COUPLED FEATURE LEARNING VIA STRUCTURED CONVOLUTIONAL SPARSE CODING FOR MULTIMODAL IMAGE FUSION

Farshad G. Veshki and Sergiy A. Vorobyov

Department of Signal Processing and Acoustics, Aalto University, Espoo, Finland

ABSTRACT

A novel method for learning correlated features in multimodal images based on convolutional sparse coding with applications to image fusion is presented. In particular, the correlated features are captured as coupled filters in convolutional dictionaries. At the same time, the shared and independent features are approximated using separate convolutional sparse codes and a common dictionary. The resulting optimization problem is addressed using alternating direction method of multipliers. The coupled filters are fused based on a maximum-variance rule, and a maximum-absolute-value rule is used to fuse the sparse codes. The proposed method does not entail any prelearning stage. The experimental evaluations using medical and infrared-visible image datasets demonstrate the superiority of our method compared to state-of-the-art algorithms in terms of preserving the details and local intensities as well as improving objective metrics.

***Index Terms*—** Multimodal image fusion, convolutional sparse coding, structured dictionary learning.

1. INTRODUCTION

Multimodal image fusion aims at merging the information from multiple images acquired using different imaging modalities into a single image, without introducing noise or artifacts [1, 2]. For instance, in medical image fusion, different information about the anatomies of tissues or the levels of biological activities captured using various medical imaging modalities are aggregated in a single fused image [1]. In surveillance applications, combining the visual information in optical images and the thermal information captured using infrared imaging techniques yield more informative images, and has applications, for example, in night vision [2].

A common approach for addressing the multimodal image fusion problem is to decompose the input images into multiscale or morphologically distinct components. This is usually done by employing deterministic mathematical models such as multiscale transforms [3–5]. Other techniques used for a similar purpose include subspace learning [6], dictionary learning [7, 8], and deep learning [9, 10]. An assumption made by all aforementioned fusion techniques is that the features (components) with similar structural properties convey the same type of information. Therefore, they are appropriate for fusion. However, the multimodal images may not obey this assumption. For example, in medical imaging, computed tomography (CT) captures hard tissues and structures (e.g., bones and implants) with a higher resolution, while using magnetic resonance (MR) imaging, the details of soft tissues (e.g., fat and bone marrow) are reflected more effectively [1]. In infrared-visible images, the details in each input image provide different types of information. Thus, a fusion based on the similarity of structural properties can lead to the omission of important information.

In a recent work [11], we demonstrated that the fusion performance can be considerably improved by replacing the conventional deterministic feature-extraction techniques with a data-driven approach for extracting correlated features in multimodal images. Specifically, a method based on coupled dictionary learning [12] and a Pearson correlation constraint has been developed to decompose the multimodal images into their correlated and independent components. In particular, the correlated features have been captured as pairs of atoms in the coupled dictionaries. Then, the fusion is performed using the most significant representations of the coupled atoms. Since the information in the independent components is specific to each modality, these components are transferred to the fused image directly. This approach has shown to be superior in terms of preserving important information while yielding an improved contrast resolution [11].

In this paper, we present a coupled feature learning (CFL) method based on convolutional sparse coding (CSC) and dictionary learning. CSC incorporates a global single-valued model that, unlike standard sparse approximation, does not require patch extraction and enables shift-invariant dictionary learning [13]. In addition, instead of minimizing linear correlations between independent components (as in [11]), we incorporate a more general model that promotes statistical independence. We also propose novel schemes for fusion of correlated features and reconstruction of the final fused image. Experimental evaluations using multimodal medical and infrared-visible image datasets show that the proposed method significantly improves the performance of state-of-the-art multimodal fusion techniques. A MATLAB implementation of our fusion method is available at `https://github.com/FarshadGVeshki/ConvCFL-MMIF`.

2. CONVOLUTIONAL COUPLED FEATURE LEARNING

The proposed model decomposes n input multimodal images $\boldsymbol{s}^i \in \mathbb{R}^N$, $i = 1, \ldots, n$, where N is the number of pixels in the images, into their correlated, shared and independent components. For simplicity of notations, we use one-dimensional arrays to represent the images. Generalization to multi-dimensional arrays is mathematically straightforward.

