UNIVERSITY OF VAASA

SCHOOL OF TECHNOLOGY AND INNOVATION

TELECOMMUNICATION ENGINEERING

Farshad Ghorbani Veshki

SUPERVISED COUPLED DICTIONARY LEARNING FOR MULTI-FOCUS IMAGE FUSION

Master's thesis for the degree of Master of Science in Technology submitted for inspection, Vaasa, August 24, 2018.

Supervisor	Mohammed Elmusrati
Instructor	Sergiy Vorobyov

ACKNOWLEDGEMENTS

First and foremost, I would like to express my sincere gratitude to Professor Sergiy Vorobyov for his contribution of idea and time, guidance and motivation throughout the entire work of this thesis.

Second, I would like to give my sincere thanks to Professor Mohammed Elmusrati for all his guidance and support during my Master's degree at the University of Vaasa.

Last, but not least, I am thankful to my wife Golnaz for collaborating with me on the biggest project which is life.

Espoo, August 23, 2018

PUBLICATION

The detailed explanation of the work of this master's thesis has been also submitted as the following manuscript.

Veshki, F. G., & Vorobyov, S. A. (2018). Multi-Focus Image Fusion using Sparse Representation Via Coupled Dictionary Learning. *IEEE Transactions on Computational Imaging*, Manuscript submitted for publication.

TABLE OF CONTENTS

TABLE OF FIGURES, ALGORITHMS AND TABLES

ABBREVIATIONS

BP	Basis Pursuit
DCT	Discrete Cosine Function
DOF	Depth of Field
DSIFT	Dense Scale Invariant Feature Transform
DWT	Discrete Wavelet Transform
IRLS	Iteratively Reweighted Least Squares
JPEG	Joint Photographic Experts Group
LARS	Least Angle Regression
LASSO	Least Absolute Shrinkage and Selection Operator
MAP	Maximum A-posteriori Probability
MI	Mutual Information
MOD	Method of Optimal Directions
MP	Matching Pursuit
MSE	Mean Squared Error
MWG	Multi-scale Weighted Gradient
NMI	Normalized Mutual Information
OLD	OnLine Dictionary learning
OMP	Orthogonal Matching Pursuit
PCA	Principal Component Analysis
Rand-OMP	Randomized Orthogonal Matching Pursuit
SF	Spatial Frequency
SIFT	Scale Invariant Feature Transform
SR-CM	Sparse Representation Choose-Max based image fusion
SR-FM	Sparse Representation of Focus Measure based image fusion
SR-KSVD	Sparse Representation choose-max based image fusion using K-SVD
SSIM	Structural Similarity
SVD	Singular Value Decomposition
SWT	Stationary Wavelet Transform
WLS	Weighted Least Square

UNIVERSITY OF VAASA
School of Technology and Innovation

Author:	Farshad Ghorbani Veshki
Topic of the thesis:	Supervised Coupled Dictionary Learning for Multi-Focus Image Fusion
Degree:	Master of Science in Technology
Master's Programme:	Communications and Systems Engineering
Supervisor:	Mohammed Elmusrati
Instructor:	Sergiy Vorobyov
Year of entering the University:	2016
Year of completing the thesis:	2018
Number of pages:	57

ABSTRACT

Among all methods that have tackled the multi-focus image fusion problem, where a set of multi-focus input images are fused into a single all-in-focus image, the sparse representation based fusion methods are proved to be the most effective. Majority of these methods approximate the input images over a single dictionary representing only the focused feature space. However, ignoring the blurred features sets limits on the sparsity of the obtained sparse representations and decreases the precision of the fusion.

This work proposes a novel sparsity based fusion method that utilizes a joint pair of dictionaries, representing the focused and blurred features, for the sparse approximation of source images. In our method, more compact sparse representations (obtained by using both features in the sparse approximation), and classification tools (provided by using the two known subspaces (focused and blurred)) are exploited to improve the performance of the existing state of the art fusion methods. In order to achieve the benefits of using a joint pair of dictionaries, a coupled dictionary learning algorithm is developed. It enforces a common sparse representation during the simultaneous learning of two dictionaries, fulfils the correlation between them, and improves the fusion performance. The detailed comparison with the state of the art fusion methods shows the higher efficiency and effectiveness of the proposed method.

KEY WORDS: Dictionary learning, sparse representations, multi-focus image fusion.

