Image Fusion with Simultaneous Orthogonal Matching Pursuit

Satya Rahul K, Mallikarjuna Lingam.K
Malla Reddy College of Engineering & Technology, Secunderabad
JNTU Kakinada, Kakinada.

Abstract: Image fusion integrates the information from multiple images of one scene to get an informative image which is more suitable for human visual perception or further image-processing. In addition, the simultaneous orthogonal matching pursuit technique is introduced to guarantee that different source images are sparsely decomposed into the same subset of dictionary bases, which is the key to image fusion. The proposed method is tested on several categories of images and compared with some popular image fusion methods. The experimental results show that the proposed method can provide superior fused image in terms of several quantitative fusion evaluation indexes.

Keywords: Discrete Wavelet Transform (DWT), Discrete cosine transform(DCT), Curvelet Transform(CVT), Quality index

I. INTRODUCTION

For one scene, many images can be simultaneously acquired by various sensors with development of numerous imaging sensors. Those images usually contain complementary information which is dependent on the natural properties of the sensors and the way the images are obtained. Image fusion can effectively extend or enhance information of the scene by combining the images captured by different sensors. The fused image can improve edge detection, image segmentation and object recognition in medical imaging, machine vision and military applications. In past two decades, many techniques and software for image fusion have been developed. According to the stage at which the information is combined, image fusion algorithms can be categorized into three levels, namely pixel-level, feature-level, and decision-level. The pixel-level fusion combines the raw source images into a single image. Compared to feature or decision-level fusion, pixel level fusion can preserve more original information. Feature-level algorithms typically fuse the source images using their various feature properties, such as regions or edges. Thus, this kind of methods is usually robust to noise and misregistration. Decision level fusion algorithms combine image descriptions directly, for example, in the form of relational graphs. But the decision-level fusion methods are very much application dependent. In this paper, we only focus on the pixel-level image fusion problem. The goal of pixel-level image fusion is to combine visual information contained in multiple source images into an informative fused image without the introduction of distortion or loss of information. In the past decades, many pixel-level image fusion method have been proposed SOMP-DCT,SOMP-Hybrid and SOMP-Trained.

The outline of this paper is as follows: in Section 2, The SOMP based multi-sensor image fusion scheme is described and SOMP-Trained. Experimental results are presented in Section 4, where the proposed method is compared with some popular ones, especially the methods based on multi-scale transforms.

II. FUSION SCHEME

A. Simultaneous orthogonal matching pursuit

Two of the most frequently discussed approaches are the matching pursuit (MP) and the orthogonal matching pursuit (OMP) algorithms. They are greedy algorithms that select the dictionary atoms sequentially.

For the image fusion problem, multiple source images need to be decomposed simultaneously. However, for the MP and OMP, the decomposed sparse coefficients of different images may correspond to different subset of atoms of the dictionary. This is similar to a decomposition of each input image patch using a different family of wavelet. Thus, the fusion rule will be hard to design. For image fusion, we hope that the different source images are decomposed into the same subset of dictionary atoms. In this paper, the SOMP technique is employed to solve this problem. The SOMP is a variant of OMP and it assumes that different signals can be constructed from the same sparse set of basis atoms, but with different coefficients. In the SOMP algorithm, a fixed dictionary $D = [d_1, d_2, ..., d_T]$ is used to represent each of the $K$ signals $(X_k)_{k=1}^{K}$. Each SOMP iteration selects the column index that accounts for the greatest amount of residual energy across all signals.

SOMP Algorithm:

Input: Dictionary $D \in \mathbb{R}^{n \times T}$, signals $\{X_k\}_{k=1}^{K}$, $x_k \in \mathbb{R}^{n}$, threshold $\epsilon$, an empty matrix $\Phi$.

Output: Sparsity coefficients $\{c_k\}_{k=1}^{K}$, $\epsilon \in \mathbb{R}^{T}$. 

ISSN (Online): 2347 - 2812, Volume-3, Issue -9, 2015
Therefore, directly applying the 2D images for y when in the overlapping patches. In fact, the task of which indicates the t information of the source images. Hence, the local mean of each

-thesized to represent the sharpness or the edges of an image. Therefore, transformed coefficients (each corresponds to a transform basis) of an image are meaningful to detect and emphasize salient features. The larger the absolute value of the coefficients is, the more information it contains.

Averaging and absolute maximum are usually used as the fusion rules. The averaging rule means that the corresponding coefficients are averaged by the weight which is depended on the activity-level. It preserves the contrast information of the source images. However, the detailed features such as the edge or the lines would get smoother. For the coefficients combining method, we hope that it could integrate the visual information contained in all the source images into fused image with no distortion or loss of information. However, transforming all the visual information from input source images into the fused image is almost impossible. As a result, a more practical way for image fusion is that the fused image is represented from only the most important input information. Therefore, the choosing absolute maximum (maxabs) rule is usually used for image fusion. The combined coefficients are obtained by selecting entry of coefficients with maximum absolute value. As mentioned before, the local mean of each patch is similar as the low frequency coefficients of the wavelet transform. Thus the local mean of each patch is combined by averaging technique which is often used in multiscale transform based image fusion schemes.

