A Novel Image Fusion Scheme by Integrating Local Image Structure and Directive Contrast

Zhang-Shu Xiao
Biomedical Engineering Research Institute
Xi'an Jiaotong University
Xi'an, P.R. China
xzs0410@yahoo.com.cn

Chong-Xun Zheng
Biomedical Engineering Research Institute
Xi'an Jiaotong University
Xi'an, P.R. China
cxzheng@mail.xjtu.edu.cn

Abstract - In multi-resolution analysis of images, the total structures are characterized by the approximate component, and the specific features such as lines, edges and contours are contained in the detail components. This paper proposes a fusion scheme to combine the structure information which represents the background of the source images and the directive contrast information which contains the features of lines, edges and contours along different directions in the source images into a single image. Here, the structure information can be attained by similar structures index. And the contrast information is defined as the ratio of the local maxima of the detail components in a Gaussian window and the mean of the corresponding approximate components. The visual and statistical results show that the proposed method has improved the fusion performance compared to the existing fusion schemes, especially in term of the visual effects.

Keywords: image fusion; structure information; directive contrast; structure similarity index; visual effects.

1 Introduction

Image fusion can be broadly defined as the process of combining multiple input images into a smaller collection of images, usually a single one, which contains the ‘relevant’ information from the inputs, in order to enable a good understanding of the scene, not only in terms of position and geometry, but more importantly, in terms of semantic interpretation. In this context, the word ‘relevant’ should be considered in the sense of ‘relevant with respect to the task the output images will be subject to’, in most cases high-level tasks such as interpretation or classification. For instance, doctors can manually combine CT and MRI medical images of a patient with a tumor to make a more accurate diagnosis, but it is inconvenient and tedious to finish this job. And more importantly, using the same images, doctors with different experiences make inconsistent decisions. Thus, it is necessary to develop the efficiently automatic image fusion system to decrease doctors’ workload and improve the consistency of diagnosis. In recent years, image fusion has been a focused research field with plethora algorithms proposed, using a lot of image processing and information fusion techniques. The simplest way to obtain a fused image from two or more medical images is to average them. Although mostly preserving the original meaning of the images, it is prone to reduce the contrast of the fused image. With developments of Marr’s vision and applications of multi-resolution image processing techniques, the potential benefits of multi-resolution image fusion schemes have been explored in order to improve the contrast of the fused image. P.J. Burt [1] [2] proposed the Laplacian pyramid based and gradient pyramid based image fusion methods. A.Toet[3] defined a contrast based on the ratio of low-pass pyramid, and given the contrast pyramid based image fusion methods. H.Li [4] employed wavelet pyramid to develop a scheme which can exact the localized characteristic of input images. Y. Chibani[5] used the multi-scale pyramid, which is over-complete representation of the original images, to merge different images into a single one to adapt the invariance with respect to elementary geometric operations such as translation, scaling, and rotations. E.P. Simoncelli[6] investigated the steerable pyramid based image scheme which can extract detail information with more directions. Pu and Ni[7] suggested a contrast-based image fusion method using the DWT. They use the directive contrast as the activity measure. More multi-resolution image fusion schemes refer to [8].

Most of present image fusion methods aim at obtaining as many as information from the different modality images. The fusion criterion is to minimize different error between the fused image and the input images. However, not all the exacted information can help human being observe the useful objects. As Marr’s vision stated, human visual system (HVS) is not only sensitive to the structure features of the scene, but also sensitive to the local contrast. Motivated by this theory, Zhou Wang [10] investigated the Structure SIMilarlarity index (SSIM) to measure the structure similarity of two input images. This paper proposes a fusion scheme to composite the overall structure information from the background of input image and the contrast information on behave of the features of the image such as lines, edges and transitions. Here, the structure information can be attained by SSIM methods of
Zhou Wang. And the contrast information is defined as the ratio of the local maxima of the detail components in a Gaussian window and the mean of the corresponding approximate components, which is different from the methods proposed by Pu and Ni [7] and is more robust especially taking account into the noised input images. The visual and statistical results show that the proposed method has improved the fusion performance compared to the existing fusion schemes. Some medical image fusion presents its good promising applications.