1. INTRODUCTION

Conventional photography cameras can capture focused images only from objects that are close to their plane of focus, with a limited depth of field (DOF). In order to have an all-in-focus image from multiple objects at different distances from the camera, multiple images with different focal settings should be taken from the scene. Then comparing all images, the sharpest representation of each object or area should be recognized, and finally the focused parts should be fused into one single image. This problem is called **multi-focus image fusion problem** and has various applications in medical imaging, remote sensing, computer vision, etc. (Nencini, Garzelli, Baronti, & Alparone, 2007; Calhou & Adali, 2009; Pajares & Cruz, 2004).

To capture a sharp image from an object, the focus plane of camera should be set so that the target is located inside the focused area, due to the camera's focal settings, including the focal distance and DOF. In this way, the image of each point from the object is approximately a point on the image plane (see Figure 1). Since the plane of focus can be set at one certain distance at a time, the image of objects at different distances than the target will be captured blurred.

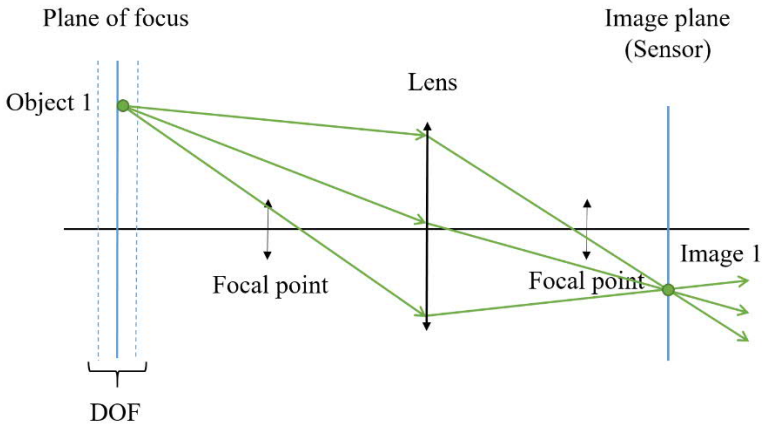


Figure 1. Image of a point near the plane of focus (inside DOF).

As it can be seen in Figure 2, the image of each point from every object which is not close enough to the camera's plane of focus is formed as a circle on the image plane. This is the main reason of blurredness in out of focus images.

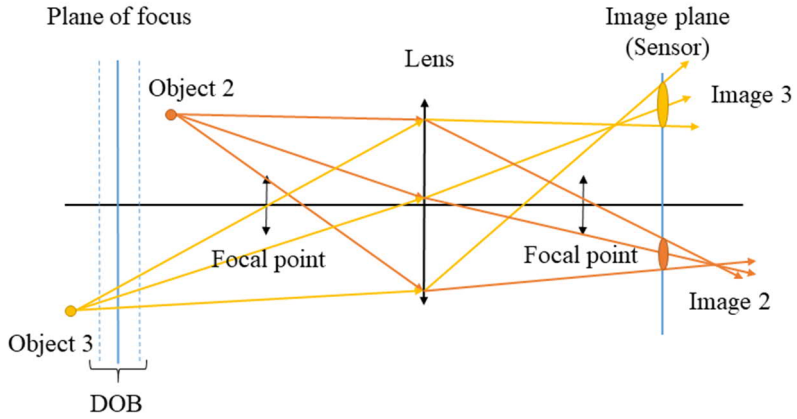


Figure 2. Image of a point away from camera's plane of focus.

In Figure 3, it is shown that in order to produce an all-in-focus image, the focused regions in multi-focus source images, taken by different focal settings, should be detected and fused. It should be mentioned that all source images are assumed to be taken from the same scene and be completely registered.

In recent decades, various methods have been developed to address the multi-focus image fusion problem. Majority of these methods can be divided into two general categories: **transform domain** and **spatial domain** fusion methods.

The transform domain fusion methods decompose the multi-focus source images into their transform coefficients, perform the fusion, and then invert the transform to reconstruct the all-in-focus image. Methods such as *discrete wavelet transform* (DWT) (Tian & Chen, 2012) and *stationary wavelet transform* (SWT) (Pradnya P. & Ruikar, 2013) are examples of the transform domain fusion methods.