2.1. Problem Formulation

The correlated components are captured using a set of common sparse feature maps $\boldsymbol{\Gamma} \in \mathbb{R}^{N\times K}$ and coupled convolutional dictionaries $\boldsymbol{D}^i \in \mathbb{R}^{M\times K}$, $i = 1, \ldots, n$. The shared and independent components are represented using a common dictionary $\boldsymbol{C} \in \mathbb{R}^{M\times L}$ and separate sparse feature maps $\boldsymbol{X}^i \in \mathbb{R}^{N\times L}$, $i = 1, \ldots, n$. The decomposition problem can then be formulated as the following optimization problem

$$\underset{\substack{\{\boldsymbol{D}^i\}_{i=1}^n, \boldsymbol{C},\\ \{\boldsymbol{X}^i\}_{i=1}^n, \boldsymbol{\Gamma}}}{\text{minimize}} \frac{1}{2}\sum_{i=1}^{n}\Big\|\sum_{k=1}^{K}\boldsymbol{D}_k^i * \boldsymbol{\Gamma}_k + \sum_{l=1}^{L}\boldsymbol{C}_l * \boldsymbol{X}_l^i - \boldsymbol{s}^i\Big\|_2^2 + \lambda_1\sum_{k=1}^{K}\|\boldsymbol{\Gamma}_k\|_1 + \lambda_2\sum_{i=1}^{n}\sum_{l=1}^{L}\|\boldsymbol{X}_k^i\|_1 \quad \text{s.t.} \quad \boldsymbol{C}_l, \boldsymbol{D}_k^i, \in \boldsymbol{\mathcal{D}} \quad \forall k, l, i, \quad (1)$$

where $\boldsymbol{\mathcal{D}} = \{\boldsymbol{d} \in \mathbb{R}^M | \|\boldsymbol{d}\|_2 \leq 1\}$ is the set of dictionary filters and $\lambda_1 > 0$ and $\lambda_2 > 0$ are regularization parameters. Subscripts are used to denote the columns of matrices.

The overlapping nonzero entries in $\{\boldsymbol{X}^i\}_{i=1}^n$ indicate that one of the dictionary filters $\{\boldsymbol{C}_l\}_{l=1}^L$ is used for approximation of multiple images at the same location, thus, it represents a shared feature. In addition, when only one of the entries in $\{\boldsymbol{X}^i\}_{i=1}^n$ is nonzero at one location, it means that one of the filters in $\{\boldsymbol{C}_l\}_{l=1}^L$ is used for only one source image. Thus, it represents an independent feature. Note that the convolutional filters are assumed to be statistically independent. As first demonstrated in [14], dictionary learning promotes statistical independence. The proof relies on the fact that accurate sparse codes preserve the information (joint entropy) in the source signal. Moreover, the sparsity regularization minimizes the entropy in each of the sparse codes (simply by maximizing the probability of one event (being zero) and minimizing the probability of all other events (being nonzero)). Therefore, by enforcing the equality of the joint entropy and the sum of the entropies of the individual sparse codes, sparse dictionary learning promotes statistical independence.

2.2. Optimization

Problem (1) is typically solved by alternating between minimization over the sparse codes and the dictionary filters. Since we address both steps in Fourier domain (using [15]), we first zero-pad all of the dictionary filters to the size of the sparse coefficient maps ($\mathbb{R}^N$).

2.2.1. Sparse Coding Step

Using the consensus ADMM method [16], (1) can be addressed with respect to the sparse feature maps $\{\boldsymbol{\Gamma}, \{\boldsymbol{X}^i\}_{i=1}^n\}$ by solving the following optimization problem

$$\underset{\{\boldsymbol{X}^i\}_{i=1}^n, \boldsymbol{\Gamma}}{\text{minimize}} \frac{1}{2}\sum_{i=1}^{n}\Big\|\sum_{k=1}^{K}\boldsymbol{D}_k^i * \boldsymbol{\Theta}_k^i + \sum_{l=1}^{L}\boldsymbol{C}_l * \boldsymbol{Y}_l^i - \boldsymbol{s}^i\Big\|_2^2 + \lambda_1\sum_{k=1}^{K}\|\boldsymbol{\Gamma}_k\|_1 + \lambda_2\sum_{i=1}^{n}\sum_{l=1}^{L}\|\boldsymbol{X}_k^i\|_1 \quad \text{s.t.} \quad \boldsymbol{\Gamma} = \boldsymbol{\Theta}^i,\ \boldsymbol{X}^i = \boldsymbol{Y}^i \quad i = 1, \ldots, n.$$