C. Fusion scheme

The image fusion scheme using image sparse representation theory is summarized in Fig. 1. In image fusion, it is essential that the source images are registered, which means that the objects in all images are geometrically aligned. Blanc et al. showed that a small geometrical distortion may produce a noticeable effect on the quality of fused images. In addition, Zhang and Blum discussed the use of image registration in image fusion in. In fact, the task of registration is very challenging, particularly when images are captured with cameras where extrinsic and/or intrinsic parameters are different. Thus, many researchers study the image registration problem separable with the image fusion, and many effective methods for image registration have been proposed. In this paper, we assumed the K source images, \( I_1, I_2, \ldots, I_K \) with size of \( M \times N \), are geometrically registered. The whole fusion procedure of the proposed method in this paper takes the following steps:
1. Divide each source image $I_i$, from left-top to right-bottom, into every possible image patches of size $\sqrt{n} \times \sqrt{n}$, which is the same as the size of the atom in dictionary. Then all the patches are transformed into vectors via lexicographic ordering and $\{v_{k}^{i}\}_{k=1}^{N}$ is obtained.

2. Decompose the vectors at each position, $i$, with different source images, $\{v_{k}^{i}\}_{k=1}^{K}$ into theirs sparse representations $a_{1}^{i}, a_{2}^{i}, ..., a_{k}^{i}$ using the SOMP.

3. Combine the sparse coefficient vectors using the max-abs rule:

$$a_{i}^{k}(t) = \alpha_{i}^{k}(t), k = \arg \max_{k=1,...,K} \left| \alpha_{i}^{k}(t) \right| \quad (4)$$

Where $a_{i}^{k}(t)$ is the $t^{th}$ value of vector $a_{i}^{k}$, $t=1,2,...,T$. The fused vector $v_{f}^{i}$ is obtained by:

$$v_{f}^{i} = Dx_{f}^{i} \quad (5)$$

4. Iterate all the vectors, and reconstruct the fused image $I_{f}$. Firstly, each vector $v_{f}^{i}$ is reshaped into a block with size $8 \times 8$. Then, the block is added to $I_{f}$ at the responding position. Thus, for each pixel position, the pixel value is the sum of several block values. Finally, the pixel value is divided by the adding times at its position to obtain the final reconstructed result $I_{f}$.

Notice that two steps in this algorithm make the fusion scheme be shift-invariant, which is of great importance for image fusion. Firstly, the overcompleteness of the dictionary makes the sparse representation is shift-invariant. Secondly the “sliding window” operation is a shift-invariant scheme.

### III. EXPERIMENTS

#### A. Experimental setup

The size of the “sliding window” and the global error of SOMP are two important parameters. As to the size of the “sliding window”, the larger the widow, the bigger the vector transformed from the patch. And the size of the corresponding dictionary also increases according to the sparse theory introduced in Section 2. Thus the process of the SOMP becomes slower. As the size of the “sliding window” decreases, the process of the SOMP becomes faster, but the information contained in the patches would be not sufficient. It may miss some of the important features of the source images. In the following, the patch size is set to $8 \times 8$ which has been proved to be appropriate setting for the image denoising application. In this paper, we assume that the source images are all clear. So a small global error is set, i.e., $e=0.01$.

Fig. 1. Procedure of image fusion based on sparse representation.

Three kinds of overcomplete dictionaries are used to test the performance of the proposed method. The first one is the overcomplete DCT bases. We sample the cosine wave in different frequencies to result 16 vectors with length of 8. The 2-D overcomplete separable version of the DCT bases consist of all possible tensor products of 1-D bases. Then, the overcomplete DCT dictionary are constructed by lexicographically ordering the 2-D overcomplete DCT bases into vectors with length of 256. The 2-D overcomplete separable version of the DCT bases are presented in Fig. 4a. The second one is the hybrid dictionary which consists of DCT bases, wavelet ‘db1’ bases, Gabor bases, and ridgelet bases, as shown in Fig. 4b. The 1-D ‘db1’ bases function is

$$\psi_{i}^{(m)}(x) = \psi(2^{i}x - j) = \begin{cases} 1 & \text{for } 0 \leq x \leq 1/2 \\ -1 & \text{for } 1/2 \leq x < 1 \\ 0 & \text{Otherwise} \end{cases} \quad (6)$$

