A Region-to-Pixel Based Multi-sensor Image Fusion

Sourav Pramanik\textsuperscript{a,*}, Swagatika Prusty\textsuperscript{a}, Debotosh Bhattacharjee\textsuperscript{b}, Piyush Kanti Bhunre\textsuperscript{a}

\textsuperscript{a}National Institute of Science and Technology, Computer Science and Engineering Department, Berhampur, India
\textsuperscript{b}Jadavpur University, Computer Science and Engineering Department, Kolkata, India

Abstract

A region based multi-sensor image fusion approach is proposed in this paper. At the initial stage of our algorithm, noise is suppressed from the input images by applying a $3 \times 3$ filter mask. In the next phase, regions are segmented from the input images by computing similarity map image followed by marker based watershed algorithm. Thereafter, regions are fused by computing the relative importance of a pixel in the region. Here, the relative importance of a pixel in the region is calculated as the second central moment of that pixel in the neighborhood with respect to the asymmetry or skewness of the whole region. After that a decision map is implemented based on the relative importance of a pixel in the region for fusion of the two correspondence regions. Finally, all the fused regions are combined to produce a final fused image. To check the robustness of our algorithm, we have tested it on 120 multi-sensor image pairs collected from Manchester University UK database and compared with some state-of-the-art region based fusion techniques. The experimental result shows the superiority of our proposed method in terms of visual and objective perception evaluation indexes.

© 2013 The Authors. Published by Elsevier Ltd. Open access under CC BY-NC-ND license. Selection and peer-review under responsibility of the University of Kalyani, Department of Computer Science & Engineering

Keywords: regional fusion; second central moment; third central moment; weight calculation; multi-sensor image;

1. Introduction

Let us have N two dimensional source images $I_1, I_2, \ldots, I_N$ of equal size $K \times L$ depicting the same true scene $S$ but exhibit different type of distortions. The images that we have used in this work have been captured in the multi sensor environment and they are registered to each other. The main goal of image fusion is to combine source images and produce a fused image $F$ which describes an improved perceptually enhance image over any of the source image $I_N$. For the advancement in technology and detailed analysis of the scene, it is not always possible to get all the physical and geometrical information from a single image. Therefore, it is necessary to procure information from different images. The process of combining images to produce a composite improved image is called image fusion. It has wide area of applications, such as aerial and satellite imaging, robotic vision, medical imaging etc. From last few years, image fusion has received lots of attention for image analysis, intelligent surveillance system, landing guidance, concealed weapon detection, computer vision, etc.

\textsuperscript{*} Corresponding author

\textit{E-mail address:} sourab@nist.edu (S.Pramanik)
In the past two decades, many techniques for generic image fusion in the spatial and in the frequency domain have been developed. In the frequency domain, all the source images are first transformed in the new domain, called frequency domain, then fused them and the result is converted back to the original image domain by an inverse transform. The most popular frequency domain methods are Laplacian pyramid [1], morphological pyramid [2], discrete wavelet transform [3], dual-tree wavelet transform [4], curvelet transform [5], the nonsubsampled contourlet transform [6], are applied to image fusion. In all the methods mentioned above the fusion coefficients are calculated with either pixel based or region based fusion rules. The problem of these methods is that during inverse transform process some information may loss. As in the spatial domain method, fused image is obtained by averaging or selection of arbitrary pixels from source images. But in the both cases the undesired affect such as blurring, pixel discontinuity take place. In [7], argued that the local structural characteristics of objects in an image cannot be completely expressed by arbitrary pixels. Hence, instead of arbitrary pixel based fusion, the region based fusion methods with actual features are more effective. Due to the information loss in the inverse transform process in the transform domain, some region segmentation and fusion based methods have been developed in the spatial domain. In [8], average image is first computed from the source images then traditional graph-based segmentation method is used to segment the image and segmented regions are fused via spatial frequency features. In [9], fused images by segmenting the target region in the IR image and replace the corresponding position in the visual image. In [10], image segmentation and fusion are performed by computing similarity map image.