Using scaled Lagrangian multipliers $\boldsymbol{U}^i \in \mathbb{R}^{N\times K}$ and $\boldsymbol{V}^i \in \mathbb{R}^{N\times L}$,$i = 1, \ldots, n$, the augmented Lagrangian is written as

$$\frac{1}{2}\sum_{i=1}^{n}\Big\|\sum_{k=1}^{K}\boldsymbol{D}_k^i * \boldsymbol{\Theta}_k^i + \sum_{l=1}^{L}\boldsymbol{C}_l * \boldsymbol{Y}_l^i - \boldsymbol{s}^i\Big\|_2^2 + \lambda_1\sum_{k=1}^{K}\|\boldsymbol{\Gamma}_k\|_1 + \lambda_2\sum_{i=1}^{n}\sum_{l=1}^{L}\|\boldsymbol{X}_k^i\|_1 + \frac{\rho}{2}\sum_{i=1}^{n}\Big(\sum_{k=1}^{K}\|\boldsymbol{\Theta}_k^i - \boldsymbol{\Gamma}_k + \boldsymbol{U}_k^i\|_2^2 + \sum_{l=1}^{L}\|\boldsymbol{Y}_l^i - \boldsymbol{X}_l^i + \boldsymbol{V}_l^i\|_2^2\Big), \quad (2)$$

where $\rho > 0$ is the penalty parameter. The ADMM iterations consist of minimizing (2) alternatively with respect to $\{\boldsymbol{\Theta}^i, \boldsymbol{Y}^i\}_{i=1}^n$, $\{\boldsymbol{\Gamma}, \{\boldsymbol{X}^i\}_{i=1}^n\}$ and $\{\boldsymbol{U}^i, \boldsymbol{V}^i\}_{i=1}^n$. The details of each subproblem are explained in the following. Denoting $\boldsymbol{Z}^i = \{\boldsymbol{\Theta}^i, \boldsymbol{Y}^i\}$, $\boldsymbol{F}^i = \{\boldsymbol{D}^i, \boldsymbol{C}\}$ and $\boldsymbol{W}^i = \{\boldsymbol{\Gamma} - \boldsymbol{U}^i, \boldsymbol{X}^i - \boldsymbol{V}^i\}$, we can update $\{\boldsymbol{\Theta}^i, \boldsymbol{Y}^i\}_{i=1}^n$ by solving n parallel optimization problems, which can be written as follows

$$(\boldsymbol{Z}^i)^+ = \underset{\boldsymbol{Z}^i}{\text{argmin}}\ \frac{1}{2}\Big\|\sum_{p=1}^{P}\boldsymbol{F}_p^i * \boldsymbol{Z}_p^i - \boldsymbol{s}^i\Big\|_2^2 + \frac{\rho}{2}\sum_{p=1}^{P}\|\boldsymbol{Z}_p^i - \boldsymbol{W}_p^i\|_2^2, \quad (3)$$

where $P = K+L$ and $(\cdot)^+$ denotes the updated variables. The problem in (3) is a standard convolutional fitting problem that can be addressed using available CSC methods (e.g., [15]).

Updating $\{\boldsymbol{\Gamma}, \{\boldsymbol{X}^i\}_{i=1}^n\}$ can be efficiently addressed in an elementwise manner using the shrinkage operator $\mathcal{S}_\kappa(a) = \text{sign}(a)\max(0, |a| - \kappa)$. The updates are written as follows

$$\boldsymbol{\Gamma}^+ = \mathcal{S}_{\lambda_1/\rho}\Big(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Theta}^i + \boldsymbol{U}^i\Big),\ (\boldsymbol{X}^i)^+ = \mathcal{S}_{\lambda_2/\rho}(\boldsymbol{Y}^i + \boldsymbol{V}^i),\ i = 1, \ldots, n.$$

