Information Fusion 22 (2015) 119–126

Contents lists available at ScienceDirect

Information Fusion

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

Focused pooling for image fusion evaluation

Vladimir Petrović [b,*], Vladimir Dimitrijević [a]

[a] Faculty of Electrical Engineering, University of Belgrade, 11000 Belgrade, Serbia
[b] Imaging Science and Biomedical Engineering, University of Manchester, Manchester M13 9PT, UK

ARTICLE INFO

Article history:
Received 21 February 2013
Received in revised form 7 April 2014
Accepted 15 May 2014
Available online 3 June 2014

Keywords:
Image fusion
Performance evaluation
Pooling

ABSTRACT

This paper presents the results of an investigation into optimal spatial pooling of localised quality scores for use in objective evaluation of multisensor image fusion. We propose and evaluate a two stage focused pooling method with a localised aggregation of pixel-level performance estimates into regional fusion performance scores as the first step followed by a global pooling of regional scores into a global fusion performance score. We investigate a selection of linear and non-linear global pooling methods and show that quality driven methods which take into account regional fusion performance levels exhibit optimal performance. The proposed pooling algorithm is general and applicable to any fusion performance and quality metric based on structural preservation estimates, local differences between input and fused images. Specifically, we evaluate the proposed method in conjunction with three well-known structural preservation fusion metrics against their baseline pooling methods. We show, through correlation with an extensive subjectively annotated dataset of fused images, that regional aggregation of local performance scores over 3–6° of visual angle with selection of the worst performing region as the global score can improve performance for all the structural fusion metrics tested.

1. Introduction

Multisensor camera arrays image their environment using two or more different sensors to ensure a wider spectral coverage and reliable imaging even in adverse environmental conditions. The price of additional robustness is a considerable increase in the amount of image data that needs to be presented or processed simultaneously. Image fusion deals with this data overload by combining images of the same scene acquired using different sensors into a single, composite fused image [1] (see Fig. 1). Consequently, image fusion has been an active area of research in fields such as medical imaging, avionics and night vision with a plethora of methods available from simple averaging of signals to complex multiresolution algorithms with advanced feature analysis and selection [1,2].

Universally, the goal of these algorithms is to preserve the content of their multisensor inputs while performing significant data reduction. While some information loss is inevitable, crucially quality of fused image quality can be further compromised by specific fusion induced degradations such as ringing artifacts which constitute false information introduced into the fused image.

As a result of this inherent imperfect operation, reliable objective image fusion performance metrics became the object of research attention. Today a number of such metrics exist [3–15], described in more detail in Section 2, broadly based on comparing localised structural similarity [3–11] or global image statistics [12–15] between the input and fused images. The former group, of structural similarity metrics have in recent studies been shown to perform robustly in a wide variety of conditions and outperform statistical and information theoretic models both in the context of image fusion [16–19] and general image and video quality [20].

This evaluation approach, illustrated in Fig. 1, produces localised estimates of structural similarity/preservation between input and fused images. In the last stage, referred to as *spatial pooling* [21], these are combined into a single, global performance score. While most conventional pooling methods involve either some form of weighted summation [3,4] or simply taking the mean of all local estimates, appropriate selection of the pooling model, as we will show, can improve metric accuracy expressed as the ability to predict ground truth subjective ratings of different fused images.

We propose a general, two stage pooling algorithm, applicable to all structural similarity fusion measures rather than a specific metric, exploring the idea of quality driven pooling where image regions are weighted based on their local quality. The results of our investigation show that departing from conventional pooling strategies, analysed in more detail in Section 2, that treat all areas

* Corresponding author. Tel.: +44 1612755140.
E-mail addresses: v.petrovic@manchester.ac.uk (V. Petrović), vdimitrijevic@visaris.com (V. Dimitrijević).

http://dx.doi.org/10.1016/j.inffus.2014.05.002
1566-2535/

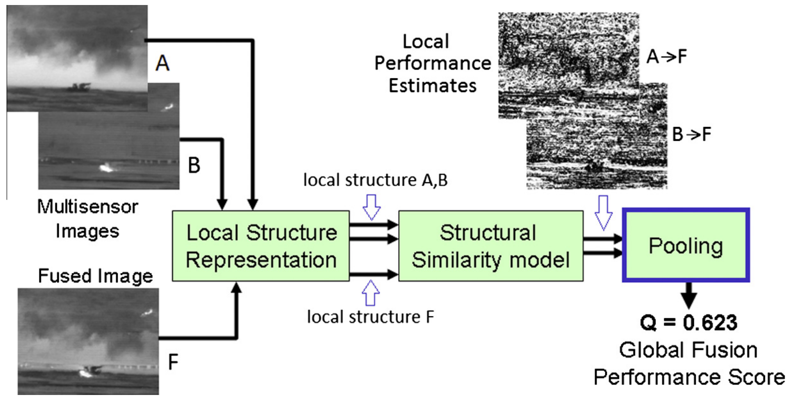

Fig. 1. Localised image fusion performance evaluation approach.