2 Structure similarity index

Based on the fact that natural images are highly structured, which means the pixels with the similar features usually form an interesting region or object, the structure similarity index (SSIM) to assess the image quality was proposed by Wang and Bovik[9]. Compared to the classic Minkowski error metric (mean square error-MSE) which measures differences of corresponding pixels between the reference and the test images, the SSIM tried to measure the structure differences of both images. The most advantage of the SSIM is qualitatively consistent with human visual system which is highly adapted to exact the structural information from the scene. Wang and Bovik combined three features of an image to yield an overall similarity measure.

\[
SSIM(x,y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \cdot \frac{2\sigma_{xy} + C_2}{\sigma_{x}^2 + \sigma_{y}^2 + C_2} \cdot \frac{\mu_x + \mu_y + C_3}{\sigma_x + \sigma_y + C_3}
\]  

Here, the two images: \( x = \{x_i | i = 1,2,\cdots,N\} \) and \( y = \{y_i | i = 1,2,\cdots,N\} \). \( \mu_x, \mu_y \) are mean luminance of the test image \( x \) and the reference image \( y \), and \( \sigma_x, \sigma_y \) are the standard deviations of them within a local window. \( \sigma_{xy} \) is the covariance of \( x \) and \( y \). They are defined as follows:

\[
\mu_x = \frac{1}{N} \sum_{i=1}^{N} x_i
\]

\[
\sigma_x = \left( \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \mu_x)^2 \right)^{1/2}
\]

\[
\sigma_{xy} = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \mu_x)(y_i - \mu_y)
\]

Therefore, the first component \( l(x,y) \) with a value range of \([0,1]\) measures how similar the mean luminance of image \( x \) and image \( y \) is. The second component \( c(x,y) \) also with a value range of \([0,1]\) measures how similar the contrast of both images is. And the third component \( s(x,y) \) can be viewed as the measurement of the structure similarity of the corresponding images. Its dynamic range is \([-1,1]\). Here \( C_1, C_2, C_3 \) are small constants given by:

\[
C_1 = (K_1 L), \quad C_2 = (K_2 L), \quad C_3 = C_2 / 2
\]

\( L \) is the dynamic range of the pixel values (\( L=255 \) for 8bits/pixel gray scale images), and \( K_1<<1 \) and \( K_2<<1 \) are two scalar constants.

3 The improved wavelet directive contrasts

Usually, the contrast of an image is defined as

\[
C = (L - L_b) / L_b = L_{II} / L_b
\]

Here the intensity of image pixel- \( L_b \) is the intensity of the background of the pixel, or local low frequency component, \( L_{II} = L - L_b \) is supposed as the local high frequency component. Wavelet transform decomposes the image into different level of pyramid structure of wavelet coefficient based on scale and direction. Then vertical, horizontal and diagonal contrast based on the directive wavelet coefficients was defined by Pu and Ni [7].

Here, \( A' \) is the approximation component of the wavelet transform, which contains the low frequency information of the processing image, while \( D_{v}', D_{h}', D_{d}' \) are the detail components, which contain the vertical, horizontal and diagonal edge information, respectively.

In this paper, we construct the improved directive contrasts which are defined as follows:

\[
C_{v}' = \frac{D_{v}'}{A'}, \quad \text{vertical contrast}
\]

\[
C_{h}' = \frac{D_{h}'}{A'}, \quad \text{horizontal contrast}
\]

\[
C_{d}' = \frac{D_{d}'}{A'}, \quad \text{diagonal contrast}
\]

Here \( M' \) is the matrix of local mean value of the approximate coefficient at level \( j \), while the max(\( D_{v}' \)), max(\( D_{h}' \)), and max(\( D_{d}' \)) are the most maximum coefficients of corresponding detail components at level \( j \) within the local window, respectively. Therefore, we obtain three new contrasts \( C_{v}' \), \( C_{h}' \), and \( C_{d}' \) in the wavelet domain, which represent the most significant features relatively to the background of the local window along vertical, horizontal, and diagonal directions respectively. Compared to previous contrasts, the advantages of the proposed contrasts are:

1) Using the mean value instead of the single pixel approximate coefficient as the background component can...
to avoid computing instability since approximate coefficient is zero or near zero and increase the robustness of the algorithm.

(2) Using the maximum coefficient of the local window as the detail components can avoid selecting zero or near zero directive detail coefficients of the single due to the sparsity of wavelet coefficients, while the maximum coefficient can represent the main features of the local window.

4 Image fusion scheme

4.1 Selection of fusion modes

In pattern-selective image fusion, the fused image is assembled from selected component patterns of the input images. Therefore, modes of combination or the fusion rules play a vital role in deciding the quality of the fused image. P.J. Burt [2] suggested adopting selection or average modes respectively according to match measure. Wang and Bovik’s [9] experiments exhibit that the SSIM is an effective metric to quantify the degree of similarity of the structure information between the test image and the reference image. And in the multi-resolution wavelet decomposition, the approximate component of the last level represents the total structure of input images, and the detail components contain features such as lines, edges, contours along different directions. Therefore, different fusion rules should be adopted for merging the approximate component and the detail information. The SSIM method discussed above which is capable of expressing structure information is suitable for combining the approximate component of input images. And as for detail information, we can use the proposed directive contrasts as the match measure to merge them. Without loss of generality, given the input images \( x \), \( y \), the fused image \( f \), and the largest decomposition scale \( L \), at first, each of the input images is decomposed by wavelet transform into vertical, horizontal and diagonal detail coefficients \( D^v_{x,y}, D^h_{x,y}, D^d_{x,y} \) and \( D^{v,y}_{x,y}, D^{h,y}_{x,y}, D^{d,y}_{x,y} \) where \( j=1,2,3 \), and the approximate coefficients at the largest scale \( A^l_x \) and \( A^l_y \). And then, the detail coefficients and the approximate coefficients can be processed in different ways to attain the fused image.