Finally, the updates for the scaled Lagrangian variables $\{\boldsymbol{U}^i, \boldsymbol{V}^i\}_{i=1}^n$ are given as

$$(\boldsymbol{U}^i)^+ = \boldsymbol{\Theta}^i - \boldsymbol{\Gamma} + \boldsymbol{U}^i, \quad (\boldsymbol{V}^i)^+ = \boldsymbol{Y}^i - \boldsymbol{X}^i + \boldsymbol{V}^i, \quad i = 1, \ldots, n.$$

2.2.2. Dictionary Update Step

Using the consensus ADMM, (1) can be reformulated with respect to the dictionary filters $\{\boldsymbol{C}, \{\boldsymbol{D}^i\}_{i=1}^n\}$ as

$$\underset{\{\boldsymbol{D}^i\}_{i=1}^n, \boldsymbol{C}}{\text{minimize}} \frac{1}{2}\sum_{i=1}^{n}\Big\|\sum_{k=1}^{K}\boldsymbol{G}_k^i * \boldsymbol{\Gamma}_k + \sum_{l=1}^{L}\boldsymbol{H}_l^i * \boldsymbol{X}^i - \boldsymbol{s}^i\Big\|_2^2 + \boldsymbol{\Omega}\Big(\{\boldsymbol{C}, \{\boldsymbol{D}^i\}_{i=1}^n\}\Big) \quad \text{s.t.} \quad \boldsymbol{C} = \boldsymbol{H}^i, \quad \boldsymbol{D}^i = \boldsymbol{G}^i, \quad i = 1, \ldots, n$$

where $\boldsymbol{\Omega}(\cdot)$ is an indicator function of the constraint set in (1). The augmented Lagrangian is written as follows

$$\frac{1}{2}\sum_{i=1}^{n}\Big\|\sum_{k=1}^{K}\boldsymbol{G}_k^i * \boldsymbol{\Gamma}_k + \sum_{l=1}^{L}\boldsymbol{H}_l^i * \boldsymbol{X}^i - \boldsymbol{s}^i\Big\|_2^2 + \boldsymbol{\Omega}(\{\boldsymbol{C}, \{\boldsymbol{D}^i\}_{i=1}^n\}) + \frac{\sigma}{2}\sum_{i=1}^{n}\Big(\sum_{k=1}^{K}\|\boldsymbol{G}_k^i - \boldsymbol{D}_k^i + \boldsymbol{R}_k^i\|_2^2 + \sum_{l=1}^{L}\|\boldsymbol{H}_l^i - \boldsymbol{C}_l + \boldsymbol{T}_l^i\|_2^2\Big), \quad (4)$$

where $\boldsymbol{R}^i \in \mathbb{R}^{N\times K}$ and $\boldsymbol{T}^i \in \mathbb{R}^{N\times L}$,$i = 1,\ldots,n$, are scaled Lagrangian variables. The ADMM iterations consist of minimizing (4) alternatively with respect to $\{\boldsymbol{G}^i, \boldsymbol{H}^i\}_{i=1}^n$, $\{\boldsymbol{C}, \{\boldsymbol{D}^i\}_{i=1}^n\}$ and $\{\boldsymbol{R}^i, \boldsymbol{T}^i\}_{i=1}^n$.

Indeed, denote $\boldsymbol{E}^i = \{\boldsymbol{G}^i, \boldsymbol{H}^i\}, \boldsymbol{S}^i = \{\boldsymbol{\Gamma}, \boldsymbol{X}^i\}$ and $\boldsymbol{Q}^i = \{\boldsymbol{D}^i - \boldsymbol{R}^i, \boldsymbol{C} - \boldsymbol{T}^i\}, i = 1,\ldots,n,$. Then updating $\{\boldsymbol{G}^i, \boldsymbol{H}^i\}_{i=1}^n$ can be addressed by solving n parallel optimization problems

$$(\boldsymbol{E}^i)^+ = \underset{\boldsymbol{E}^i}{\operatorname{argmin}} \frac{1}{2}\|\sum_{p=1}^{P} \boldsymbol{E}_p^i * \boldsymbol{S}_p^i - \boldsymbol{s}^i\|_2^2 + \frac{\sigma}{2}\sum_{p=1}^{P}\|\boldsymbol{E}_p^i - \boldsymbol{Q}_p^i\|_2^2. \tag{5}$$

Problem (5) is similar to (3) and can be efficiently addressed using available CSC methods (e.g., [15]).

