Information Fusion 31 (2016) 100–109

Contents lists available at ScienceDirect

Information Fusion

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

Infrared and visible image fusion via gradient transfer and total variation minimization

Jiayi Ma[a], Chen Chen[b], Chang Li[c], Jun Huang[a,*]

[a] Electronic Information School, Wuhan University, Wuhan 430072, China
[b] Department of Electrical and Computer Engineering, UIUC, Urbana, IL 61801, United States
[c] School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan 430074, China

ARTICLE INFO

Article history:
Received 8 June 2015
Revised 24 October 2015
Accepted 2 February 2016
Available online 9 February 2016

Keywords:
Image fusion
Infrared
Registration
Total variation
Gradient transfer

ABSTRACT

In image fusion, the most desirable information is obtained from multiple images of the same scene and merged to generate a composite image. This resulting new image is more appropriate for human visual perception and further image-processing tasks. Existing methods typically use the same representations and extract the similar characteristics for different source images during the fusion process. However, it may not be appropriate for infrared and visible images, as the thermal radiation in infrared images and the appearance in visible images are manifestations of two different phenomena. To keep the thermal radiation and appearance information simultaneously, in this paper we propose a novel fusion algorithm, named *Gradient Transfer Fusion* (GTF), based on gradient transfer and total variation (TV) minimization. We formulate the fusion problem as an ℓ^1-TV minimization problem, where the data fidelity term keeps the main intensity distribution in the infrared image, and the regularization term preserves the gradient variation in the visible image. We also generalize the formulation to fuse image pairs without pre-registration, which greatly enhances its applicability as high-precision registration is very challenging for multi-sensor data. The qualitative and quantitative comparisons with eight state-of-the-art methods on publicly available databases demonstrate the advantages of GTF, where our results look like sharpened infrared images with more appearance details.

1. Introduction

Multi-sensor data often provides complementary information about the region surveyed, and image fusion which aims to create new images from such data offering more complex and detailed scene representation has then emerged as a promising research strategy for scene analysis in the areas of pattern recognition, remote sensing, medical imaging and modern military [1–3]. Particularly, multi-sensor data such as thermal infrared (IR) and visible images has been used to enhance the performance in terms of human visual perception, object detection, and target recognition [4,5]. For example, visible sensors capture reflected lights with abundant appearance information, and it is better for establishing a discriminative model. In contrast, IR sensors capture principally thermal radiations emitted by objects, which are less affected by illumination variation or disguise and hence, it can overcome some of the obstacles to discover the target and work day and night. However, IR image typically has lower spatial resolution than visible image, where appearance features such as textures in a visible image often get lost in the corresponding IR image since textures seldom influence the heat emitted by an object. Therefore, it is beneficial for automatic target detection and unambiguous localization to fuse the thermal radiation and texture information into a single image, which is the main focus of this paper.

Image fusion is a process that can be conducted at varying levels, depending on information representations and applications. These levels are usually categorized into signal, feature, and symbol levels; however, despite their distinctions, they can be combined [6]. Pixel-level fusion, or fusion at the signal level, is the lowest-level type of fusion and involves the combination of raw source images into a single image [7]. By contrast, higher-level fusion involves combining information that usually occurs in the form of feature descriptors and probabilistic variables [8]. For example, source images are usually fused by feature-level algorithms according to the obtained feature properties, such as edges, shapes, textures, or regions. By contrast, symbol-level algorithms fuse a number of symbolic representations based on decision rules that

* Corresponding author. Tel.: +86 13871380842.
E-mail addresses: jyma2010@gmail.com (J. Ma), cchen156@illinois.edu (C. Chen), lichang@hust.edu.cn (C. Li), junhwong@whu.edu.cn (J. Huang).

http://dx.doi.org/10.1016/j.inffus.2016.02.001
1566-2535/

Fig. 1. Schematic illustration of image fusion. From left to right: the visible image, the infrared image, the fusion result of a recent MST-based method LP-SR [18], and the fusion result of our proposed GTF.

promote common interpretation and resolve differences. This study addresses the problem of pixel-level fusion, which is widely used in most image fusion applications, as the original information is contained in the input data and the algorithms are rather easy to implement and time efficient.