4.2 Fusion of the approximate coefficients

(1) The SSIM(\( A^l_x \), \( A^l_y \) | \( w \)) is computed. Here \( w \) is a local window. We use a circular symmetric Gaussian window in order that the SSIM map has a locally isotropic property to reduce undesirable ‘blocking’ artifacts.

(2) Given a threshold \( T \) (a positive constant less than 1, usually larger than 0.75), the fused image \( f \) can be attained. If SSIM(\( A^l_x \), \( A^l_y \) | \( w \)) is equal or larger than \( T \), the corresponding regions of the approximate images have the similar structures in general, which means that an average fusion mode should be adopted in order to not only make full use of the structure information from the source images but also reduce the image noise. Otherwise, the selection mode is more suitable for transferring the corresponding source images with locally larger features into fused image. we summarize the image fusion rules in equation (9).

\[
\begin{align*}
    f &= \begin{cases} 
        A^l_x, & \text{SSIM}(A^l_x, A^l_y) < 0.75 & \text{or} & A^l_x \leq A^l_y \\
        A^l_y, & \text{SSIM}(A^l_x, A^l_y) \geq 0.75 & \text{or} & A^l_y \leq A^l_x 
    \end{cases}
\end{align*}
\]

Here \( \lambda(w) \) is the local weight with a value range of \([0.4, 0.6]\). The average value is 0.5, which means that the fused image takes the mean of the two input images in the corresponding locations.

The difference between the proposed image fusion scheme and the existing methods to merge the approximate images at the coarsest scale lies in the use of the match measure. In previous fusion methods, a simple average and weighted average way is just used to merge the approximate components of the input images. This will always blur the structures of the fused image, and sometimes even cancel out the complementary structures in different input images due to different image formation. Here, we use the SSIM as the match measure to determine which mode is suitable for the corresponding regions of the approximate images. When the value of the SSIM(\( A^l_x \), \( A^l_y \) | \( w \)) is equal or larger than the threshold \( T \), which means the two input images have similar structures, average mode should be used to combine the input images in order to reduce the noise and improve the robustness. When the value of the SSIM(\( A^l_x \), \( A^l_y \) | \( w \)) is less than the threshold \( T \), which means the input images have different even complementary structures, and the select mode is more reasonable for merge them in order to avoid blurring the structure of the fused image. Since the SSIM measure synthesizes several features \( l(x,y), c(x,y) \), and \( s(x,y) \) to form a universal match measure, it can effectively select the structure information which is easily sensible to human visual system into the fused image.

4.3 Fusion of the detail coefficients

(1) To construct the directive contrasts using the detail coefficients according to equation (9).

(2) To merge the directive contrasts using the maximum selection mode according to equation (10).

(3) To make consistency verification by applying a majority filter in order to remove possible wrong selection decisions caused by impulsive noise. Consistency verification is motivated by the fact that significant image features tend to be stable with respect to variations in space. Thus, when comparing the corresponding image features in multiple input images, consideration of the space dependencies between transform coefficients can lead to a more robust fusion strategy.
4.4 Integrating approximate and detail coefficients.

Now, combining fusion process of the approximate coefficients and detail coefficients, we attain the fusion pyramid. Furthermore, by inverse wavelet transform of the fused pyramid, the result image is attained. Since this method integrates the local structures of the approximate components and the local contrasts of the detail components, it is called ‘LStrC’ method.