In this paper, we have proposed a novel region based fusion algorithm. In our method, for region wise segmentation of the source images, we have first computed a correlation image or similarity image via multi-scale structural similarity index map (MSSSIM), then marker based watershed algorithm is applied on similarity image to obtain segmentation map. Thereafter, we have calculated weight for each pixel in the region which will assist the fusion process. The weight for a pixel in the region is calculated as the second central moment of that pixel in the neighborhood with respect to the asymmetry or skewness of the whole region. In the next phase, to handle discontinuity in the fused region, we have taken majority vote over some pixel in a given block. Finally, all the fused regions are combined in order to get final fused image. To verify our proposed algorithm, we have tested it on 120 pair of multisensory images collected from Manchester University UK database. Our proposed algorithm performs better compared with existing algorithm in terms of objective evaluation, contrast measure and overall information measure.

The rest of the paper is organized as follows: section II describes the overall system design, section III describes the detail of the experiments conducted along with results and section IV concludes the paper.

2. Proposed Method

In this section, we have presented our details algorithm for region based multisensory image fusion. Let us have \( N \) two dimensional source images \( I_1, I_2, \ldots, I_N \) of equal size \( K \times L \) depicting the same true scene \( S \) but exhibit different type of distortions. The main aim is to fuse those images in some way so that the fused image is an enhanced image over any of the source images. In the present work, we have taken two source images as input and produce a single fused image. This section is divided into five parts: problem formulation, preprocessing, region segmentation, weight calculation and fusion.

2.1. Problem Formulation

Let us have a set \( I \) of \( N \) source images \( I_1, I_2, \ldots, I_N \) of equal size \( K \times L \), describing the same true scene \( S \) but different distortion. Let \( T_\Theta : I \rightarrow I \) be a transformation, for a given image \( I_i \in I \), \( T(I_i) \) is also an image. Basically the transformation \( T \) is defined as the preprocessing task on the input images before fusion. For a predefined goal, \( T_\Theta \) denote the desired transformation parameterized by \( \Theta \). Let \( R_\phi : I \rightarrow (I_{R_1}, I_{R_2}, I_{R_3}, \ldots, I_{R_n}) \) be a transformation, for a given image \( I_i \in I \). In this transformation, input image \( I_i \) is segmented into \( n \) regions and \( R_\phi \) denote the desired transformation parameterized by \( \varphi \). Each segmented regions \( I_{R_i} \) are then transformed into row vector \( V_{R_i} \), for \( i = 1, 2, \ldots, n \). Let \( W_\phi : V_{R_i} \rightarrow \overline{V_{R_i}} \) be a transformation, for a given row vector \( V_{R_i} \). \( \overline{V_{R_i}} \) is the weighted region vector for the given input region vector which will assist the fusion process and \( W_\phi \) denote the desired transformation parameterized by \( \phi \). If \( V_{R_i} \) contains \( m \) elements then \( \overline{V_{R_i}} \) also contains \( m \) elements. The aim of our proposed algorithm is to fused source images region wised based on the weighted vector.
The images that we have used in this work have been captured in the multi-sensor environment. So it is desired that images contain some short of noise. In this section, we have tried to suppress noise to the maximum extend from source images before fusion. To remove noise, we have used a 3 × 3 filter mask which is shown in fig.1 [11]. In this filter mask, all the coefficients are considered heuristically but it is experimentally proved that this mask gives better result in comparison with mean filter or median filter. Fig.2 shows (a) and (b) are the input images, (c) and (d) are the corresponding filtered images using fig.1 mask.

2.3. Region Segmentation

The images used in this work are usually describing same true scene but bearing different distortions and distortion also varies from one region to another. So it is not possible to get same corresponding segmented regions from two source images using a traditional image segmentation algorithm. In [12], use a joint gradient image of two input images as a segmentation map. But due to the different distortions in the multisensory images, it is quite obvious that joint gradient image contains duplicate gradient information. Hence, it can mislead the segmentation process. In [13], are done segmentation of the source images using graph based segmentation algorithm. Then those segmented graphs are merged to get segmentation map. In multi-sensor image, distortion varies region wise and two images have different distortions. Thus, merging of two different segmented graphs, it can produce over segmentation.

