Journal of Engineering Science and Technology
Vol. 14, No. 3 (2019) 1231 - 1247

SPARSE BASED IMAGE FUSION USING COMPACT SUB-DICTIONARIES

ASHWINI K.*, AMUTHA R.

SSN College of Engineering, OMR, Kalavakkam, Tamil Nadu, India
*Corresponding Author: ashwini88.k@gmail.com

Abstract

Image fusion schemes are desirable to obtain a high-quality image by integrating complementary information from multiple source images. The main aim of this paper is to propose a novel image fusion technique that provides a highly informative image, which is useful in various applications like computer vision, medical diagnosis, remote sensing, etc. Traditional Sparse Representation (SR) based fusion method makes use of a single highly redundant dictionary for image fusion. This increases complexity and may also lead to visual artefacts in the fused image. Fusion scheme using dictionary-based sparse representation is proposed in this paper. A large number of image patches are pre-classified based on projections using R-Transform and a set of compact sub-dictionaries are learnt from them. At the fusion stage, one of the sub-dictionaries is chosen to fuse the given set of source images. Quantitative and Qualitative evaluation of the proposed fusion scheme on multi-focus and multi-modal images shows the superiority of the proposed scheme over other existing fusion algorithms.

Keywords: Image fusion, R-transform, Sparse representation, Sub-dictionary.

1. Introduction

Images obtained from a single camera or sensor does not always provide sufficient information needed for machine or visual perception. Image fusion techniques are often employed in cases where the highly informative image is desirable. Image fusion is a process where relevant and complementary information from two or more source images are combined into a single image [1].

Fusion techniques are broadly classified into spatial domain and transform domain techniques. In the spatial domain method, the fusion process is employed directly on image pixels or on image regions of the source images. The averaging method by Bavachan and Krishnan [2] and Principal Component Analysis (PCA) [3] method fall under this category. Zhan et al. [4] proposed image fusion based on Phase Congruency (PC). Using PC, authors were able to extract local and sharp changes in the intensity of images. They have used PC as the focus measure for computing the focus quality. The author claims that the main advantage of the PC is that it is invariant to noise. Zhan et al. [5] proposed a novel explicit image fusion method using the energy of Laplacian. Authors use the energy of Laplacian of input images to find out the portions of the input image that are in better focus. In order to reduce block effects in the final fused image, they have also used edge preserving guided image filter.

Spatial domain based fusion schemes often cause blocking artefacts or spatial distortions in the fused image. In transform domain techniques, pre-defined fusion rule is applied on transformed coefficients of source images to get the fused image. Common transform domain techniques are fusion using pyramids [6], fusion using wavelet transform [7], curvelet transform based fusion techniques [8], Dual-Tree Complex Wavelet Transform techniques (DTCWT) [9], Non Subsampled Contourlet Transform (NSCT) based fusion techniques [10], Fusion using DCT [11, 12], etc. Recently, Sparse Representations (SR) of signals have been widely used in many applications such as denoising [13], super-resolution [14], fusion [15-18], etc. Sparse representation based fusion method performs fusion in the sparser domain and it can also be classified as one of the transform based fusion technique. In the fusion scheme proposed by Yang and Li [15], a fixed dictionary was used. But this type of dictionary cannot represent any arbitrary set of signals. Elad and Yavneh [19], proposed an SR based fusion scheme, in which, a highly redundant dictionary is used for image reconstruction. This leads to computational inefficiency and visual artefacts are produced in the fused image. Zhu et. al [20] proposed a sparse based multi-modality fusion scheme. In their proposed fusion method, multi-modality images are decomposed into cartoon and texture components. Decomposed Cartoon and texture components are fused separately using energy based and sparse based fusion rules. Fused cartoon and texture components are finally integrated to obtain the final fused image. Zong and Qiu [21] proposed a medical image fusion algorithm using an online dictionary. Least angle regression (LARS) algorithm was used by the authors to sparsely code image patches and the sparse coefficients were then combined with the "choose-max" fusion rule. Finally, the fused image obtained was reconstructed by them using combined sparse coefficients and the corresponding dictionary.

Liu and Wang [22], image patches of high quality are used to learn a set of dictionaries. At the fusion stage, source images that are represented as image patches are fused by choosing one sub-dictionary for each patch. Choosing separate

dictionary for each patch of source image increases complexity as well as the computation time. To reduce the computational cost and to obtain a highly informative fused image, sparse representation based fusion scheme using compact sub-dictionaries is proposed. Image patches are initially classified based on their orientation information and sub-dictionaries are learned from them. Among the learned sub-dictionaries, a single appropriate sub-dictionary is chosen to fuse a given source image pair. In order to prove the effectiveness of the proposed fusion scheme, both quantitative and qualitative analysis are carried out and the results are compared with existing state-of-the-art fusion schemes.

The rest of the paper is structured as follows. Review on sparse representation theory is given in Section 2. Section 3 explains the proposed fusion scheme. Experimental results along with qualitative and quantitative analysis are presented in Section 4. Conclusions are presented in Section 5.

2. Sparse representation theory

The basic idea behind the sparse representation is that any natural signal can be represented as a linear combination of atoms from an over-complete dictionary [23]. The sparse approximation of a signal $x \in R^n$ is given by $x = D\alpha$, where $D \in R^{n \times m}$ ($m \gg n$) is an over-complete dictionary and $\alpha \in R^m$ is the sparse coefficient vector with very few non-zero elements. By solving the following optimization problem, the sparsest solution of α is obtained as follows

$$\alpha = \arg\min_{\alpha} \|\alpha\|_0 \text{ subject to } \|x - D\alpha\|_2 \leq \varepsilon \tag{1}$$

where ε is target error. $\|.\|_0$ is $l_0 - norm$. Finding an optimal solution for the above equation is a NP-hard problem and thus, an approximation technique such as pursuit algorithms are used to solve the above equation. Of many pursuit algorithms proposed in the literature, Orthogonal Matching Pursuit (OMP) [24] is widely used because of its efficiency and simplicity.