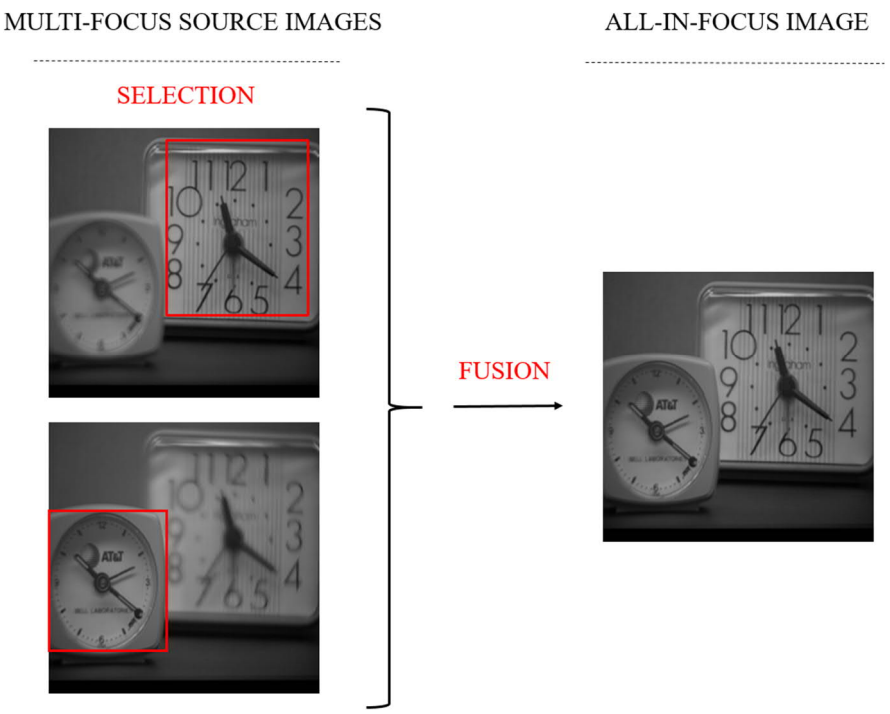


Figure 3. Multi-focus image fusion procedure.

On the other hand, the spatial domain fusion methods calculate the local features of source images' pixels as *focus measures*. Then by comparing the focus measures of all corresponding pixels (from a set of multi-focus images) they form a *decision map* and finally they fuse the source images by masking them according to the obtained decision map. *Spatial frequency* (SF) (Li & Yang, 2008), *multi-scale weighted gradient* (MWG) (Zhou, Li, & Wang, 2014) and *dense SIFT* (DSIFT) (Liu, Liu, & Zengfu, 2015) are some examples of using local features in spatial domain fusion methods.

Another approach that has been attracting significant attention in the field of image processing, similar to many other research areas of signal processing, is the approach that exploits the concepts of *sparsity* and *over-completeness* (Wan, Canagarajah, & Achim, 2008; Yang & Li, 2010; Wan, Qin, Zhu, & Liao, 2013). The conventional sparsity based multi-focus image fusion approaches obtain the sparse representations of source images over a single over-complete dictionary, learned on a training dataset of focused image patches, then fuse the sparse representations using a fusion rule, mostly based on max-l_1-norm which relates the focus level of sparse vectors to their activity level (Yang et al.

2010), and finally reconstruct the image from the fused sparse representation. Here, l_1-norm of a vector x is given as $\|x\|_1 = \sum_{\forall i}|x_i|$, and max-$l_1$-norm is defined as $x^F = \text{argmax}\|x^k\|_1$, where x^F is the most focused vector from the set of K multi-focus vectors $\{x^k\}_{k=1}^K$.