Here, we use a similarity or correlation image of two input images as a segmentation map. The similarity image is computed via multi-scale structural similarity index (MS-SSIM) [14]. Basically multi-sensor images are captured in different viewing condition. Therefore, multi-scale method is a suitable way to compute similarity details at different scale. The overall MS-SSIM is evaluated by combining luminance, contrast and structural comparison measure at different scale of two image signals A and B (1). This is a full resolution similarity map image. In fig.3a and 3b are two input images and (c) shows their similarity image.

\[
MS-SSIM(A, B) = [l_M(A, B)]^{\alpha_M} \prod_{i=1}^{M} [c_i(A, B)]^{\beta_i} [s_i(A, B)]^{\gamma_i} \tag{1}
\]

Where, \(\alpha_M, \beta_i, \gamma_i\) are used to adjust the relative importance of different components. \(M\) is the highest scale. \(l(A, B), c(A, B), s(A, B)\) are luminance, contrast and structural similarity measure. In this work, we have considered \(\alpha_M = \beta_i = \gamma_i = 1\) and \(M = 4\). The luminance \(l(A, B)\), contrast \(c(A, B)\) and structural \(s(A, B)\) are computed as follows:

\[
l(A, B) = \frac{2\mu_A\mu_B + C_1}{\mu_A^2 + \mu_B^2 + C_1} \tag{2}
\]
\[ c(A, B) = \frac{2\sigma_A \sigma_B + C_2}{\sigma_A^2 + \sigma_B^2 + C_2} \]  
\[ s(A, B) = \frac{\sigma_{AB} + C_3}{\sigma_A \sigma_B + C_3} \]

where, \( C_1, C_2, C_3 \) are very small positive constant and computed by eq(5)

\[ C_1 = (K_1 L)^2, \quad C_2 = (K_2 L)^2, \quad C_3 = \frac{C_2}{2} \]

\( L \) is the maximum range of the pixel value in the input image and \( K_1, K_2 \) are very small scalar constant and \( K_1, K_2 < 1 \). Here we consider \( K_1 = 0.002 \) and \( K_2 = 0.003 \). \([\mu_A, \mu_B], [\sigma_A^2, \sigma_B^2], [\sigma_A, \sigma_B] \) and \( \sigma_{AB} \) are the mean, variance, standard deviation and covariance of \( A \) and \( B \).

In order to get full resolution similarity image, all the comparison in eq(1) are computed within \( 9 \times 9 \) Gaussian weighting window at different scale, whose center moves pixel by pixel over the entire image. The sum of weights in the Gaussian window is 1. The statistical measure mean, variance, standard deviation and covariance are computed as,

\[ \mu_A = \sum_{i=1}^{P} w_i A_i \]  
\[ \sigma_A^2 = \sum_{i=1}^{P} w_i (A_i - \mu_A)^2 \]  
\[ \sigma_A = \sqrt{\sum_{i=1}^{P} w_i (A_i - \mu_A)^2} \]  
\[ \sigma_{AB} = \sum_{i=1}^{P} w_i (A_i - \mu_A)(B_i - \mu_B) \]

where, \( P \) is the number of elements within specified window and \( w_i \) is the weight of Gaussian window. \( \mu_B, \sigma_B^2, \sigma_B \) are also computed same way as (6), (7), (8).

In this work, we have used the multi-scale structural similarity image to obtain segmentation map via marker based Watershed algorithm. The Watershed algorithm finds catchment basins and watershed ridge lines in an image by treating it as a surface where light pixels are high and dark pixels are low [15]. But watershed algorithm suffers from over segmentation. Hence, the image segmentation using the watershed algorithm performs better if we mark the foreground objects and background objects. In fig.3, (a) and (b) are the input images, (f) is the corresponding segmentation map using our method. The steps for segmentation of similarity image are as follows:

1. Compute the gradient image for multi-scale structural similarity image.
2. Modify gradient image by marking foreground object (internal marker) and background object (external marker).
3. Apply watershed algorithm on modified gradient image for segmentation.

Here, the gradient image from the similarity image is computed by eq(10).

\[ S_g(x, y) = \sqrt{(I_{MSSSIM}(x, y) * \nabla_x)^2 + (I_{MSSSIM}(x, y) * \nabla_y)^2} \]  

Where \( I_{MSSSIM}(x, y) \) is the similarity image, \( \nabla_x \) and \( \nabla_y \) are the Sobel derivative filters in the \( x \) and \( y \) directions and \( * \) denotes the convolution operator.