Updating $\{\boldsymbol{C}, \{\boldsymbol{D}^i\}_{i=1}^n\}$ is performed as

$$(\boldsymbol{D}^i)^+ = \operatorname{proj}_{\boldsymbol{\mathcal{D}}}(\boldsymbol{G}^i + \boldsymbol{R}^i), i = 1,\ldots,n, \quad \boldsymbol{C}^+ = \operatorname{proj}_{\boldsymbol{\mathcal{D}}}\Big(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{H}^i + \boldsymbol{T}^i\Big),$$

where $\operatorname{proj}_{\boldsymbol{\mathcal{D}}}(\cdot)$ denotes the orthogonal projection onto the set $\boldsymbol{\mathcal{D}}$. This can be done by mapping the entries outside the constraint support to zero and then projecting the filters on the unit-ball.

The updates for scaled Lagrangian variables are given as

$$(\boldsymbol{R}^i)^+ = \boldsymbol{G}^i - \boldsymbol{D}^i + \boldsymbol{R}^i, \quad (\boldsymbol{T}^i)^+ = \boldsymbol{H}^i - \boldsymbol{C} + \boldsymbol{T}^i. \quad i = 1,\ldots,n.$$

We perform the sparse coding and the dictionary update steps in an interleaved manner (one iteration of each step is executed before passing the variables to the next). The updated ADMM variables (auxiliary variables and scaled Lagrangian multipliers) are used to initialize the next iteration.

2.3. Projection on the Sparse Support

After the convolutional CFL stage, we can still significantly improve the approximation accuracy by orthogonalizing the residuals on the supports (the set of indices of nonzero entries) of the sparse coefficient maps. For this purpose, we use a gradient descent (GD) approach. Based on the convolution theorem, the GD iterations are found as

$$[\boldsymbol{\Gamma}_k^+]_{\mathcal{S}(\boldsymbol{\Gamma}_k)} = [\boldsymbol{\Gamma}_k]_{\mathcal{S}(\boldsymbol{\Gamma}_k)} - \alpha\Big[\mathrm{DFT}^{-1}\big(\sum_{i=1}^{n}\bar{\hat{\boldsymbol{D}}}_k^i \odot \hat{\boldsymbol{r}}^i\big)\Big]_{\mathcal{S}(\boldsymbol{\Gamma}_k)}, \forall k,$$

$$[(\boldsymbol{X}_l^i)^+]_{\mathcal{S}(\boldsymbol{X}_l^i)} = [\boldsymbol{X}_l^i]_{\mathcal{S}(\boldsymbol{X}_l^i)} - \alpha\Big[\mathrm{DFT}^{-1}\big(\bar{\hat{\boldsymbol{C}}}_l^i \odot \hat{\boldsymbol{r}}^i\big)\Big]_{\mathcal{S}(\boldsymbol{X}_l^i)}, \forall l, i,$$

where $\hat{(\cdot)}$ denotes the discrete Fourier transform and $\mathrm{DFT}^{-1}(\cdot)$ represents its inverse, $\bar{(\cdot)}$ denotes the complex-conjugate, $\odot$ is the elementwise multiplication and operator $\mathcal{S}(\cdot)$ returns the support of an array. In addition, α is the stepsize and $\boldsymbol{r}^i$ represents the residuals associated with $\boldsymbol{s}^i$, that is

$$\boldsymbol{r}^i = \sum_{k=1}^{K}\boldsymbol{D}_k^i * \boldsymbol{\Gamma}_k + \sum_{l=1}^{L}\boldsymbol{C}_l * \boldsymbol{X}_l^i - \boldsymbol{s}^i, i = 1,\ldots,n.$$

3. MULTIMODAL IMAGE FUSION ALGORITHM

In this section, the steps of the proposed fusion method are explained. Note that the images are considered as one-dimensional arrays. The elementwise operations are applied to all pixels.