Using an appropriate dictionary is the most important aspect in sparse representation theory. Dictionary can either be a fixed or a learned one. Fixed dictionaries are obtained using analytical models such as DCT, wavelets, curvelets, etc. They are very fast to implement but are unfit to be used to represent any arbitrary family of signals. Learned dictionaries are generally constructed using a large number of the training dataset. Dictionary learning algorithm such as K-SVD [25] or MOD [26] is used on these training datasets to obtain a learned dictionary. Since this type of dictionary contains all the key directions required to represent a signal, they outperform the fixed dictionary in image reconstruction [27].

3. Proposed Fusion Method

3.1. Dictionary learning procedure

As mentioned earlier, dictionary formation is an important aspect in any sparse representation theory. The schematic representation of dictionary learning process is given in Fig. 1.

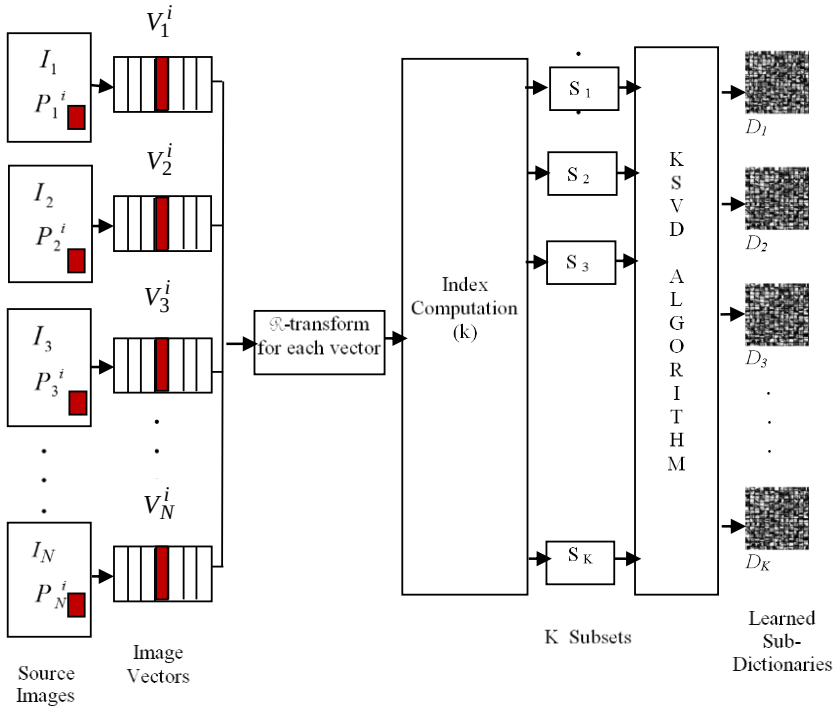


Fig. 1. Schematic representation of the dictionary learning process.

Training set for the dictionary learning process is obtained by dividing each multi-focus image into 8×8 patches with 50% overlap. The corresponding mean values are subtracted from these patches before training and are denoted as $P = \{p_1^i, p_2^2, ..., p_N^i\}$, where p_N^i denotes i^{th} patch of N^{th} image and i varies from 1 to the total number of patches in an image. The patches are then transformed into vectors $V = \{v_1^i, v_2^i, ..., v_N^i\}$ and the vectors are classified based on their projections along the lines taken at different angles. R-transform [28], an adaptation of Radon transform, is used for classification purpose.

Radon transform is a shape descriptor that provides the 2D feature of a signal. Radon transform of a signal $f(x,y)$ can be thought as a projection of signal along the given angle. It is defined as:

$$T_{R^f}^2(\delta,\theta) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,\delta(x\cos\theta + y\sin\theta - \delta)\,dxdy \quad (2)$$

where $\delta(.)$ is the Dirac delta function. δ ranges from $-\infty$ to ∞ and θ ranges from 0 to π. In order to make the radon transform invariant to translation, rotation and scaling, R-transform is given as follows

$$R_f(\theta) = \int_{-\infty}^{\infty} T_{R^f}^2(\delta,\theta)\, d\delta \tag{3}$$

where T_{R^f} is radon transform of the signal *f(x,y)*.

Using the above-defined R-transform, the vectors are classified based on their projections. For each vector, $v_N^i \in V$, R-transformed coefficient is denoted as $\{r_{N,1}^i, r_{N,2}^i, .., r_{N,K}^i\}$. $r_{N,K}^i$ is the projection of i^{th} vector in K^{th} angle for N^{th} image. The value of *K* depends upon the choice of θ. For example, θ ranging from 0° to 180° with 30° spacing gives six coefficients with *K* value ranging from 1 to 6.

The R-transform thus projects the orientations of vectors at *K* different angles and the maximum value of the transformed coefficient denotes that the vector is maximumly oriented at that angle.

The vectors in the set *V* are then classified into *K* subsets as $\{S_k \mid k=1, 2,\ldots,K\}$, k is the index of the subset to which, the vector v_N^i should be grouped. Index *k* is obtained by the finding the index of maximum transformed coefficient value using the equation given below.

$$k = \arg\max_k \left\{ r_{N,1}^i, r_{N,2}^i, \ldots, r_{N,K}^i \right\} \tag{4}$$

Once the vectors are grouped into *K* different subset, K-SVD algorithm is used to learn *K* different sub-dictionaries $\{D_1, D_2, \ldots, D_K\}$.