The 2-D wavelet basis consists of all possible tensor products of 1-D basis functions. The Gabor bases function is defined as:

$$G(x, y, k_{x}, k_{y}) = \exp \left[ \frac{-(x - k_{x})^{2} + (y - k_{y})^{2}}{2\sigma^{2}} \right] \cdot \psi(w(x - k_{x}))$$

where $x$ and $y$ represent the spatial coordinates while $k_{x}$ and $k_{y}$ represent the frequency coordinates. $X$ and $Y$ are the spatial localizations of the Gaussian window. The 2D ridgelet is defined using a wavelet function as:

$$\psi_{x, b, a} = a^{-1/2} \psi((x_{1} \cos \theta + x_{2} \sin \theta - b) / a)$$

where $w(.)$ is a wavelet function. Each kind of bases consisted of 64 non redundant bases, thus the size of the overcomplete dictionary is also 256. The third one is the trained dictionary obtained from learning natural sample using the iterative K-SVD algorithm, which has been proved effective when it is used to train the overcomplete dictionary. The training data consisted of 50,000 $8 \times 8$ patches, randomly takes from a database of 50 natural images. A fraction collection of the 50 natural images and the blocks are presented in Fig. 2d. The iteration is set to 200. The obtained dictionary is shown in Fig. 2c.
We test the proposed method on several pairs of source images including computed tomography (CT) and magnetic resonance imaging (MRI) images, infrared and visual images, and optical multi-focus images, shown in Fig. 3. The images shown in Fig. 3a and c are two CT images that show structures of bone, while the images shown in Fig.3b and d are two MRI images that show areas of soft tissue. In clinical applications, the combined images showing clearly the position of both bone and tissue can aid in diagnosis of doctors. Fig.3e–h gives two pair of infrared and visual source images. In the infrared image, the object (the people), in Fig. 3e, is clear, while in visual image the background, such as the tree and the road in Fig. 3f, is clear. Many applications need more comprehensive information containing in both of infrared and visual images. Fig. 3i–l shows two pair of multifocus source images. Fig. 3i is near focused where the small clock is in focus and clear while the larger one is out of focus and blurred. Fig. 3j is far focused, and the situations for the clocks with different sizes are contrary. The fused image should contain both the clear clocks in Fig. 3i and j.

The proposed method is compared with SOMP-DCT, SOMP-Hybrid and SOMP-Trained. For all of those methods, the most popular setting, the max-abs fusion rule, is selected. And for each method, three levels decomposition are used. For DWT and SWT based methods, the wavelet basis is ‘db6’. We note that DWT and CVT are shift-variant transform because of the decimating operator. Thus, the fused images would have Gibbs effect in some degree. So, we combine the common Cycle Spinning algorithm to the DWT and CVT based methods to lessen the Gibbs phenomenon of the fused images. All the experiments are implemented in Matlab 6.5 and on a Pentium(R) 1.7-GHz PC with 512 M RAM.

In this paper, five objective evaluation measures, $Q_0$, $Q_w$, $Q_e$ and $Q_t$, which have been proved to be validated in large degree, are considered to quantitatively evaluate the fusion performances.

Fig. 2. The overcomplete dictionaries and the training data. (a) The overcomplete DCT dictionary; (b) the hybrid overcomplete dictionary with DCT bases, wavelet ‘db1’ bases, Gabor bases, and ridgelet bases; (c) the trained overcomplete dictionary; and (d) 50 natural images and the sampled blocks.

Fig. 3. Source images. The top row: medical images; the middle row: infrared and visual images; the bottom row: multi-focus images.

1. Image quality metric based metrics include the local quality index ($Q_0$), the weighted fusion quality ($Q_W$) measure, and the edge dependent fusion quality index ($Q_E$), which assess the pixel-level fusion performance objectively. The metric $Q_0$ between the source image $A$ and the fused image $F$ is defined as follows:

$$Q_0(A, F) = \frac{2\sigma_A \sigma_F}{\sigma_A^2 + \sigma_F^2}$$

where $\sigma_A$ represents the standard deviation of $A$ and $F$; $\sigma_A$, $\sigma_F$ denote the standard deviation of $A$ and $F$; and $\sigma_A$, $\sigma_F$ represent the mean value of $A$ and $F$, respectively. $Q_0(A, B, F)$ is the average between $Q_0(A, F)$ and $Q_0(B, F)$, i.e.,

$$Q_0(A, B, F) = \frac{Q_0(A, F) + Q_0(B, F)}{2}$$

2. The metric $Q_W$ between images $A$, $B$, and $F$ is defined as follows:

$$Q_W(A, B, F) = \sum_{w \in W} c(w) \lambda(w) Q_0(A, F|W) + (1 - \lambda(w)) \lambda(w) Q_0(B, F|W)$$

where $k(w)$ represents the relative salience of $A$ compared to $B$ in the same window $w$, and $c(w)$ denotes the normalized salience of the window $w$.