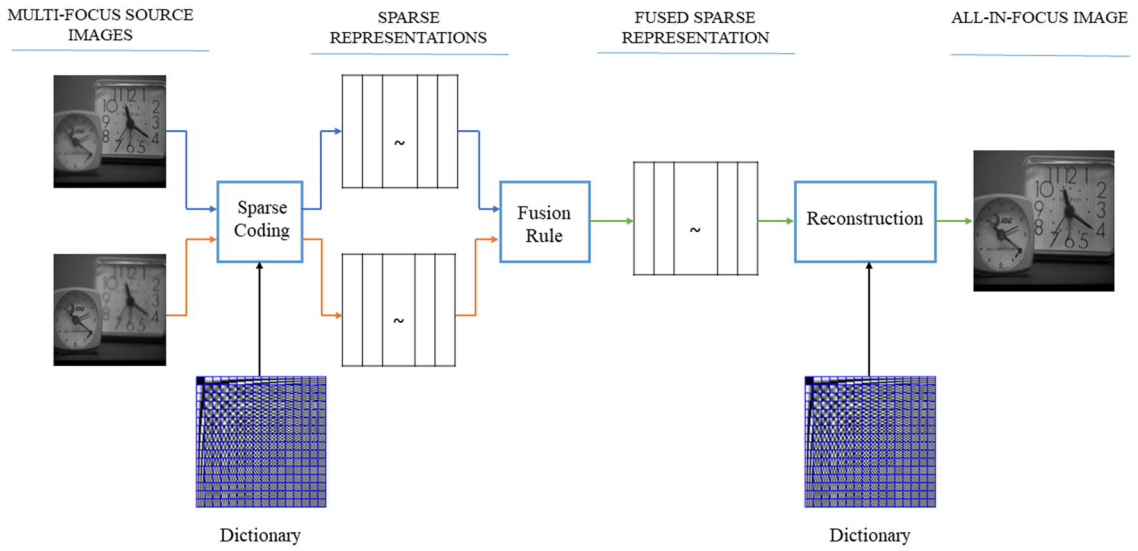


Figure 4. Diagram of multi-focus image fusion using sparse representations (Credit to (Zhang, Liu, Blum, Han, & Tao, 2018)).

The diagram in Figure 4 illustrates the general procedure of multi-focus image fusion using the sparse representations of source images.

Using a single dictionary that is only trained on focused features limits the accuracy of sparse coding. In other words, ignoring the blurred features in the sparse approximation of sets of image patches limits the ways of exploiting sparsity. This leads to a higher error in fusion performance as the selection of the most focused image patches is based on the sparsity of their sparse representation.

In this work, a fusion method based on the sparse representations of multi-focus source images over a couple of dictionaries that are learned over a pair of focused and blurred datasets is proposed. The term *"supervised"* in the title refers to the fact that the focused and blurred training datasets are produced manually. It is shown that providing both

focused and blurred features improves the accuracy of the sparse coding. The accurate sparse representations and available information on the distribution of the sparse vectors of coefficients over the two dictionaries provides tools that help to recognize which sparse vector is built from the focused features. Comparing the results of this work to the state of the art fusion methods shows that the proposed method has a superior performance both visually and quantitatively. A coupled dictionary learning method is also developed. By correlating the corresponding atoms of the two dictionaries, the fusion results are significantly improved, while the training time is dramatically reduced.

In the next chapter, all the basic concepts that are related to multi-focus image fusion based on sparse image representation are explained in details. In the third chapter, the development of proposed fusion and coupled dictionary learning methods are detailed in two separate subsections. Our simulation results are presented in Chapter 4, and finally the last chapter concludes what has been done in this master's thesis.

2. BACKGROUND

In this work, a multi-focus image fusion method based on sparse representation and coupled dictionary learning is proposed. Hence, this chapter is devoted to a review of essential concepts of this research field. The main concepts are the general theory of *sparse image representation* including *sparse approximation* and *dictionary learning* techniques, *sparse representations based fusion methods* and *fusion quality measuring indexes* that are used for the quantitative evaluation of the proposed method.

2.1. Sparse image representation

The concept of sparse representation is based on the fact that signals can be modeled as a weighted sum of a small number of elements taken from a large and *"good"* enough set[1] of atoms (basis functions (here vectors)), without losing much information. Thus, signals are modeled as vectors of coefficients representing the weights of their building elements. The term *"sparse"* stands for the fact that only few coefficients are nonzero (Wan et al., 2008; Yang et al., 2010; Wan et al., 2013).

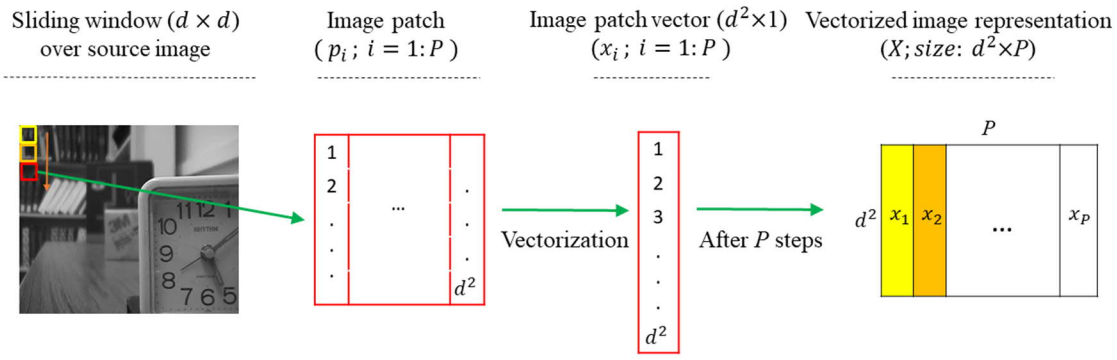


Figure 5. Image vectorization procedure.