3.2. Fusion rule

Schematic representation of the proposed fusion method is shown in Fig. 2. Let I_A and I_B denotes source image pair to be fused and I_F denotes the final fused image. The detailed fusion scheme is explained as follows:

Step 1: Apply the sliding window with a step length of one pixel on I_A, I_B and get image patches of size 8x8. Let $\left\{P_A^i, P_B^i\right\}_{i=1}^N$ denote a set of patches in $\{I_A, I_B\}$. *N* is the total number of patches in each image.

Step 2: Rearrange patches $\left\{P_A^i, P_B^i\right\}_{i=1}^N$ into column vectors. Vector set $\left\{\hat{v}_A^i, \hat{v}_B^i\right\}_{i=1}^N$ is obtained by subtracting their corresponding mean values from $\left\{v_A^i, v_B^i\right\}_{i=1}^N$.

$$\hat{v}_A^i = v_A^i - \bar{v}_A^i \tag{5}$$

where $\bar{v}_A^i$ is the mean value of v_A^i.

Step 3: Among the vectors in the set $\left\{\hat{v}_A^i, \hat{v}_B^i\right\}$ choose a vector v_m, which has the maximum variance. Then perform R-transform on the selected vector v_m. Out of six transformed coefficient, select the index *k* of the coefficient, which has the maximum value.

Step 4: Calculate the sparse coefficient vectors $\left\{\alpha_A^i, \alpha_B^i\right\}_{i=1}^N$ with the chosen dictionary D_k using Orthogonal Matching Pursuit algorithm.

$$\alpha_A^i = \arg\min \left\|\alpha_A^i\right\|_0 s.t. \left\|\hat{v}_A^i - D_k \alpha_A^i\right\|_2 \leq \varepsilon \tag{6}$$

$$\alpha_B^i = \arg\min \left\|\alpha_B^i\right\|_0 s.t. \left\|\hat{v}_B^i - D_k \alpha_B^i\right\|_2 \leq \varepsilon \tag{7}$$

Step 5: Fused sparse vector α_F^i is obtained by 'Max-L1' fusion rule [14] given as follows.

$$\alpha_F^i = \begin{cases} \alpha_A^i & if \ \left\|\alpha_A^i\right\|_1 > \left\|\alpha_B^i\right\|_1 \\ \alpha_B^i & otherwise \end{cases} \tag{8}$$

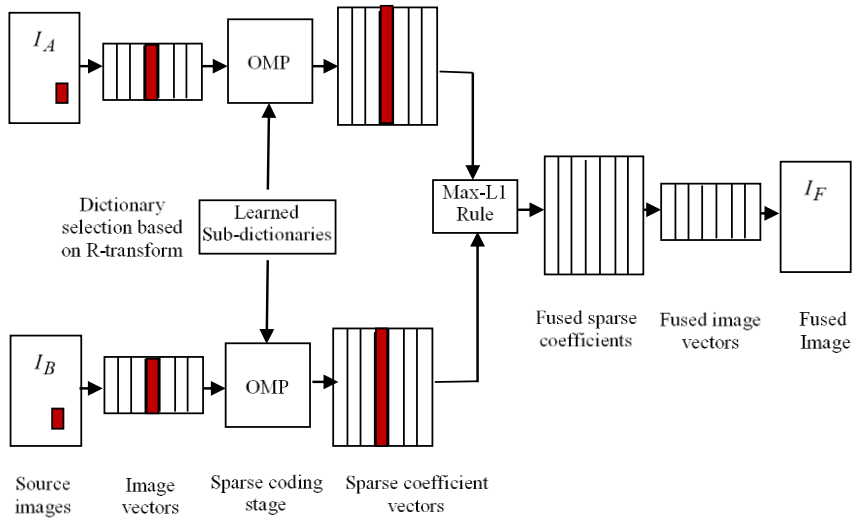


Fig. 2. Schematic representation of proposed fusion method.

Step 6: The fused vector is computed by

$$v_F^i = D_k \alpha_F^i + \bar{v}_F^i \tag{9}$$

where $\bar{v}_F^i = \dfrac{\hat{v}_A^i + \hat{v}_B^i}{2}$.

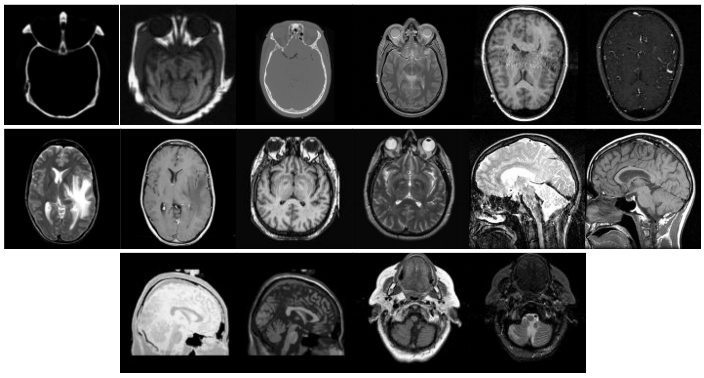
Step 7: Reshape each vector in v_F^i into a patch and then merge all the patches to get the final fused image I_F.

4. Results and Discussion

The 8 medical image pairs and 12 multi-focus image pairs shown in Fig. 3 (obtained from USTC dataset [29]) are used to test the proposed fusion scheme. Dictionary is learnt using the first 10 pairs of multi-focus images in Fig. 3. The range θ for computing the R transform of the vector is fixed at 0° to 180° with 30-degree spacing. The proposed method is implemented with dictionaries of size 128 and 256.

In order to prove the effectiveness of the proposed method, a comparative analysis of the proposed scheme with existing state-of-the-art fusion schemes is carried out. DTCWT [9], CVT [8], NSCT [10], LP [6], PC Fusion [4], Jihmsp [5] and Adaptive Sparse Representation (ASR) [22] fusion methods are considered for comparison. For DTCWT, CVT, NSCT and LP methods, the decomposition level is set to 4. Simulation results of the proposed and existing fusion schemes are obtained with MATLAB R2014a running on a 64 bit, INTEL (R) 2.67 GHz processor with 4GB RAM.