5 Experimental study and performance assessments

In many applications of image fusion, the human beings such as diagnostic doctors of the medical images, the interpreters of the remote sensing images, are the final observers of the fusion results. Thus, the human perception of the merged image is of paramount importance and therefore, fusion results are always evaluated by subjective criteria, but these subjective assessments are difficult to reproduce or verify in that the quality measure depends highly on some personal factors such as the observer’s professional knowledge, experience and perception capabilities of the human visual systems. Furthermore, the evaluation process is time-consuming and expensive. Therefore, with the development of image fusion, the investigation of the objective measures that can quantify the performance of fusion algorithms is an urgent need. There are two common ways for objective
evaluation of the fused image. One is to compare the fused image to the ideal image. The criterion is the mean square error (MSE) between the fused image and the ideal image. Then, another problem arises, namely how to get the ideal image in various specific applications of image fusion. Thus, this way has only theoretic analysis value, and it is impractical for many situations of image fusion. A more practicable way is to develop performance measures which can evaluate the quality of the fused image without the ideal image. That way is named as non-reference image quality assessment. For example, Petrovic and Xydeas[10] proposed an objective edge based performance measure $Q_{AB/F}$ which computes the amount of edge information that is transferred from source images to the fused image. Piella and Heijmans [11] proposed a new quality metric (here, we call it $Q_y$), $P$ stand for Piella, the following definitions as the same with it) for image fusion based on research by Wang and Bovik [9] on a structural similarity (SSIM) measure. C.Yang et al.[12] given an improved SSIM-based measure ($Q_y$) which can evaluate the amount of the structure information in the fused image transferred from the input images by take account of the structure similarity between both input images. The measure carries out a quantitative correlation analysis between the source images and the fusion image, which can give a good assessment of various image fusion algorithms.

In this section, by fusing CT/MRI images we try to compare the subjective visual effects and the objective performance of proposed fusion scheme to that of Laplacian pyramid based fusion (calling it 'LP' method) and gradient pyramid (GP) fusion proposed by P.J. Burt [1][2], the ratio of low-pass pyramid (CP) based fusion suggested by Toat[3], the conventional discrete wavelet transform (DWT) image fusion given by H.Li[4], and steerable pyramid (StrP) based fusion proposed by E.P. Simoncelli [6]. For medical diagnosis, CT provides better information on denser tissue with less distortion than MRI, and MRI gives better information on soft tissue than CT. Thus, in order to make full use of the information of the two modality images for medical diagnosis, doctors usually observe the images manually and fuse them in their minds. However, this process is very tedious and tired. Moreover, doctors with different medical knowledge and clinic experience will get different diagnostic results. In order to reduce personal inconsistence and doctors’ workload, to merge CT/MRI images automatically is an urgent need in clinic diagnosis. Fig. 1 (a), (b) are the source images of CT and MRI of a patient with a brain tumor. Fig.1 (c), (d), (e), (f), and (g) are the fused results using the methods based on CP, LP, DWT, GP, and StrP respectively. Fig.1 (h) is attained by the proposed method LStrC. Fig.1 (c) shows that the fused image based on CP method is not so good. And the results of LP, GP, StrP, and DWT almost have the same visual effects. The proposed fusion scheme presents lightly better visual effect but has less distortion especially in respect of the large structure features such as the outlines of skulls and brain tumor. The large structure information is more important than details for doctors to diagnose the tumor status. Therefore, in view of the medical diagnosis, the proposed method provides better results compared with the others.

Above, we compare the perception results of LStrC fusion methods to several classic image fusion schemes. To further evaluate quantitatively the ability of different fusion methods in respect with exacting the large features we adopt the edge information metric $Q_{AB/F}$ proposed by V. Petrovic [10] and the structure information measure suggested by C.Yang et al.[12] to compare the statistical results of the above discussed image fusion algorithms. The data in table 1 show that LStrC not only has maximum value of $Q_{AB/F}$, but also has maximum value of $Q_y$. This means that the fused image attained by LStrC methods not only contains more edge information but also keeps the structure features better. That is to say the general performance of the proposed algorithm is best among these algorithms. This compared result is consistent with the visual effects of the fused images.

<table>
<thead>
<tr>
<th>Metrics</th>
<th>CP</th>
<th>LP</th>
<th>GP</th>
<th>DWT</th>
<th>StrP</th>
<th>LStrC</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Q_{AB/F}$</td>
<td>0.556</td>
<td>0.564</td>
<td>0.572</td>
<td>0.561</td>
<td>0.576</td>
<td>0.610</td>
</tr>
<tr>
<td>$Q_y$</td>
<td>0.550</td>
<td>0.569</td>
<td>0.576</td>
<td>0.565</td>
<td>0.583</td>
<td>0.613</td>
</tr>
</tbody>
</table>

6 Conclusions and acknowledge

In this paper, we investigate a new image fusion scheme by fusing the wavelet approximate coefficient based on local structure similarity match measure and by fusing the wavelet detail coefficients according to the new directive contrasts. CT/MRI image fusion experiments show that the LStrC method can effectively extract larger structure features with less distortion compared to the existing methods. As we know, large structure features are important for doctors in clinic medical diagnosis. Therefore, our proposed method provides doctors an effective and efficient tool to make full use of complementary information from different source images to attain a more accurate diagnosis. This work was supported by the key project of Shaanxi Normal University (No.995285). Moreover, the author would like to thank Zhou. Wang who provides the free program of SSIM algorithm in his homepage: http://www.cns.nyu.edu/~zwang/.

References


1953