[1] *"Good"* set of atoms are discussed in details in Subsection *2.3.*

For example, in the Fourier transform, signals are decomposed to a finite number of basis sinusoidal signals (frequencies). Considering the fact that the sinusoidal signals can have infinite number of frequencies, the Fourier transform of a signal is a sparse representation of that signal.

In the image processing applications, the signals are vectorized small pieces (image patches, e.g. 8×8), taken from the source images as the computational cost for processing the whole image is too high. In Figure 5, the diagram of image vectorization procedure is represented.

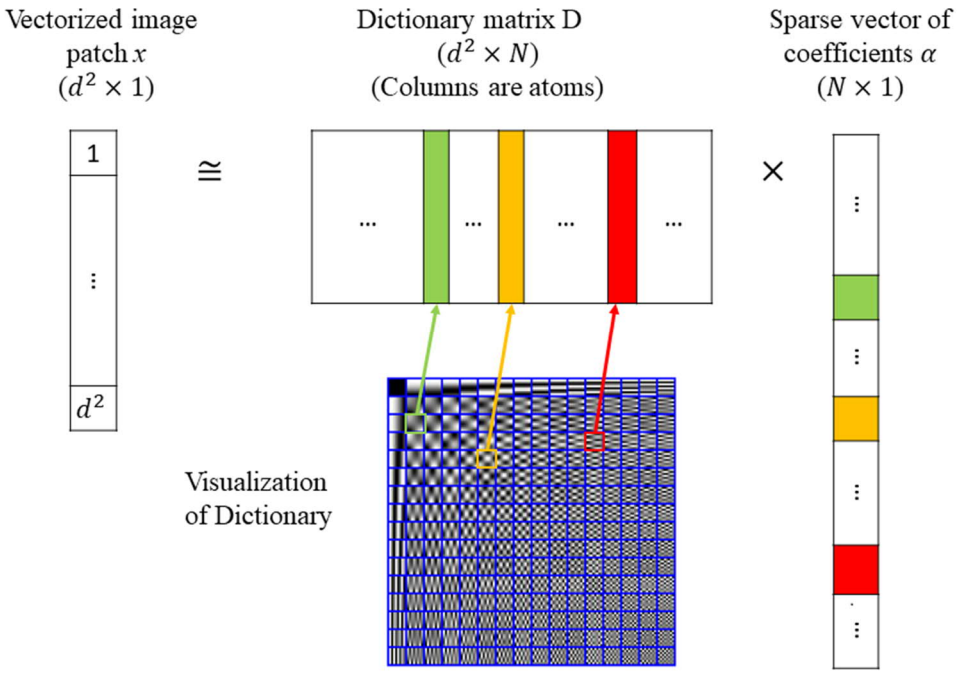


Figure 6. Sparse coding procedure.

The aim of sparse coding is to approximate each vectorized patch using only a few weighted samples from a set of predefined basis vectors. The predefined basis vectors are called *atoms* and a set of atoms is a *dictionary*. For instance, in *joint photographic expert groups* (JPEG) method, the *discrete cosine transform* (DCT) matrix is used as the dictionary, where the atoms are spectrums of cosine function.

The diagram in Figure 6 shows the procedure of sparse coding of a vectorized image patch over an over-complete dictionary. The term *"over-complete"* comes from the fact that the number of atoms in such dictionaries is greater than the number of dimensions of the atoms. Here it means that *N* is greater than d^2. As it is illustrated, the image patch *x* can be approximated by a linear combination of few atoms from the dictionary *D*. The required atoms for the best representation then can be found by solving the following optimization problem (Elad, 2007).

$$\min_{\alpha} \|\alpha\|_0 \quad s.t. \;\; x \approx D\alpha. \tag{1}$$

where $\|\cdot\|_0$ denotes the operator that counts the number of non-zero entries in a vector. The problem (1) is known to be NP-hard and takes an unacceptable time to be computed using conventional computers. Alternatively, *relaxation*, *greedy methods* and *adaptive dictionary learning* can be used to address the sparse coding problem. In the next two subsections, the widely used sparse approximation, and dictionary learning algorithms are introduced.

2.2. Sparse approximation algorithms

Different approaches have been adopted to solve the sparse approximation problem. Majority of existing methods can be categorized into two general groups: **convex relaxation** and **greedy methods** (Tropp, 2006a; Tropp, Gilbert, & Strauss, 2006b). Some of the most popular methods of each group are introduced next.