(a)



(b)

Fig. 3. Source image pairs:
(a) Multifocus images (12 pairs), (b) Medical images (8 pairs).

4.1. Qualitative analysis

Figure 4 shows the fusion results of multi-focus image pairs. From the fused image of NSCT method, it can be observed that the image has very poor contrast. Fused image of DTCWT and CVT method contain many artefacts especially around the corners of the clock and head regions. LP method does not preserve the clearer regions of source images and also suffers from significant artefacts. The fused image of ASR method and the proposed method are almost similar and do not show much difference.

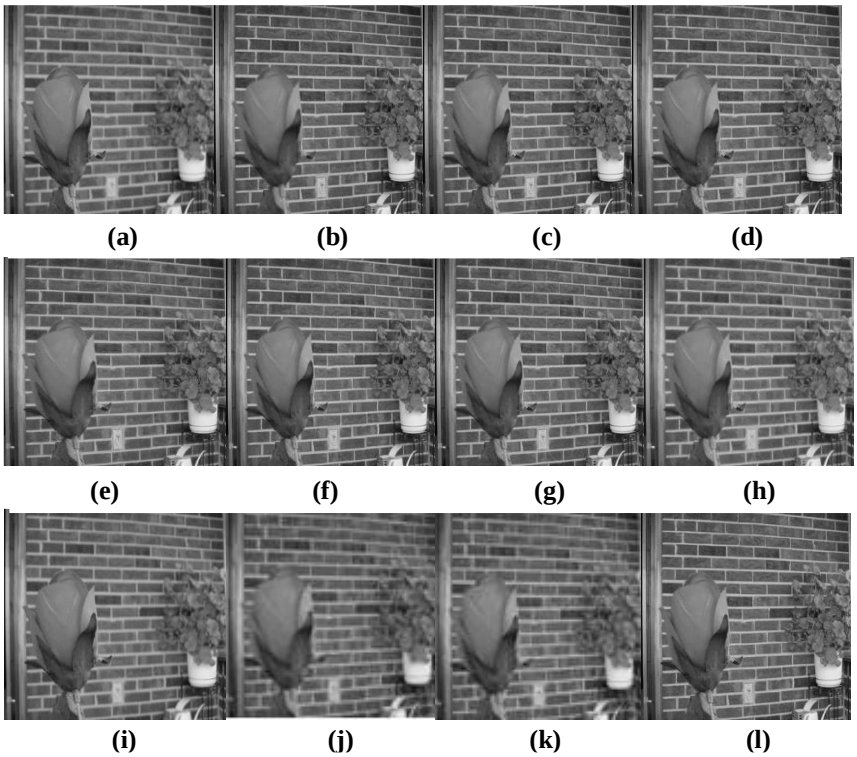


Fig. 4. Fusion results of multi-focus images: (a) Near focused image, (b) Far focused image, (c) ASR-256 [22], (d) ASR-128 [22], (e) NSCT [10], (f) DTCWT [9], (g) CVT [8], (h) LP [6], (i) PC Fusion [4], (j) Jihmsp [5], (k) Proposed 128, (l) Proposed 256.

Fusion results of CT and MRI images are shown in Fig. 5. A fusion of these images must have good contrast for accurate diagnostic purposes. The fused image of the LP method contains many blocking artefacts and is of very poor contrast. NSTC method suffers from the shadowing effect. Inner structures in fused images of DTCWT and CVT are clear but these methods also suffer from significant artefacts. ASR 128 method also has very less contrast compared to ASR 256. Fused image obtained by the proposed fusion scheme has better contrast and is more informative than the images obtained using other fusion schemes.

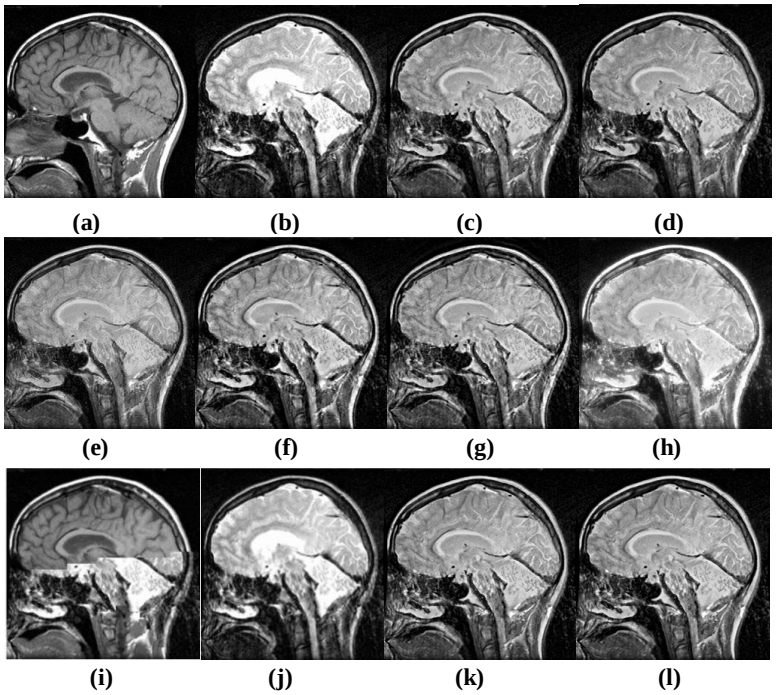


Fig. 5. Fusion results of medical images: (a) CT image, (b) MR image, (c) ASR-256 [22], (d)ASR-128 [22], (e) NSCT [10], (f) DTCWT [9], (g) CVT [8], (h) LP [6], (i) PC Fusion [4], (j) Jihmsp [5], (k) Proposed 128, (l) Proposed 256.