Gradient image \( S_g(x, y) \) is modified as:

\[ \tilde{S}_g(x, y) = \text{imimposemin}(S_g(x, y), BGM|FGM) \]  

where \( \text{imimposemin} \) is a matlab function which modifies the gradient image \( S_g(x, y) \) using morphological reconstruction so it has regional minima wherever \( (BGM|FGM) \) is nonzero [15], BGM and FGM are the background and foreground marker, \( | \) is logical OR operator.
In this work, foreground marker (FGM) and background marker (BGM) are computed using the following steps [16]:

1. Compute foreground marker (FGM)
   i. Apply morphological erosion followed by morphological reconstruction on similarity image \( I_{MSSSIM}(x,y) \) and give a name \( I_{re-ero-img}(x,y) \). In reconstruction step, marker image is the erode image \( I_{ero-img}(x,y) \) and mask image is the original similarity image \( I_{MSSSIM}(x,y) \). Here we have considered disk shape structuring element with a radius of 10 pixels in order to identify small object as well.
   ii. Apply morphological dilation followed by morphological reconstruction on \( I_{re-ero-img}(x,y) \) and give a name \( I_{re-di-img}(x,y) \). In the reconstruction process, complement of dilated image \( I_{di-img}(x,y) \) is used as the marker image and mask is the complement of \( I_{re-ero-img}(x,y) \).
   iii. Take the complement of \( I_{re-di-img}(x,y) \).
   iv. Calculate the regional maxima of \( I_{re-di-img}(x,y) \) in order to get good foreground markers \( I_{fgm-img}(x,y) \).
   v. Apply morphological closing operation on \( I_{fgm-img}(x,y) \) by the 5 \( \times \) 5 structuring element and give a name \( I_{fgm-close}(x,y) \).
   vi. Erode \( I_{fgm-close}(x,y) \) by the 5 \( \times \) 5 structuring element and give a name \( I_{fgm-close-ero}(x,y) \).
   vii. Remove all connected components from \( I_{fgm-close-ero}(x,y) \) that have fewer than 50 pixels and use this image as foreground marker (FGM).

2. Compute background marker (BGM).
   i. Compute binary threshold image of the cleaned up image \( I_{re-di-img}(x,y) \) and give a name \( I_{bn}(x,y) \).
   ii. Thin the background by computing the watershed transform of the distance transform of the \( I_{bn}(x,y) \).
   iii. Then compute watershed ridge line and used it as background marker (BGM).

2.4. Weight Calculation and Regional Fusion

This section presents regional fusion based on weight calculation and reconstruction of fused region in order to get final fused image. After segmentation of input images into some regions, regions are transformed into row vectors (lexicographical order). In each entry of the vectors, we have stored three values (pixel value, x-coordinate of the pixel value, y-coordinate of the pixel value), because to preserve local properties of a pixel about its neighborhoods. Thereafter, two same corresponding regions are considered from two different source images for fusion. For fusion of two regions, weight is computed for each pixel belonging to the regions. Basically, the notion of information in an image is usually represented by its features that are usually in the form of two-dimensional signals including 2-D neighborhood information. Hence, the weight of a pixel in the region is computed in the neighborhood of that pixel and it signifies how much information it contains. In this work, we have considered 5 \( \times \) 5 neighborhood window centered at the current pixel position. Regions are not symmetric in their size, so some pixels which are may be boundary pixels or may not be boundary pixels don’t have all neighborhoods. Therefore, all the empty coefficients in 5 \( \times \) 5 window are filled up with zero.

The weight for a pixel in the region is calculated as the second central moment of that pixel in the neighborhood with respect to the asymmetry or skewness of the whole region. The second central moment of a pixel describes the relative importance over its neighborhood and the skewness of the region describes the distribution of the pixels in the region. Therefore, the ratio of these two measures of a pixel tells us how much information or salience feature it contains.

\[
\text{Weight} = \frac{M_2}{M_3}
\]  

(12)

Where, \( M_2 \) is second central moment of a pixel in the neighborhood and \( M_3 \) is the third central moment of the region. The second central moment \( M_2 \) is computed in terms of ordinary moment,

\[
M_2 = a_2 - \mu^2
\]  

(13)

\[
a_2 = \sum_{i=1}^{n} [I(x,y)]^2 p_{x,y}
\]  

(14)

Where, \( a_2 \) is the second moment about the origin, \( \mu \) is the mean value, \( I(x,y) \) and \( p_{x,y} \) are the intensity values of pixel \((x,y)\) within window and probability of the intensity at coordinate \((x,y)\) in the window and \( n \) is the number of element within specified window.
Here, asymmetry of the region is computed by taking third central moment. The third central moment $M_3$ is computed in terms of ordinary moment,

$$M_3 = a_3 - 3a_2\rho + 2\rho^3$$  \hspace{1cm} (15)

$$a_3 = \sum_{j=1}^{m} [R(x,y)]^3 p_{x,y}$$  \hspace{1cm} (16)

Where, $a_3$ and $\rho$ are the third moment about the origin and mean value of the region, $R(x,y)$ is the region intensity value and $p_{x,y}$ is the probability of the intensity at coordinate $(x,y)$ in the region, $m$ is the number of pixels within region.