3.1. Low-pass Filtering

The input images are first decomposed into base-layers $\{\boldsymbol{s}_b^i\}_{i=1}^n$ and details-layers $\{\boldsymbol{s}_d^i\}_{i=1}^n$ using low-pass filtering. This is done using the *lowpass* function from the SPORCO library [17] (the regularization parameter is set to 10).

3.2. Fusion of the Details-layers

The details-layers $\{\boldsymbol{s}_d^i\}_{i=1}^n$ are decomposed into the correlated, shared and independent components using the convolutional CFL method explained in Section 2.

3.2.1. Fusion of Coupled Features

The coupled features are fused based on the highest visual significance, which can be measured, for example, using *variance* (denoted as $\mathrm{var}(\cdot)$). This can be formulated as follows

$$\boldsymbol{D}_k^F = \boldsymbol{D}_k^{i^*}, \quad i^* = \underset{i=1,\ldots,n}{\operatorname{argmax}}\Big(\mathrm{var}(\boldsymbol{D}_k^i)\Big), \quad k = 1,\ldots,K,$$

where $\boldsymbol{D}^F$ is the dictionary of fused coupled features.

3.2.2. Fusion of Shared and Independent Components

The fusion of shared and independent component $\boldsymbol{X}^F$ is found by combining the redundant sparse codes $\{\boldsymbol{X}^i\}_{i=1}^n$ using maximum-absolute-value rule. This can be written as

$$\boldsymbol{X}_l^F(j) = \boldsymbol{X}_l^{i^*}(j), i^* = \underset{i=1,\ldots,n}{\operatorname{argmax}}\Big(|(\boldsymbol{X}_l^i(j)|\Big), j = 1,\ldots,N, l = 1,\ldots,L.$$

This allows to transfer the independent features along with the shared features with the most significant representation coefficients into the fused image.

The fused details-layer $\boldsymbol{s}_d^F$ is then reconstructed using

$$\boldsymbol{s}_d^F = \sum_{k=1}^{K}\boldsymbol{D}_k^F * \boldsymbol{\Gamma}_k + \sum_{l=1}^{L}\boldsymbol{C}_l * \boldsymbol{X}_l^F.$$

3.3. Fusion of the Base-layers

We form two images $\boldsymbol{s}_b^{max}$ and $\boldsymbol{s}_b^{min}$ representing the maximum and the minimum allowed local intensities, using

$$\boldsymbol{s}_b^{max} = \max_{i=1,\ldots,n}(\boldsymbol{s}_b^i), \quad \boldsymbol{s}_b^{min} = \omega\Big(\max_{i=1,\ldots,n}(\boldsymbol{s}_b^i)\Big) + (1-\omega)\Big(\min_{i=1,\ldots,n}(\boldsymbol{s}_b^i)\Big),$$

where $0 \le \omega \le 1$, and $\max(\cdot)$ and $\min(\cdot)$ are the elementwise maximum and minimum operators, respectively.

It is favorable to incorporate $\boldsymbol{s}_b^{max}$ into the final fused image. However, this can cause a loss of information due to the limited range (0 to 1) of the standard images. To achieve a compromise between contrast resolutions and local intensities, we propose the following approach. First, the difference between the local maximum and minimum intensities (local variations) of $\boldsymbol{s}_d^F$ (for example, in a 3×3 neighborhood) is stored in $\boldsymbol{s}_d^v$. Then the fused base-layer $\boldsymbol{s}_b^F$ is computed as

$$\boldsymbol{s}_b^F(j) = \begin{cases} \boldsymbol{s}_b^{max}(j), & \text{if } \boldsymbol{s}_d^v(j) \leq 1 - \boldsymbol{s}_b^{max}(j) \\ \boldsymbol{s}_b^{min}(j), & \text{if } \boldsymbol{s}_d^v(j) \geq 1 - \boldsymbol{s}_b^{min}(j) \\ 1 - \boldsymbol{s}_b^v(j), & \text{if otherwise} \end{cases}, \quad j = 1, \ldots, N.$$

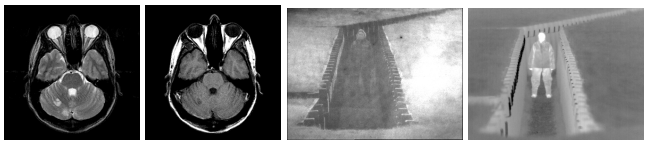
A Gaussian filter may be used to smooth $\boldsymbol{s}_b^F$ so that discontinuities are not introduced.