4.2. Quantitative analysis

Proposed fusion scheme is quantitatively analysed by computing the performance metrics, namely standard deviation [30], spatial frequency [31], $Q^{AB/F}$ [32], Q_w [33, 34], Q_Y [35] and Q_{CB} [36]. These metrics are explained as follows.

Standard Deviation (SD) (σ) is considered as one of the best metrics to measure the contrast in an image. Higher the standard deviation value better is the image contrast.

SD of the fused image $F(i, j)$ is given by

$$\sigma = \sqrt{\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}(F(i, j) - \mu)^2} \tag{10}$$

where μ - mean of the grayscale *F*. *M* and *N* are the dimensions of the image *F*.

Spatial Frequency (*SF*) reflects the amount of edge information preserved in the final fused image. It is given by

$$SF = \sqrt{RF^2 + CF^2} \tag{11}$$

RF and *CF* indicate the row and column frequency of the fused image *F*. *RF* and *CF* are calculated as,

$$RF = \sqrt{\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=2}^{N}\left[F(i,j) - F(i,j-1)\right]^2} \tag{12}$$

$$CF = \sqrt{\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=2}^{N}\left[F(i,j) - F(i-1,j)\right]^2} \tag{13}$$

Petrovic's metric $Q^{AB/F}$ evaluates the relative amount of edge information that is transferred from the input images *A* and *B* to the composite fused image *F*. It is defined as

$$Q^{AB/F} = \frac{\sum_{m=1}^{M}\sum_{n=1}^{N} Q^{AF}(n,m)\,w^A(n,m) + Q^{BF}(n,m)\,w^B(n,m)}{\sum_{i=1}^{M}\sum_{j=1}^{N}(w^A(i,j) + w^B(i,j))} \tag{14}$$

where $Q^{AF}(n,m) = Q_g^{AF}(n,m)Q_\alpha^{AF}(n,m)$. Edge strength and orientation preservation values at pixel (n,m) are denoted by $Q_g^{AF}(n,m)$ and $Q_\alpha^{AF}(n,m)$. The weighting factors $w^A(n,m)$ and $w^B(n,m)$ indicates the significance of Q^{AF} and Q^{BF} .

Xydeas and Petrovic [32] proposed the universal image quality index (UIQI) based fusion metric. Q_w is given as:

$$Q_w = \sum_{w\in W} C(w)\lambda(w)Q_0(A,F/w) + (1 - \lambda(w))Q_0(B,F/w) \tag{15}$$

$Q_0(A,F/w)$ and $Q_0(B,F/w)$ are calculated using the method in a local sliding window *w*. The saliency weight is calculated by:

$$\lambda(w) = \frac{s(A/w)}{s(A/w) + s(B/w)} \tag{16}$$

The saliency measure $s(A/w)$ and $s(B/w)$ are calculated using the variance of *A* and *B* in window *w*, respectively. Normalized salience $C(w)$ is obtained by:

$$C(w) = \frac{\max(s(A/w), s(B/w))}{\sum_{w'\in W}\max(s(A,w'), s(B,w'))} \tag{17}$$

Yang's fusion metric Q_Y evaluates the structural information of input images that were preserved in the fused image. Q_Y is given by

$$Q_Y = \begin{cases} \lambda(w)SSIM(A,F \mid w) + (1-\lambda(w)SSIM(B,F \mid w), SSIM(A,B \mid w) \geq 0.75 \\ \max\{SSIM(A,F \mid w), SSIM(B,F \mid w)\}, SSIM(A,B \mid w) < 0.75 \end{cases} \quad (18)$$

SSIM is a structural similarity between two images, w is a local window and $\lambda(w)$ is the local weight.

Chen-Blum metric (Q_{CB}) is a human perception based fusion metric. It measures the amount of information transferred from input images to the fused image. Q_{CB} is given by

$$Q_{CB} = \overline{Q_P}(x,y) \quad (19)$$

Q_P is the global quality index computed for each pixel? Q_P is given as follows

$$Q_P(x,y) = \lambda_A(x,y)Q_{AF}(x,y) + \lambda_B(x,y)Q_{BF}(x,y) \quad (20)$$

λ_A and λ_B is the saliency map based on image contrast. Q_{AF} and Q_{BF} are the information preservation values. Chen and Blum [36] given a more detail calculation of Q_{CB} . The value of $Q^{AB/F}$, Q_Y , Q_w and Q_{CB} are between 0 and 1.

The value of $Q^{AB/F}$, Q_Y , Q_w and Q_{CB} are between 0 and 1. These metrics are computed for existing and proposed methods. From the computed results of multi-focus image pairs, the average value is calculated for each metric and is tabulated in Table 1. The highest value of the corresponding metric is indicated in bold font. From Table 1, it is clear that $Q^{AB/F}$, Q_Y and Q_w values are high for the proposed scheme with dictionary size 256. Most of the information in source images is thus, well preserved in the fused image of the proposed method.

However, the Standard Deviation (SD) and the spatial frequency (SF) values are higher for the DTCWT method. This is because DTCWT is a transform based fusion scheme that uses an over-complete wavelet transform with good shift invariance and directional sensitivity.

In addition, the fusion process involves all the transformed coefficients thus, retaining good contrast in the fused image. However, the main disadvantage of this method is that of increased memory and computation cost.

Sparse based methods are a form of lossy techniques and the sparse coefficients that are involved in the fusion process are only an approximate representation of the input images over a learnt dictionary. The main advantage of sparse methods is that the coefficients involved in the fusion process are sparser and thus, requires less memory compared to the traditional transform based techniques.