Finally, based on weight, pixels are selected from both the region. Our assumption is that the pixel with large weight, consists most of the important information (like edges, corner, contrast, etc.) of an image, because we know that the useful features in the image usually are larger than its neighboring pixels. The pixel selection rule for region fusion is implemented using method given in eq(17). During pixel selection, if one pixel within a specified window $(5 \times 5)$ in the fused region is selected from the region $R_A$ and all other pixels are selected from the region $R_B$, then we simply handle this problem by replacing that pixel with the corresponding pixel in $R_B$. 
Table 1. Average OPM, SD, MI score of four different methods

<table>
<thead>
<tr>
<th>Methods</th>
<th>OPM</th>
<th>SD</th>
<th>MI</th>
</tr>
</thead>
</table>

\[
F_R = \begin{cases} 
R_A(x,y) & \text{if Weight}(R_A) \geq \text{Weight}(R_B) \\
R_B(x,y) & \text{if Weight}(R_A) < \text{Weight}(R_B) 
\end{cases} \tag{17}
\]

Where, \( R_A \) and \( R_B \) are the regions of the input image \( A \) and \( B \). \( F_R \) is the corresponding fused region.

After fusion of all the regions from both the input images, final fused image is obtained by combining all the fused regions.

3. Experimental Results

In this section, we show the experimental results of our method in terms of performance index. We have tested our proposed method on 120 multi-sensor image pairs collected from Manchester University UK database [17] and compared with three state-of-the-art methods from objective and visual perspective. The three state-of-the-art methods: region segmentation and expectation maximization (RSEM) [13], image segmentation and fuzzy logic (ISFL) [12], region segmentation and spatial frequency (RSSF) [8], are used because all the methods are region based, difference type of fusion rule, recency, and easy to implement. Due to unavailability of the ground truth fused image, performance of the fusion algorithm is computed using two input images and the fused image. In this work, we use three fusion performance metrics: objective performance metric (OPM) [18], standard deviation (SD) [19], mutual information (MI) [20]. These metrics measure how well a feature (e.g., edge, contrast, amount of mutual information) transfers from the source images into the fused image. Fig. 2 shows the result of filter mask. In Fig. 2, (a) and (b) are the input images and (c), (d) are the corresponding filtered images. In Fig. 3 we present the similarity results and segmentation results using our method and three other methods. Fig. 3(a), (b) are the input images and (c), (d), (e) are the similarity images using our method, RSSF and ISFL. Fig. 3(f - i) are the segmentation map images using our method, RSEM, RSSF and ISFL. Fig.4 shows the subjective fusion result of four fusion methods. Fig.4 (a), (b) are the input images and (c), (d), (e), (f) are the fused images using our method, RSSF, RSEM and ISFL. In table1, we show the comparative results of different fusion techniques for all test image pair. All the values in table 1 are the average OPM, SD, and MI over 120 image pairs for four fusion techniques. From this table, it can easily be observed that the integrated ranking of our method takes the leading position.

4. Conclusion

In this paper, we have proposed a fusion algorithm based on region segmentation and weight calculation. In this work, the similarity image is computed for segmentation of source images. The salient information is adequately identified from the regions by calculating the second central moment of a pixel in the neighborhood with respect to the asymmetry of the whole region for fusion implement. To validate this new approach, the approach was tested on 120 sensor image pairs collected from Manchester University UK database. The experimental results of our method show that our method achieved superior results over other three methods. In addition, our method has a number advantages over some other region based fusion algorithm have been introduced, such as easy to implement without parameter setting, better from visual perceptio. Region classification and fusion approach may be employed in future for better fusion result in terms of visual and objective perception.
Fig. 4. (a-b) are the input images, (c) corresponding fused image using our method, (d) fused image using RSSF, (e) fused image using RSEM, (f) fused image using ISFL.

Acknowledgements

Authors are thankful to Department of Computer Science and Engineering, Jadavpur University and National Institute of Science and Technology, India for providing necessary infrastructure to conduct experiments relation to this work.

References