The final fused image $\boldsymbol{s}^F$ is then reconstructed as

$$\boldsymbol{s}^F = \boldsymbol{s}_b^F + \boldsymbol{s}_d^F.$$

4. EXPERIMENTAL RESULTS

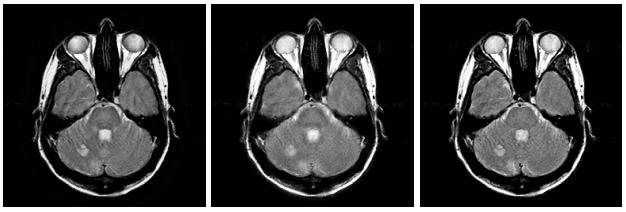
We compare our method to four recent multimodal fusion methods both visually and using objective evaluation metrics. We use two medical image fusion methods: a method based on the non-subsampled shearlet transform (NSST) [4] and a method based on Laplacian redecomposition (LRD) [5]. We also use two infrared-visible image fusion method: a method that incorporates a hierarchical Bayesian model (Bayes) [18] and a method based on deep learning (Resnet) [9]. The multimodal medical image dataset consists of 20 pairs of images collected from [19], and the infrared-visible image dataset includes 21 pairs of images taken from `https://github.com/hli1221/imagefusion_resnet50/tree/master/IV_images`. Four metrics are used for objective evaluations, the objective image fusion performance measure $\mathrm{Q_{AB/F}}$ [20], the information measure for performance of image fusion $\mathrm{Q_{IM}}$ [21], spatial frequency (SF) [22] and the structural similarity index (SSIM) [23]. The algorithm parameters are λ_1=λ_2= 0.01, K=8, L=12, ρ=σ=10, α=0.01 and $\omega = 0.9$. Moreover, we use 150 ADMM iterations, 100 GD iterations and 8×8 filters (M in two-dimensional case).



(a) MR-T1 (b) MR-T2 (c) Visible (d) Infrared

Fig. 1: Examples of multimodal images.

Fig. 1 shows a pair of images from each dataset used. The results obtained using different methods are shown in Fig. 2. Table 1 compares the average results for objective evaluation metrics for each dataset. The results show that the LRD method leads to low contrast-resolutions, which is reflected in very low SF and $\mathrm{Q_{AB/F}}$ values for this method. NSST also loses/blurs high-resolution information, while this information is well preserved using our method (see Figs. 2a and 2b, for example). Results obtained using Resnet and Bayes show an inferior fusion of local intensities, which results in low visibility of the details in the fused images (see Figs. 2d and 2e, for example). Overall, the proposed method results in the best performance in terms of the fusion of the high-resolution information as well as the local intensities (for example, see Figs. 2c and 2f). These observations can be validated by the objective evaluation results in Table 1, where our method obtains the best results in all cases.



(a) NSST (b) LRD (c) proposed

(d) Bayes (e) Resnet (f) proposed

Fig. 2: The fusion results for the multimodal images in Fig. 1 using different methods.

Metrics	Medical			Infrared-Visible		
	NSST	LRD	proposed	Bayes	Resnet	proposed
$\mathrm{Q_{AB/F}}$	0.5646	0.5278	**0.5667**	0.4561	0.3520	**0.4941**
$\mathrm{Q_{IM}}$	0.6778	0.7341	**0.7424**	0.4044	0.3167	**0.4311**
SF	30.02	28.93	**31.93**	7.63	6.13	**11.11**
SSIM	0.5501	0.6601	**0.7047**	0.5040	0.5072	**0.5121**

Table 1: Average objective evaluation results for each dataset using different methods. The Best results are shown in bold.

5. CONCLUSION

A novel multimodal image fusion method based on convolutional sparse coding has been developed. A convolutional coupled feature learning algorithm has been proposed for the decomposition of multimodal images into correlated, shared, and independent features. Appropriate schemes have been proposed for the fusion of extracted features and reconstruction of the final image. The experimental results show significant improvements by the proposed method compared to the state-of-the-art multimodal image fusion methods.