In addition, the fusion process is much simple compared to other methods. Since most of the coefficients are zero and only very few values are used in the fusion process, the contrast and edge information are less in the fused image, resulting in low SD and SF values for the proposed method.

Table 1. Average metric value of different fusion schemes on 12 pairs of multi-focus images.

Fusion schemes	Performance metrics					
	SD	**SF**	$Q^{AB/F}$	Q_w	Q_Y	Q_{CB}
Liu and Wang [22], ASR 128	**49.3751**	**22.6860**	**0.7511**	**0.9013**	**0.9511**	**0.7605**
Liu and Wang [22], ASR 256	**49.4502**	**2.8428**	**0.7534**	**0.9035**	**0.9519**	**0.7623**
Zhang and Guo [10], NSCT	**48.1703**	**20.0871**	**0.7305**	**0.8789**	**0.8832**	**0.7287**
Lewis et al. [9], DTCWT	**49.7370**	**3.3552**	**0.7436**	**0.8953**	**0.9397**	**0.7761**
Nencini et al. [8], CVT	**49.5459**	**23.2532**	**0.7331**	**0.8989**	**0.9200**	**0.7342**
Burt and Adelson [6], LP	**48.3203**	**17.3898**	**0.6664**	**0.8128**	**0.8711**	**0.6790**
Zhan et al. [4], PC Fusion	**49.7049**	**21.4301**	**0.0005**	**6.8×10^{-7}**	**7.2×10^{-5}**	**0.8133**
Zhan et al. [5], Jihmsp	**49.6902**	**22.9352**	**0.6287**	**0.7461**	**0.7824**	**0.7849**
Proposed 128	**49.5097**	**22.9512**	**0.7554**	**0.9048**	**0.9540**	**0.7661**
Proposed 256	**49.5357**	**22.9921**	**0.7571**	**0.9048**	**0.9558**	**0.7640**

However, the values obtained for the proposed scheme is almost closer to that of the DTCWT method. There is only a 0.5 % increase in SD value and around 1.5 % increase in SF value between the proposed scheme and DTCWT method. Also, only around 0.01 increase in Q_{CB} value is observed between the proposed method and DTCWT method.

Table 2 lists the average value of each metric obtained from the computed results of the medical image set. It can be inferred from Table 2, that the proposed method with dictionary size 256 has higher $Q^{AB/F}$, Q_w, and Q_Y values compared to other existing methods. This proves that more edge details and structural information of source images are preserved in the fused image. In addition, the correlation between the source image and fused image is higher for the proposed method.

Table 2. Average metric values of different Fusion schemes on 8 pairs of medical images.

Fusion schemes	Performance metrics					
	SD	**SF**	$Q^{AB/F}$	Q_w	Q_Y	Q_{CB}
Liu and Wang [22], ASR 128	**55.8926**	**19.8428**	**0.5662**	**0.6628**	**0.8131**	**0.5518**
Liu and Wang [22], ASR 256	**56.5946**	**20.7194**	**0.5969**	**0.6931**	**0.8185**	**0.5608**
Zhang and Guo [10], NSCT	**53.4970**	**17.0160**	**0.4680**	**0.5123**	**0.7450**	**0.5160**
Lewis et al. [9], DTCWT	**58.5727**	**22.4307**	**0.5588**	**0.6251**	**0.7027**	**0.4332**
Nencini et al. [8], CVT	**55.9599**	**21.7054**	**0.5206**	**0.5815**	**0.6729**	**0.4813**
Burt and Adelson [6], LP	**67.6112**	**17.8434**	**0.4798**	**0.5448**	**0.6522**	**0.3642**
Zhan et al. [4], PC Fusion	**61.7877**	**22.0931**	**0.0009**	**0.02×10^{-3}**	**0.0794**	**0.6832**
Zhan et al. [5], Jihmsp	**58.6324**	**21.5508**	**0.4108**	**0.4739**	**0.5975**	**0.6063**
Proposed 128	**56.9346**	**21.1285**	**0.6116**	**0.7075**	**0.8186**	**0.5659**
Proposed 256	**57.1749**	**21.4309**	**0.6200**	**0.7144**	**0.8230**	**0.5698**

Figures 6(a) and (b) compares the average Q values (average of $Q^{AB/F}$, Q_w, Q_Y and Q_{CB}) of the existing and proposed method for aircraft image pair and leaf image pair. Figures 6(c) and (d) shows the average Q value for two medical image set. Compared to all other existing fusion methods, the average Q value of the proposed scheme is higher for clock image and leaf image. For medical image, set 1 the average Q value of the proposed method is almost closer to Jihmsp method.