2.2.1. Relaxation methods

As it is mentioned before, one approach to solve the problem (1) is to replace the operator $\|\cdot\|_0$ by the l^1-norm. This is called convex relaxation and it is the basis of methods such as *basis pursuit* (BP) in signal processing (Chen, Donoho, & Saunders, 2001) and *least absolute shrinkage and selection operator* (LASSO) in statistics (Tibshirani, 2011). In both methods the problem of sparse approximation is changed to

$$\min_{\alpha} \frac{1}{2}\|x - D\alpha\|_2^2 + \|\alpha\|_1 \quad \textbf{(2)}$$

where the first term is the total square error and the second term is the l^1-norm of the coefficient vector x. Thus, the accuracy and sparsity are compromised in (2) (Chen et al., 2001).

Another widely used method based on convex minimization is the method of *iteratively reweighted least squares* (IRLS). In this method the operator $\|\cdot\|_0$ is replaced by a l^p-norm in the minimization problem for $p < 1$, and the corresponding problem is addressed by an iterative *weighted least square* (WLS) method (Chartrand & Yin, 2008).

2.2.2. Greedy methods

Majority of greedy sparse approximation algorithms use heuristic methods that in each iteration look for the best matching atom from the dictionary for approximating a target vector. The target vector r is initialized by vectorized image patches ($r = x$) and then it is updated to the residue vector $r \coloneqq r - D\alpha$ at the end of each iteration. That is why such methods are also called *matching pursuit* (MP). The iterations are continued until either the total error $\|x - D\alpha\|_2^2$ becomes smaller than maximum tolerance error ϵ or the number of entries of α reaches the maximum allowed number T_0 (Mallat & Zhang, 1993). Thus, instead of finding global optimum, MP approximates the solution of (1) in terms of solving:

$$\min_{\alpha} \|x - D\alpha\|_2^2 \; s.t. \; \|\alpha\|_0 < T_0 \,. \quad \textbf{(3)}$$

The method of *orthogonal matching pursuit* (OMP), which is an improved version of MP, is another commonly used sparse approximation method in signal processing. It significantly improved the performance of MP by adding an orthogonalization step, so that the coefficient of each selected atom (all non-zero entries) are found by least squares method. Thus, the residue vector is always orthogonal to all atoms that are already chosen, which means that each atom can be selected only one time. This reduces the computation

time and guarantees the convergence of the algorithm (Pati, Rezaiifar, & Krishnaprasad, 1993).

Different variants of MP have been developed to address specific problems in different research areas. Another variant with applications in image processing is the method of *randomized orthogonal matching pursuit* (Rand-OMP). This method is based on the fact that the average of multiple sparse representations of a signal is more accurate than the sparsest one alone.

Obtaining multiple sparse representations from one signal is realized by selecting the atoms based on a probability distribution that gives each atom a probability of being selected proportional to how much it matches the residue vector. In this way, by adding randomness, different representations are obtained in consecutive executions of algorithm (Elad & Yavneh, 2009).

2.3. Dictionary learning algorithms

The dictionary that is required for sparse approximation can be taken from a fixed basis, for instance DCT basis or *Wavelet* basis, which is of course simpler and faster. However, using dictionaries that are customized for the specific types of data in use leads to a more compact and accurate sparse representations.

The customization of dictionary can be realized through *"learning"* the dictionary over a training data of the same type of data that the dictionary is customized for. Three of most popular dictionary learning methods, namely K-SVD, *method of optimal directions* (MOD), and *online dictionary learning* (OLD) are introduced next.

2.3.1. K-SVD

K-SVD finds a suitable dictionary for the training data by alternating between two phases: sparse approximation and dictionary update. For the sparse approximation phase any

pursuit algorithm can be exploited, however, OMP is the most common. The dictionary update phase is a generalization of K-mean algorithm, modified for updating the dictionary atom using *singular value decomposition* (SVD) method (Aharon, Elad, & Bruckstein, 2006).

The problem of learning a dictionary can be formulated as:

$$\min_{D,\,A} \|X - DA\|_2^2 \;\; s.t. \;\; \|\alpha_i\|_0 < T_0 \;, \; \forall i \qquad \textbf{(4)}$$

where $A = [\alpha_1, \dots, \alpha_P]$ is the matrix of sparse representations and α_i is the *i*-th column of A.