3. The metric $Q_b$ is defined as follows:

$$Q_b(A, B, F) = Q_{\text{corr}}(A, B, F) \cdot Q_{\text{corr}}(A', B', F')^a$$

where $A_0$, $B_0$, and $F_0$ are the corresponding edge images of $A$, $B$, and $F$, respectively. Parameter $a$ reflects the contribution of the edge images compared to the original images. $a$ is set to 1 in this paper. The larger the $Q_b$, $Q_0$, and $Q_w$ values, the better the fused results.

B. Experimental results

The fused images of with different fusion methods based on are shown in Figs.4a. The fusion result of the proposed method with over complete DCT dictionary, denoted by SOMP-DCT. Fig. 4b and 4c show areas of soft tissue. In clinical applications, the combined images showing clearly the position of both bone and tissue can aid in diagnosis of doctors. Fig.3e–h gives two pair of infrared and visual source images. In the infrared image, the object (the people), in Fig. 3e, is clear, while in visual image the background, such as the tree and the road in Fig. 3f, is clear. Many applications need more comprehensive information containing in both of infrared and visual images. Fig. 3i–l shows two pair of multifocus source images. Fig. 3i is near focused where the small clock is in focus and clear while the larger one is out of focus and blurred. Fig. 3j is far focused, and the situations for the clocks with different sizes are contrary. The fused image should contain both the clear clocks in Fig. 3i and j.

The proposed method is compared with SOMP-DCT, SOMP-Hybrid and SOMP-Trained. For all of those methods, the most popular setting, the max-abs fusion rule, is selected. And for each method, three levels decomposition are used. For DWT and SWT based methods, the wavelet basis is ‘db6’. We note that DWT and CVT are shift-variant transform because of the decimating operator. Thus, the fused images would have Gibbs effect in some degree. So, we combine the common Cycle Spinning algorithm to the DWT and CVT based methods to lessen the Gibbs phenomenon of the fused images. All the experiments are implemented in Matlab 6.5 and on a Pentium(R) 1.7-GHz PC with 512 M RAM.

In this paper, five objective evaluation measures, $Q_0$, $Q_w$, $Q_e$ and $Q_t$, which have been proved to be validated in large degree, are considered to quantitatively evaluate the fusion performances.

Fig. 2. The overcomplete dictionaries and the training data. (a) The overcomplete DCT dictionary; (b) the hybrid overcomplete dictionary with DCT bases, wavelet ‘db1’ bases, Gabor bases, and ridgelet bases; (c) the trained overcomplete dictionary; and (d) 50 natural images and the sampled blocks.
are the fusion results of the proposed methods of SOMP-DWT and SOMP-CVT respectively. Careful inspection of Fig. 4a reveals that the SOMP-DCT fused image contains significant reconstruction artifacts and it also losses contrast in some degree. We can see that the results of our method exhibit the best visual quality. The important features from both source images are faithfully reserved in the fused image and no reconstruction artifacts are produced. Since the proposed method also has well shift-invariant property. The image contents like tissues are clearly enhanced. Other useful information like brain boundaries and shape are almost perfectly preserved. Moreover, we also conducted the method by replacing the SOMP with the OMP, while the fusion rules and other settings are the same. We can see that the proposed SOMP-DWT and SOMP-CVT provides better visual results.

Existing method Fig. 5. The left column shows the results of SOMP-DCT, the middle one shows the results of SOMP hybrid, and the right one shows the results of SOMP-trained. Because of the lack of space, only the fused results by our are presented in fig. 6 SOMP-DWT and SOMP-CVT. We can see that the features and detailed information are presented well in the result images. For example, the second row of Fig. 5 depicts the fused images of the infrared and visual images for hybrid over complete dictionary. It is clear that the fence from the visual image is well transferred into the fused images. In addition, the details of the tree in the visual image are visually pleasing and the human figure is much bright in the fused images. The third row of Fig. 5 depicts the fused images of the multi-focus images.

IV. CONCLUSION

This paper presents a novel multi-sensor image fusion algorithm based on signal sparse representation theory. The fusion process is conducted by the simultaneous orthogonal matching pursuit (SOMP). The experiments demonstrate that the proposed method provides superior fused image in terms of the pertained quantitative fusion evaluation indexes. In addition, tuning the reconstruction error parameter of the sparse representation according to the noise standard deviation, the proposed fusion scheme is easy to extend to combining of image fusion and restoration problems. However, because the SOMP process is a greedy matching pursuit method, and the “sliding window” scheme is also time-consuming, the computation load of the proposed scheme is heavier than traditional multiscale transform methods.

REFERENCES


