Information Fusion 27 (2016) 85–94

Contents lists available at ScienceDirect

Information Fusion

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

Jointly registering and fusing images from multiple sensors

Yinghao Li [a], Zhongshi He [a,*], Hao Zhu [b], Weiwei Zhang [c], Yuhao Wu [a]

[a] College of Computer Science, Chongqing University, No. 174 Shazhengjie, Shapingba, Chongqing 400044, PR China
[b] College of Automation, Chongqing University of Posts and Telecommunications, Chongqing 400065, PR China
[c] College Of Computer and Communication Engineering, Zhengzhou University of Light Industry, Zhengzhou, Henan 450000, PR China

ARTICLE INFO

Article history:
Received 5 July 2014
Received in revised form 18 May 2015
Accepted 19 May 2015
Available online 28 May 2015

Keywords:
Mixture models
Multi-image
Multi-sensor
Image registration
Image fusion

ABSTRACT

In this paper, a novel approach is proposed for jointly registering and fusing a multisensor ensemble of images. Based on the idea that both groupwise registration and fusion can be treated as estimation problems, the proposed approach simultaneously models the mapping from the fused image to the source images and the joint intensity of all images with motion parameters at first, and then combines these models into a maximum likelihood function. The relevant parameters are determined through employing an expectation maximization algorithm. To evaluate the performance of the proposed approach, some representative image registration and fusion approaches are compared on different multimodal image datasets. The criterion, which is defined as the average pixel displacement from its true registered position, is used to compare the performances of registration approaches. As for evaluating the fusion performance, three fusion quality measures which are metric $Q^{ab/f}$, mutual information and average gradient are employed. The experimental results show that the proposed approach has improved performance compared to conventional approaches.

1. Introduction

Image fusion is an important technique for various image-based applications such as computer vision, remote sensing and medical diagnosis. It aims to combine different source images of the same scene into a single fused image, which is supposed to contain as much information from the source images as possible. Since the source images might be collected at different times, from different perspectives or by different imaging sensors, an image registration algorithm is usually applied to transform these images into a common coordinate system or space. That is to say, image registration is critical for the success of image fusion.

Through literature-review, there have been many approaches dealing with image registration problem, which can be broadly classified into feature-based and intensity-based approaches. Feature-based approaches first select a set of corresponding features, which can be significant regions, contours or control points (such as landmarks and line intersections), etc., from both reference image and distorted image. Then these two sets of features are matched through estimating the optimal transformations [1–4]. The performance of feature-based approaches heavily relies upon the appropriate choice of features and accurate estimation of feature correspondences. Rather than selecting features for image registration, intensity-based approaches define similarity measures directly based on the joint intensity distribution of two images. The registration problem is thus considered as an optimization process to minimize or maximize the similarity measures. Correlation coefficient (CC) [5,6], mutual information (MI) [7], and maximum likelihood (ML) [8,9] are common used similarity measures. Although great deals of researches have been devoted to registering images, most of them are only suitable for solving pairwise registration problems.

To address this deficiency, several researches begin to focus on registration problem with more than two images. Some approaches select an image as a common reference or a template, and then register the rest of images to it in a pairwise manner [10,11]. However, in real practice, this type of approaches will suffer from high computational complexity and lead to bias in aligning all the other images to the priori chosen template. To circumvent these defects, some groupwise registration approaches use a global criterion to measure the joint information from the entire group of images. They simultaneously estimate the transformation of each image in the group by optimizing the global criterion. By this way, these approaches eliminate the requirement of choosing a priori template and generate a consistent solution. In addition, more information which comes from different source

* Corresponding author. Tel.: +86 23 65112390.
E-mail addresses: lyhcqu@126.com (Y. Li), zshe@cqu.edu.cn (Z. He), haozhu1982@gmail.com (H. Zhu), anqikeli@163.com (W. Zhang), ggglowu@gmail.com (Y. Wu).

http://dx.doi.org/10.1016/j.inffus.2015.05.007
1566-2535/

images could be used simultaneously to produce better registration results. The first groupwise registration approach was proposed by Woods et al. [12]. It constructed the global cost function by adding sums of squared intensity differences (SSD) between all possible image pairs, and then minimized this cost function to generate the transformation of each image. Later, Zöllei et al. [13] adopted congealing framework [14], in which the cost function is the total voxel-wise entropy of the input image volumes, for groupwise registration of magnetic resonance (MR) images. Wu et al. [15] used attribute vector, which could capture the geometric information at different scales, to improve the correspondence detection in groupwise registration. Although these approaches register a group of images simultaneously, they only focus on registering monotone images. If the distorted images come from different modalities, i.e. multimodal images, these approaches may fail to estimate the transformations of these images accurately.