In sparse approximation phase, K-SVD finds the best *A* (using for example OMP) over *D*. Then in the dictionary update phase, having *A*, it updates the dictionary atoms **one by one** so that the total square error $\|X - DA\|_2^2$ is minimized (Aharon et al.,2006).

To be more precise, after finding *A*, for each atom in the dictionary at time t ($D_k^t, k = 1{:}N$), the subset of vectorized patches that are using that atom X^k and their sparse representation A^k are found. Then the residue vector set R^k is formed as

$$R^k = X^k - D^t_{(n=1:N, n\neq k)} A^k \qquad \textbf{(5)}$$

where $D^t_{(n=1:N,n\neq k)}$ is the dictionary D excluding D_k^t at time t (D_k^t is replaced by a zero vector of the same size). Then, in order to find D_k^{t+1}, which is the updated D_k^t, K-SVD solves the following minimization problem.

$$D_k^{t+1} = \underset{D_k^t}{\operatorname{argmin}} \left\| R^k - D_k^t A_k^k \right\|_2^2 \qquad \textbf{(6)}$$

where A_k^k is the *k*-th row of A^k (the coefficients of D_k^t in A^k). Thus, D_k^t is updated to minimize the error in approximation of the residue vectors. Problem (6) is solved here using SVD. The two phases are repeated for a predefined number of iterations (Aharon et al., 2006)**.**

2.3.2. Method of Optimal Directions

The *method of optimal directions* (MOD) is similar to K-SVD in iterating between sparse approximation and dictionary update phases. Also, it solves the same problem (4). The only difference is that in the dictionary update phase, instead of updating the atoms one by one using SVD, the MOD updates the dictionary by solving the equation $X = DA$ analytically. The updated dictionary is found as $D = XA^{+}$ where A^{+} is the pseudoinverse of A (Engan, Aase, & Husoy, 1999).

2.3.3. Online dictionary learning

The *online dictionary learning* (OLD) method, similar to *online machine learning*, is developed to address the problems where the training data is received sequentially in time, and not at once. Thus, it is suitable in the cases when the processing data is very large, or the dictionary is needed to be updated constantly.

The OLD solves the problem, again, using an iterative two-phase approach. First, the sparse approximation phase solves (4) over A, keeping D unchanged. This is performed based on a linear regression method, suitable for high dimensional data, which is called *least-angle regression* (LARS) (Mairal, Bach, Ponce, & Sapiro, 2010).

Then, by solving the same problem (4), using the iterative method of *coordinate-descent* this time over D and keeping A unchanged, the dictionary is updated (Mairal et al., 2010).

2.4. Multi-focus image fusion using sparse representations

During the last decade, numerous multi-focus image fusion methods based on sparse representations have been proposed and the results have shown that their performance is better than those of other methods in this field (Wan et al., 2008; Yang et al., 2010; Wan et al., 2013; Nejati, Samavi, & Shirani, 2015).

In general, main approaches used in the aforementioned methods can be divided into two categories

- Fusion via sparse representations,
- Fusion via classification of sparse representations.

An example for each categories is given below.

2.4.1. Fusion via sparse representations

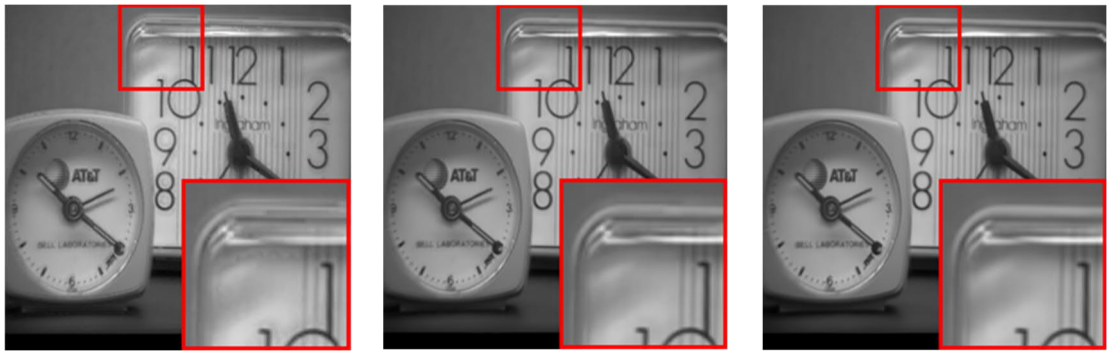
In the methods that are based on the fusion of sparse representations, first, the sparse representations of multi-focus source images are obtained. Next, using a selection rule, the most focused sparse vector, among sparse representations of all corresponding (spatially) source image patches, is found. Going across all sets of corresponding multi-focus vectors, the fused sparse matrix is formed. At the end, using the same dictionary that is used for sparse coding of the source images, the all-in-focus image can be reconstructed from the fused sparse matrix. The diagram of this procedure is given in Figure 4.

