists the same region, but the region is affected by other regions, which means that intensities in it become different more or less. There is no doubt that the former one may obtain a better visual perception, but the previously mentioned measures may favor the latter one. An illustrative example is shown in Fig. 1(a) and (b) are the two images with focus on different locations. The focus in (a) is on the right clock, while the focus in (b) is on the left clock. In the two images, three regions are marked by red rectangles, and we denote these regions from up to bottom as A-I, A-II, A-III in (a) and B-I, B-II, B-III in (b). The six regions share the same size of 11×11. (c)–(h) show the intensity distributions of A-I, B-I, A-II, B-II,