6. REFERENCES

[1] B. Huang, F. Yang, M. Yin, X. Mo, and C. Zhong, "A review of multimodal medical image fusion techniques," *Comput. Math. Methods. Med.*, vol. 2020, 2020.

[2] X. Zhang, P. Ye, H. Leung, K. Gong, and G. Xiao, "Object fusion tracking based on visible and infrared images: A comprehensive review," *Inf. Fusion*, vol. 63, pp. 166–187, 2020.

[3] G. Li, Y. Lin, and X. Qu, "An infrared and visible image fusion method based on multi-scale transformation and norm optimization," *Inf. Fusion*, vol. 71, pp. 109–129, 2021.

[4] W. Tan, P. Tiwari H. M. Pandey, C. Moreira, and A. K. Jaiswal, "Multimodal medical image fusion algorithm in the era of big data," *Neural Comput. Appl.*, 2020.

[5] X. Li, X. Guo, P. Han, X. Wang, H. Li, and T. Luo, "Laplacian re-decomposition for multimodal medical image fusion," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 9, pp. 6880–6890, 2020.

[6] Y. Yang, S. Cao, S. Huang, and W. Wan, "Multimodal medical image fusion based on weighted local energy matching measurement and improved spatial frequency," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–16, 2021.

[7] Y. Liu, X. Chen, R. K. Ward, and Z. J. Wang, "Medical image fusion via convolutional sparsity based morphological component analysis," *IEEE Signal Process. Lett.*, vol. 26, no. 3, pp. 485–489, 2019.

[8] H. Li, X. He, D. Tao, Y. Tang, and R. Wang, "Joint medical image fusion, denoising and enhancement via discriminative low-rank sparse dictionaries learning," *Pattern Recognit.*, vol. 79, pp. 130–146, 2018.

[9] H. Li, X. Wu, and T. S. Durrani, "Infrared and visible image fusion with resnet and zero-phase component analysis," *Infrared Physics and Technology*, vol. 102, 2019.

[10] Z. Wang, X. Li, H. Duan, Y. Su, X. Zhang, and X. Guan, "Medical image fusion based on convolutional neural networks and non-subsampled contourlet transform," *Expert Systems with Applications*, vol. 171, 2021.

[11] F. G. Veshki, N. Ouzir, S. A. Vorobyov, and E. Ollila, "Coupled feature learning for multimodal medical image fusion," *arXiv:2102.08641*, 2021.

[12] F. G. Veshki and S. A. Vorobyov, "An efficient coupled dictionary learning method," *IEEE Signal Process. Lett.*, vol. 26, no. 10, pp. 1441–1445, 2019.

[13] C. Garcia-Cardona and B. Wohlberg, "Convolutional dictionary learning: A comparative review and new algorithms," *IEEE Trans. Comput. Imaging*, vol. 4, no. 3, pp. 366–381, 2018.

[14] B. Olshausen and D. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 381, pp. 607–609, 1996.

[15] F. G. Veshki and S. A. Vorobyov, "Efficient ADMM-based algorithms for convolutional sparse coding," *IEEE Signal Process. Lett.*, vol. 29, pp. 389–393, 2022.

[16] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.

[17] B. Wohlberg, "SParse Optimization Research COde (SPORCO)," Software library available from `http://purl.org/brendt/software/sporco`, 2017.

[18] Z. Zhao, S. Xu, C. Zhang, J. Liu, and J. Zhang, "Bayesian fusion for infrared and visible images," *Signal Process.*, vol. 177, pp. 1–12, 2020.

[19] Harvard Medical School, "The Whole Brain Atlas ," `http://www.med.harvard.edu/AANLIB/`, [Online; accessed 16-sep-2021].

[20] C. Xydeas and V. Petrovic, "Objective image fusion performance measure," *Electron. Lett.*, vol. 36, no. 4, pp. 308–309, 2000.

[21] G. Qu, D. Zhang, and P. Yan, "Information measure for performance of image fusion," *Electronics Letters*, vol. 38, no. 7, pp. 313–315, 2002.

[22] A.M. Eskicioglu and P.S. Fisher, "Image quality measures and their performance," *IEEE Transactions on Communications*, vol. 43, no. 12, pp. 2959–2965, 1995.

[23] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.