of the signal equally, a focused pooling approach where local estimates are pooled regionally and certain regions of the scene are ignored completely actually produce improved objective metric accuracy. Specifically, we evaluate the proposed method in conjunction with three well-known structural preservation fusion metrics against their baseline pooling methods and show that quality driven focused pooling can improve their accuracy.

In Section 2, we provide more details on the current field of fusion performance metrics with a particular focus on state-of-the-art structural preservation metrics and pooling strategies used with them. We outline and analyse the proposed focused pooling approach in Section 3 and present its results on an extensive calibrated dataset of fused imagery in Section 4. We conclude in Section 5.

2. Objective fusion evaluation

Image fusion performance evaluation is an active field of research [1–19] with many different metrics available to compare different fusion alternatives.

Statistical similarity metrics compare image statistics between the inputs and fused images. Mutual information (MI) and variants such as normalised MI (NMI), long used for multisensor processing, measure statistical dependence between signals and have been used as measures of image fusion performance [1,12]. However, questions remain over their accuracy and reliability [16–18,23]. Cvejic et al. [13] meanwhile demonstrated that their accuracy could be improved by using the Tsallis entropy formulation. Several statistical measures to estimate hyper-spectral fusion including modified versions of MI to evaluate fusion symmetry were proposed in [9].

Visual Information Fidelity (VIF), a wavelet-based mutual information measure was used to measure similarity between segmented regions of input and fused images in a metric proposed in [15]. Segmentation is also used in [10] in a metric inspired by the human vision system (HVS) where an explicit model of the contrast sensitivity function is used to process the input and fused images prior to evaluating the mean square error (MSE) over each segmented region. More recently, VIF has been used as a basis of a four stage fusion metric in [11] in a complex evaluation customised for the image fusion context.

Structural preservation metrics meanwhile measure localised differences between input and fused images. They usually proceed in three steps, see Fig. 1: (i) extraction of local structure descriptors (representation) across the scene, (ii) comparison of spatially corresponding local structures between input and fused images, yielding local quality/performance estimates, and (iii) spatial pooling of local estimates into a global measure of fusion performance.

Gradient preservation $Q^{AB/F}$ metric [3] operates on the principle that fusion algorithms that transfer input gradient information into the fused image more accurately perform better. Specifically, given input images A and B and fused image F, Sobel edge operator is used to find local gradient strength $\mathbf{g}$ and orientation $\boldsymbol{\alpha} \in [0,\pi]$ at each pixel in A, B and F (local structure representation in Fig. 1). Relative change in gradient magnitude and orientation between each input and the fused image, are then defined as:

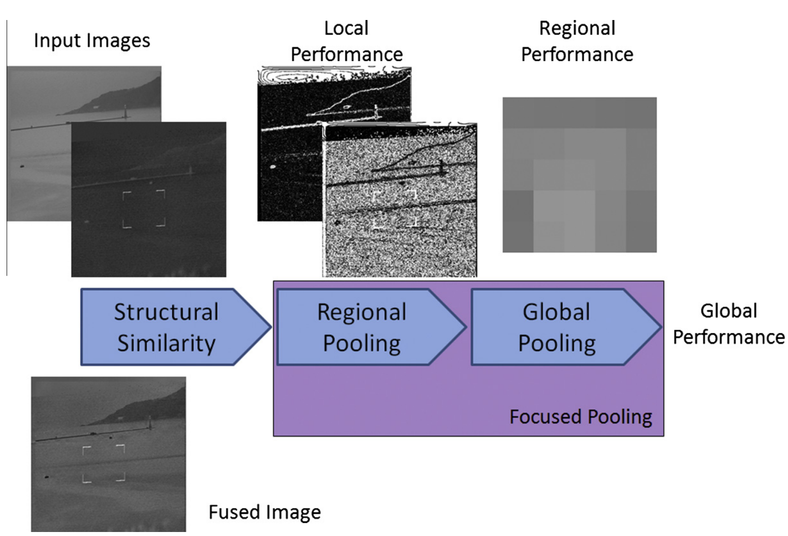

Fig. 2. Pooling steps in structural similarity fusion evaluation.