In pixel-image fusion, the first problem to be solved is to determine the most important information in the source images to transfer the obtained information into a fused image with the least change possible, especially distortion or loss. To address this problem, many methods have been proposed in the past decades, including methods based on pyramid [9–11], wavelets [12,13], curvelet transform [14], multi-resolution singular value decomposition [15], guided filtering [16], multi-focus [17], sparse representation [18], etc. Averaging the source images pixel by pixel is the simplest strategy. However, a number of undesirable effects, such as reduced contrast, arise from this direct method. In order to solve this problem, multi-scale transform (MST) based methods have been proposed which involve three basic steps: i) the source images are first decomposed into multi-scale representations with low and high frequency information; ii) the multi-scale representations are then fused according to some fusion rules; iii) the inverse transform of composite multi-scale coefficients is finally used to compose the fused image [19]. The MST-based methods are able to provide much better performance as they are consistent with human visual system and real-world objects usually consist of structures at different scales [6]. Examples of these methods include Laplacian pyramid [9], discrete wavelet transform [20], non subsampled contour let transform [21], etc. The MST-based methods have achieved great success in many situations; however, they use the same representations for different source images and try to preserve the same salient features such as edges and lines in the source images. For the problem of infrared and visible image fusion, the thermal radiation information in an infrared image is characterized by the pixel intensities, and the target typically has larger intensities compared to the background and hence can be easily detected; while the texture information in a visible image is mainly characterized by the gradients, and the gradients with large magnitude (e.g. edges) provide detail information for the scene. Therefore, it is not appropriate to use the same representations for these two types of images during the fusion process. Instead, to preserve the important information as much as possible, the fused image is desirable to keep the main intensity distribution in the infrared image and the gradient variation in the visible image. To this end, in this paper we proposed a new algorithm, named Gradient Transfer Fusion (GTF), based on gradient transfer and total variation (TV) minimization for infrared and visible image fusion.

More precisely, we formulate the fusion as an optimization problem, where the objective function consists of a data fidelity term and a regularization item. The data fidelity term constrains that the fused image should have the similar pixel intensities with the given infrared image, while the regularization term ensures that the gradient distribution in the visible image can be transferred into the fused image. The ℓ^1 norm is employed to encourage the sparseness of the gradients, and the optimization problem can then be solved via existing ℓ^1-TV minimization techniques [22,23]. To illustrate the main ideas of our method, we show a simple example in Fig. 1. The left two images are the visible and infrared images to be fused, where the visible image contains detailed background and the infrared image highlights the target, i.e. the building. The third image is the fusion result by using a recent MST-based method [18]. We see that the details of the background are kept and the target also becomes brighter. However, it fails to highlight the target as the background is also bright after fusion. This demonstrates the importance of keeping the thermal radiation distribution in the infrared image, and the advantage will be magnified when the scene contains false targets (e.g., decoys) which often occurs in military applications. The rightmost image in Fig. 1 is the fusion result of our GTF algorithm. Clearly, our result preserves the thermal radiation distribution in the infrared image and hence, the target can be easily detected. Meanwhile, the details of the background (i.e. the trees) in the visible image are also kept in our result.

The second problem to be solved in pixel-level image fusion is the accurate alignment of the source images on a pixel-by-pixel basis. This alignment process guarantees that the original information from each source corresponds to the same real-world physical structures. Registration involves various challenges, especially when multi-sensor data, such as IR and visible images, are used. Thus, researchers and scholars have explored the issue of image registration as a process independent from image fusion and have proposed a good number of considerably successful techniques of image registration. In the literature, there are in general two types of approaches for image registration: area-based methods [24,25] and feature-based methods [26–28], as discussed in recent survey papers [29]. The former finds the matching information by using the original pixel intensities in the overlapped region of two images with a specified similarity metric, while the latter seeks correspondence between local features under descriptor similarity and/or spatial geometrical constraints. The area-based methods are preferable in case of few prominent details where the distinctive information is provided by pixel intensities rather than by local shapes and structures, but they suffer from heavy computational complexities, image distortions, as well as illumination changes. By contrast, feature-based methods are more robust to geometric and photometric distortions. As such, the registration problem reduces to determining the correct correspondence and finding the underlying spatial transformation between two sets of extracted features. However, it is difficult to establish accurate alignments for infrared and visible image pairs, since locating reliable features for such images is very challenging and the registration accuracy directly relies on the extracted features. In this paper, we further generalize the proposed GTF so that it is able to fuse unregistered image pairs.

Specifically, we address the registration problem by introducing an additional variable, i.e. the spatial transformation, to the GTF formulation and imposing it on the image with larger field of view. The fused image and the spatial transformation are alternatively optimized until convergence under the assumption that the