To solve the multimodal images registration problem effectively, some approaches are proposed based on the idea that the co-occurring intensities form clusters, which represent the most probable intensity correspondences between the images, in the space of joint intensity. Studholme and Cardenas [16] used a generic statistical method to define a function of joint density and registered multimodal images by approximating the function. Their global measure is a sum of total self-information with geometric constraints on the registration solution. Recently, Orchard and Mann [17] presented a clustering method for multimodal groupwise registration. Their approach modeled the distribution of points in the space of joint intensity based on a Gaussian mixture model (GMM), and then estimated the model parameters by using an expectation maximization (EM) algorithm [18]. The motion parameters were also estimated by using an iterative Newton-type method. These approaches register a group of multimodal images, but the distorted images come from different sensors or viewpoints and have different characters of the scene. If a fused image, which contains all characteristics of distorted images, can join in the registration of these images, the registration accuracy could be improved.

A large amount of image fusion approaches have been proposed in literature. These approaches can be classified into three categories: decision-based, feature-based and pixel-based. Decision-based approaches extract information from each source image and then make decisions for them. Finally, they combine those decisions to generate the final decision [19]. Feature-based approaches extract features from each source image and then perform fusion directly based on those features [20,21]. As for pixel-based approaches, they fuse source images by combining multiple source image pixels into a single fused image pixel. There are mainly two types of approaches in pixel-level fusion. The first one converts the source images into a consistent transform domain and then performs fusion by combining their transform coefficients [22]. Some popular transform algorithms are pyramid transform [23] and wavelet transform [24]. Rather than applying transform algorithms before image fusion, the second one performs fusion directly on the gray values of source images [25–27].

Although many approaches have been proposed for image fusion, they always assume that image registration approaches have been applied already and the source images are perfectly co-registered. However, in most cases, completely accurate registration cannot be achieved in advance and the registration errors will have a bad influence to the subsequent fusion results. Therefore, the performance of image fusion is decided by both registration and fusion. Combining these two parts provides a new perspective for solving image fusion problems. Chen et al. [28] proposed an approach for joint image registration and fusion. This approach treated both image registration and image fusion as the problem of parameter estimation and combined them into a single ML formulation. However, this approach only focuses on pairwise registration and fusion. It is not suitable for jointly registering and fusing more than two images. Furthermore, the joint intensity of source images and fused image, which might contain more information, is not modeled in this approach.

In this paper, we present an efficient approach for multisensor groupwise images registration and fusion. The proposed approach models the mapping from fused image to source images by a linear transform with Gaussian mixture noise. Meanwhile, it models the joint intensity of source images and fused image by using a GMM. By this way, the intensity distribution of fused image could be used in groupwise registration. The groupwise registration and fusion problems are solved simultaneously by combining these models into a ML formulation and estimating relevant parameters. The EM algorithm is used to find the ML estimate of these parameters.

The rest of the paper is organized as follows. The problem formulation of joint groupwise image registration and fusion is presented in Section 2. The parameter estimation method is given in Section 3. The experiment settings which include experimental data and evaluation method are introduced in Section 4. The experimental results are reported and analyzed in Section 5. Conclusions and further research directions are drawn in the last section.

2. Problem formulation

In this section, we present the ML formulation of joint groupwise images registration and fusion. Assume that D source images will be simultaneously registered and fused into one image. Therefore, each pixel x is associated with D intensity values or transforming coefficients. A vector of intensities I_x is used to represent each pixel and each fused image pixel is denoted as F_x. I_x and F_x are combined into a $D+1$ dimensional vector L_x as follows in order to model the joint intensity distribution of source images and fused image conveniently.

$$L_x = [I_x; F_x] \tag{1}$$

Groupwise registration is to give these images a set of motion parameters which refer to the transforms applied to them. If each image of the group has M motion parameters, the total number of motion parameters is $MD + M$ (the fused image also has M motion parameters). We use θ to represent this parameter set and the joint intensity vector with motion parameters could be written as

$$L_x^\theta = \left[I_x^\theta; F_x^\theta\right] \tag{2}$$

Given the source images and fused image, the relationship between I_x^θ and F_x^θ is modeled by using a formation model of sensory image [26]. For every pixel of the source images, the mapping from F_x^θ to I_x^θ is modeled as

$$I_x^\theta = \beta_x F_x^\theta + \alpha_x + w_x \tag{3}$$

where β_x is a vector of sensor selectivity factors, α_x is a vector of sensor offsets, and w_x is a random distortion vector which is modeled by a GMM. In this model, the constraint of $\beta_x(d,1)$ ($\beta_x(d,1)$ denotes the number in the dth row and the first column of vector β_x) acknowledges that sensor d may be able to "see" certain objects ($\beta_x(d,1) = 1$), may fail to "see" other objects ($\beta_x(d,1) = 0$), or may "see" certain objects with a polarity-reversed representation ($\beta_x(d,1) = -1$) [28]. A vector A_x^θ is used to represent the relationship between I_x^θ and F_x^θ, and it is

$$A_x^\theta = I_x^\theta - \beta_x F_x^\theta - \alpha_x = w_x \tag{4}$$

After given registration and fusion models, the ML formulation which combines these two models is presented. Let us represent the unknown parameters as $\boldsymbol{\rho}$. For our approach, the complete data