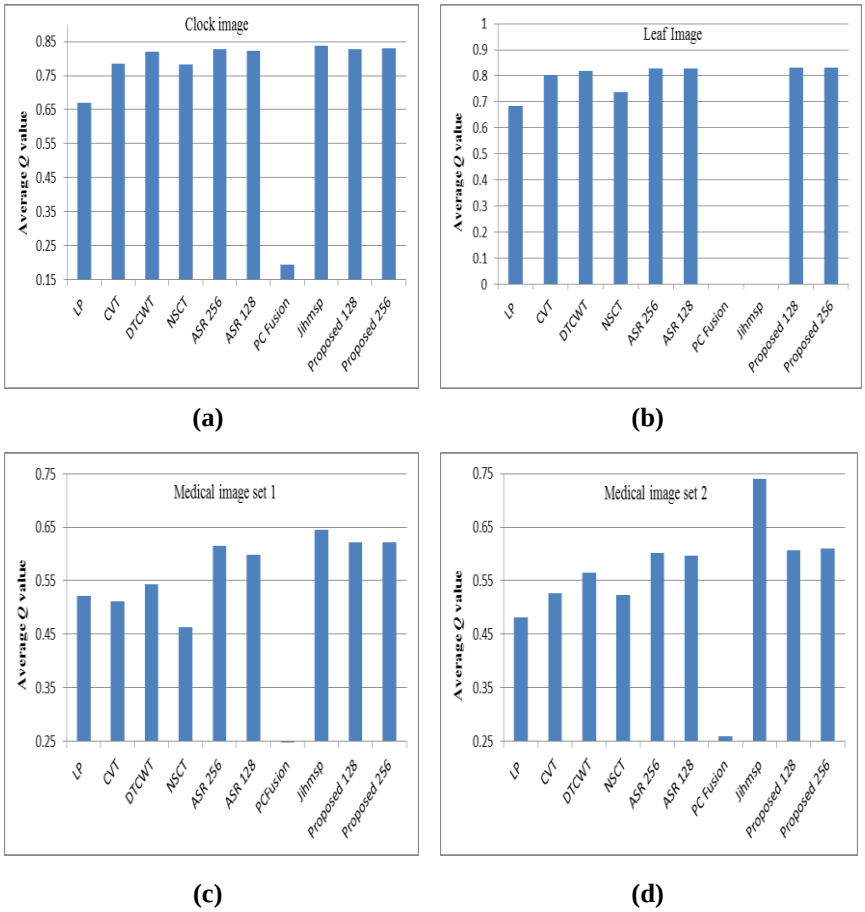


Fig. 6. Average *Q* metric value comparison with different fusion schemes.

The running time of sparse based methods is usually higher compared to non-sparse based methods. It would thus, be relevant to compare the running time of the proposed method with existing sparse based methods alone. Figure 7 shows the running time comparison of the proposed method with ASR scheme.

From Fig. 7, it can be inferred that the proposed method for dictionary size 128 as well as 256, takes very less time to perform fusion compared to the ASR methods. The time taken to fuse a given image pair is only 109 seconds for the proposed method with dictionary size 256, whereas the ASR method takes around 295 seconds.

The computation time of the proposed method is almost 50% less than the ASR method. The proposed scheme can thus, be considered as a computationally efficient method for image fusion.

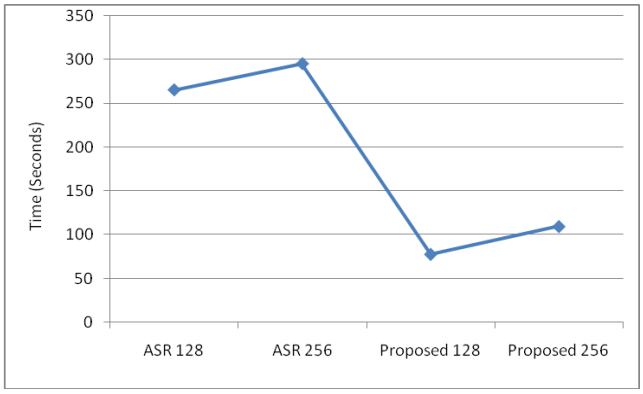


Fig. 7. Comparison of running time for sparse based methods.

5. Conclusions

A new SR based image fusion method is proposed in this paper. Set of compact sub-dictionaries are learned using the K-SVD dictionary algorithm. The training set for dictionary learning is obtained from the image patches that are classified based on R-transform. In the fusion stage, one sub-dictionary is chosen based on the R-transform of the image patch that has maximum variance. Sparse coefficients of the source image to be fused are obtained using the selected sub-dictionary. The final fused image is obtained by combining the coefficients of source images using Max-L1 rule. Quantitative and qualitative results show that the proposed scheme outperforms some of the existing SR based and transform based fusion scheme.

The proposed fusion method uses the K-SVD dictionary learning algorithm for classification of image patches that are to be fused. Dictionaries are learnt offline and are used as and when required for the fusion process. In future, more accurate sparsification of images can be obtained using an adaptive dictionary learning algorithm. Adaptive dictionaries can sparsify images better since they are learnt for the particular image instance or class of images. In addition, the presented work involves only grayscale images. The efficiency of the proposed algorithm infusing the colour images will also be analysed in future.

Nomenclatures

D	Over-complete dictionary
I_A	Source image A
I_B	Source image B
I_F	Fused image
P	Image patch set
$Q^{AB/F}$	Petrovic's metric
Q_{CB}	Chen-blum metric
Q_W	Universal image quality index
Q_y	Yang's fusion metric
R_f	Radon transform value
V	Image vectors
X	One dimensional signal

Greek Symbols

α	Sparse coefficient vector
δ	Dirac delta function
ε	Target error
θ	Projection angle

Abbreviations

ASR	Adaptive Sparse Representation
DCT	Discrete Cosine Transform
DTCWT	Dual Tree Complex Wavelet Transform
DWT	Discrete Wavelet Transform
NSCT	Non Subsampled Contourlet Transform
OMP	Orthogonal Matching Pursuit
SD	Standard Deviation
SF	Spatial Frequency
SR	Sparse Representation
UIQI	Universal Image Quality Index

