Joint sparsity model improves multimodal image fusion

Haitao Yin and Shutao Li

A novel multimodal image-fusion method exploits a model that can effectively separate complementary information from source images.

With extraordinary advances in sensor technology, numerous imaging devices and techniques, such as visible cameras, IR cameras, and magnetic resonance (MR) imaging, have been developed for military and civilian applications. For a given scene, different devices often provide complementary information. MR can provide detailed information about soft tissue, but not dense structures. By contrast, computed tomography (CT) reveals dense structures, such as bones, but not soft tissue. Fusion can integrate complementary information from different devices into a single image to provide a complete and accurate description of a scene. For example, if an MR image is combined with one generated by CT, the soft tissue and bones are well visualized, which helps in making diagnoses.

Because natural scenes exhibit a multiscale structure, multiscale-transform methods, such as discrete-wavelet transform (DWT), are widely used. The major step in these techniques is to decompose source images through a multiscale transform and then to combine the decomposed coefficients at corresponding positions according to specific rules. The fused image is constructed through the inverse transform of the coefficients. The coefficients can characterize salient features, i.e., complementary information, such as edges and lines. However, each multiscale transform is generally suitable for describing only certain features. Because the goal is to integrate salient features from multiple images, it is important to extract them effectively and completely. We previously proposed an approach that uses sparse coefficients as salient features, calculating them by means of an overcomplete dictionary. Unlike multiscale transforms, the overcomplete dictionary can be trained from a set that effectively represents all features, such as edges, lines, corners, and ridges.

The joint sparsity model (JSM) is useful for recovering sparse signals. In the JSM, multiple signals provided by different sensors for the same scene form an ensemble. All signals in an ensemble have a common sparse component, and each signal has its own unique sparse component. The common and unique sparse components can be determined by

$$\min_{\Theta} \left\| \Theta \right\|_0 \text{ subject to } \left\| X - D\Theta \right\|_2 \leq \epsilon,$$

(1)

where $\Theta = [\alpha_1^T, \alpha_2^T, \alpha_3^T, \ldots, \alpha_{\Lambda}^T]^T$ consists of the common sparse component ($\alpha_i$) and unique sparse components ($\alpha_1, \alpha_2, \ldots, \alpha_{\Lambda}$), $X = [x_1^T, x_2^T, \ldots, x_{\Lambda}^T]^T$ denotes the ensemble, $\Lambda$ is the number of source images, and the matrix $D$ consists of an overcomplete dictionary $\Phi$. Because the JSM can effectively separate the common and unique components of multiple source images, we propose using it as the basis for a method of multimodal image fusion (see Figure 1). The sparse coefficients are the salient features, and the individual components represent the complementary information.

To represent local information, the new method focuses on patches. We extract two $\sqrt{n} \times \sqrt{n}$ patches from the source images $I_1, I_2$ and order them as the column vectors $x_1^i, x_2^i$. Subtracting the mean values $m_1^i, m_2^i$ yields $\tilde{x}_i^j$ ($i = 1, 2$). Using

Continued on next page
the JSM, we obtain the common sparse component $\alpha^c_j$ and the unique sparse components $\alpha^1_j, \alpha^2_j$. Then, summing the common component $\Phi \alpha^c_j$, the unique components $\Phi \alpha^1_j, \Phi \alpha^2_j$, and the average of the mean values, we obtain the estimated $j$th patch $x^F_j$ of the fused image. Finally, we calculate the fused image $I^F$ by averaging the estimated patches. Two rules, ‘max-abs’ and ‘average,’ are commonly used in multiscale-transform- and sparse-representation-based methods. However, in the proposed method, we use the JSM to accurately calculate the complementary information. Because our approach can directly determine the fused image from the separated complementary information, it can avoid problems associated with rule design. In addition, by accurately separating the complementary information, we can completely merge the information from different source images into a single image.

We compared our JSM-based approach with four multiscale-transform-based methods, including DWT, stationary-wavelet transform, dual-tree complex-wavelet transform, and nonsubsampled-contourlet transform. In addition, we compared the proposed method with a sparse-representation-based technique known as simultaneous orthogonal matching pursuit. Figure 2 shows the visible and IR source images and the fused images produced by the different methods. The visible image clearly shows the road and signboard in the background, while the IR image depicts the outlines of the pedestrian and cars: see Figure 2(a) and (b). On visual inspection, we observe that the proposed JSM-based method produces fewer black shadows over the

Continued on next page
signboard than the other techniques: see Figure 2(c–h). Furthermore, the JSM-based method produces a smaller region of black artifacts around the pedestrian than the other methods—see Figure 2(c–h)—effectively preserving complementary information. The $Q^{AB/F}$ index, which is used to objectively assess fused images, also shows that the proposed method generates the best results.

Figure 3 presents additional results for fluorodeoxyglucose positron emission tomography (PET-FDG) and T1-weighted MR. Treating the PET-FDG image as colored, we transform it into intensity-hue-saturation (IHS) space. Then we combine the intensity component of the PET-FDG image with the T1-weighted MR image to create a new intensity component. We obtain the final results through the inverse IHS transform of the new intensity component and the hue and saturation components of the PET-FDG image into red-green-blue space. As shown in Figure 3, our proposed method preserves details from T1-weighted MR more effectively than other approaches.

In summary, we have described a JSM-based method that is superior to existing multiscale-transform- and sparse-representation-based methods. The new technique preserves details more effectively by accurately extracting and integrating complementary information. The experimental results suggest that our approach is applicable to monitoring safety in cities and making medical diagnoses. Future work will focus on developing a technique based on the sparse-representation structure.

Author Information

Haitao Yin and Shutao Li
Hunan University
Changsha, China

Haitao Yin received a BS and MS in applied mathematics from the College of Mathematics and Econometrics (2007 and 2009, respectively). He is currently pursuing a PhD in the College of Electrical and Information Engineering. His research interests include image processing and sparse representation.

Shutao Li received a BS, MS, and PhD in electrical engineering (1995, 1997, and 2001, respectively). He is currently a full professor at the College of Electrical and Information Engineering. His professional interests include information fusion, pattern recognition, and image processing.

References