References

1. Goshtasby, A.; and Nikolov, S. (2007). Image fusion: Advances in the state of the art. *Information Fusion*, 8(2), 114-118.
2. Bavachan, B.; and Krishnan, P. (2014). A survey on image fusion technique. *International Journal of Research in Computer and Communication Technology*, 3(3), 48-52.
3. Wan, T.; Zhu, C; and Qin, Z. (2013). Multifocus image fusion based on robust principal component analysis. *Pattern Recognition Letters*, 34(9), 1001-1008.
4. Zhan, K.; Li, Q.; Teng, J.; Wang, M.; and Shi, J. (2015). Multifocus image fusion using phase congruency. *Journal of Electronic Imaging*, 24(3), 033014-1-033014-12.
5. Zhan, K.; Teng, J.; Li, Q.; and Shi, J. (2015). A novel explicit multi-focus image fusion method. *Journal of Information Hiding and Multimedia Signal Processing*, 6(3), 600-612.
6. Burt, P.; and Adelson, E. (1983). The laplacian pyramid as a compact image code. *IEEE Transactions on Communications*, 31(4), 532-540.
7. Li, H.; Manjunath, B.S.; and Mitra, S.K. (1995). Multisensor image fusion using the wavelet transform. *Graphical Models and Image Processing*, 57(3), 235-245.
8. Nencini, F.; Garzelli, A.; Baronti, S.; and Alparone, L. (2007). Remote sensing image fusion using the curvelet transform. *Information Fusion*, 8(2), 143-156.
9. Lewis, J.J.; O'Callaghan, R.J.; Nikolov, S.G.; Bull, D.R.; and Canagarajah, N. (2007). Pixel and region based image fusion with complex wavelets. *Information Fusion*, 8(2), 119-130.
10. Zhang, Q.; and Guo, B.-l. (2009). Multifocus image fusion using the nonsubsampled contourlet transform. *Signal Processing*, 89(7), 1334-1346.

11. Phamila, Y.A.V.; and Amutha, R. (2014). Discrete cosine transform based fusion of multi focus images for visual sensor networks. *Signal Processing*, 95, 161-170.
12. Phamila, A.V.Y.; and Amutha, R. (2013). Low-complexity multi-focus image fusion in discrete cosine transform domain. *Optica Applicata*, 43(4), 693-706.
13. Elad, M.; and Aharon, M. (2006). Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Transactions on Image Processing*, 15(12), 3736-3745.
14. Yang, J.; Wright, J.; Huang. T.S. and Ma, Y. (2010). Image super-resolution via sparse representation. *IEEE Transactions on Image Processing*, 19(11), 2861-2873.
15. Yang, A.; and Li, S. (2010). Multifocus image fusion and restoration with sparse representation. *IEEE Transactions on Instrumentation and Measurement*, 59(4), 884-892.
16. Yang, B.; and Li, S. (2012). Pixel-level image fusion with simultaneous orthogonal matching pursuit. *Information Fusion*, 13(1), 10-19.
17. Yu, N.; Qiu, T.; Bi, F.; and Wang, A. (2011). Image features extraction and fusion based on joint sparse representation. *IEEE Journal of Selected Topics in Signal Processing*, 5(5), 1074-1082.
18. Yang, G.; Xu, X.; and Man, H.; (2011). Optimum image fusion via sparse representation. *Proceedings of the 20th Annual Wireless and Optical Communications Conference (WOCC)*. Newark, New Jersey, United States of America, 1-4.
19. Elad, M.; and Yavneh, I. (2009). A plurality of sparse representations is better than the sparsest one alone. *IEEE Transactions on Information Theory*, 55(10), 4701-4714.
20. Zhu, Z.; Yin, H.; Chai, Y.; Li, Y.; and Qi, G. (2018). A novel multi-modality image fusion method based on image decomposition and sparse representation. *Information Sciences*, 432, 516-529.
21. Zong, J.-j.; and Qiu, T.-s. (2017). Medical image fusion based on sparse representation of classified image patches. *Biomedical Signal Processing and Control*, 34, 195-205.
22. Liu, Y.; and Wang, Z. (2015). Simultaneous image fusion and denoising with adaptive sparse representation. *IET Image Processing*, 9(5), 347-357.
23. Olshausen, B.A.; and Field, D.J. (1996). Emergence of simple-cell receptive field properties by learning a sparse coding for natural images. *Nature*, 381, 607-609.
24. Elad, M.; Figueiredo, M.A.T.; and Ma, Y. (2010). On the role of sparse and redundant representations in image processing. *Proceedings of the IEEE*, 98(6), 972-982.
25. Aharon, M.; Elad, M.; and Bruckstein, A. (2006). K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on Signal Processing*, 54(11), 4311-4322.
26. Engan, K.; Aase, S.O.; and Husoy, J.H. (2000). Multi-frame compression: Theory and design. *Signal Processing*, 80(10), 2121-2140.

27. Elad, M. (2009). *Sparse and redundant representations*. From theory to applications in signal and image processing. New York, United States of America: Springer Science+Business Media.
28. Tabbone, S.; Wendling, L.; and Salmon, J.-P. (2006). A new shape descriptor defined on the Radon transform. *Computer Vision and Image Understanding*, 102(1), 42-51.
29. Johnson, K.A.; and Becker, J.A. (2016). The whole brain atlas. Retrieved January 12, 2016, from http://www.med.harvard.edu/AANLIB/.
30. Jagalingam, P.; and Hegde, A.V. (2015). A review of quality metrics for fused image. *Aquatic Procedia*, 4, 133-142.
31. Huang, W.; and Jing. Z. (2007). Evaluation of focus measures in multi-focus image fusion. *Pattern Recognition Letters*, 28(4), 493-500.
32. Xydeas, C.S.; and Petrovic, V. (2000). Objective image fusion performance measure. *Electronics Letters*, 36(4), 308-309.
33. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; and Simoncelli, E.P. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4), 600-612.
34. Piella, G.; and Heijmans, H. (2003). A new quality metric for image fusion. *Proceedings of the International Conference on Image Processing*. Barcelona, Spain, 173-176.
35. Yang, C.; Zhang, J.-Q.; Wang, X.-R.; and Liu, X. (2008). A novel similarity based quality metric for image fusion. *Information Fusion*, 9(2), 156-160.
36. Chen, Y; and Blum, R.S. (2009). A new automated quality assessment algorithm for image fusion. *Image and Vision Computing*, 27(10), 1421-1432.