In the work of Yang et al. (2010, 884-892), a similar approach using DCT dictionary has been proposed. However, in later experiments, it has been shown that the use of an adaptively learned dictionary can lead to better fusion results. The improvement is obtained in two phases of the fusion method. At first, the use of adaptively learned dictionary improves the accuracy of fusion by yielding more compact (sparser) representations. This helps the selection rule to be more effective in the formation of a decision map with higher accuracy.

The second contribution of the adaptively learned dictionary is in the reconstruction phase, where the sparse approximations are then much more accurate in comparison to the cases where fixed basis dictionaries are used (Zhang et al. 2018). In addition, the use of sparser representations reduces the computational cost of fusion, simply because less nonzero entries needs to be computed.

A popular fusion rule, that is also common in many other fields of signal processing and statistics, is the max-l^1-norm rule. This rule suggests that sparse vectors that show higher activity level from the atoms of dictionary, in the approximations of their corresponding data, contain higher amount of information. Here it means that they are the most focused image patches. The measurement of the so called activity level is realized by calculating the l^1-norm of the sparse vectors (Zhang et al. 2018).

A disadvantage of the local information measuring based fusion methods, such as the Laplacian energy, spatial Frequency and entropy in spatial domain, and l^1-norm methods, is that when a flat region in the source images is being processed, the most noisy image patches are highly likely to be mistaken as the most focused. It is because their texture is more complicated than the completely uniform image patches. A similar mistake is also probable near the edges, where in the blurred images, the spread of colors or pixel intensities form an object into a uniform region increases the local variance. A remedy for this problem is to process images using bigger sliding windows, however, this will dramatically increase the computational costs. A comparison between quality of fusion and the size of sliding window is given in Figure 7.



(a) (b) (c)

Figure 7. Comparison of fusion results using different sizes of sliding window: (a) 2^2, (b) 4^2, and (c) 8^2.

All fused images in Figure 7 are produced using the same method proposed by Yang et al. (2010, 884-892), and the only parameter that is changed is the size of sliding window. Looking at the shadowing effect on the horizontal edge of the bigger clock, which is

magnified at the right bottom corner, we can see the direct impact of the window size on the fusion quality. However, this improvement in quality is very expensive computationally as the fusion running time (including sparse approximation) using sliding window sizes of 2^2, 4^2 and 8^2 turned out to be 3.3, 18.8 and 317.4 seconds, respectively.

2.4.2. Fusion via classification of sparse representations

Another approach to address the multi-focus image fusion problem is to exploit the data classification techniques that are based on sparse representation (Wright, Ganesh, Sastry, & Ma, 2009; Zhang, Li, & Huang, 2014; Yuan, Liu, & Yan, 2012).

As an example, a fusion method that works based on the classification using correlation of sparse representations is as follows (Nejati et al., 2015).

The method finds the most focused image patches by comparing the correlations between the sparse representations of local focus measures of multi-focus image patches and two pooled feature vectors (focused or blurred).

This process requires two phases: training, and fusion. In the training phase, the two pooled feature vectors v^F and v^B, and the dictionary D are obtained. Then, using D, v^F and v^B, the fusion can be performed. The block-diagram of the training phase is illustrated in Figure 8. As it can be seen, in the first step, going across the multi-focus images S_1 and S_2, using a sliding window (of size, e.g., d^2), the local focus measures of all pixels are calculated and the focus information maps F_1 and F_2 are formed (pixel values of F_1 and F_2 are the local focus measures of pixels in S_1 and S_2, respectively).

Next, by applying the patch extraction process to F_1 and F_2, the matrices X_1 and X_2 are formed, which are used as the training data for adaptively learning the dictionary D. Then, all the patches in X_1 and X_2 are manually labeled as focused or blurred, so two patch groups X^F and X^B (the focused and blurred datasets, respectively) are produced. By sparse approximation of the grouped patches, the sparse matrices A^F and A^B are obtained, and

finally by raw-wise aggregation of A^F and A^B, using *maximum* or *average* pooling methods, two pooled feature vectors v^F and v^B are obtained (Nejati et al., 2015).

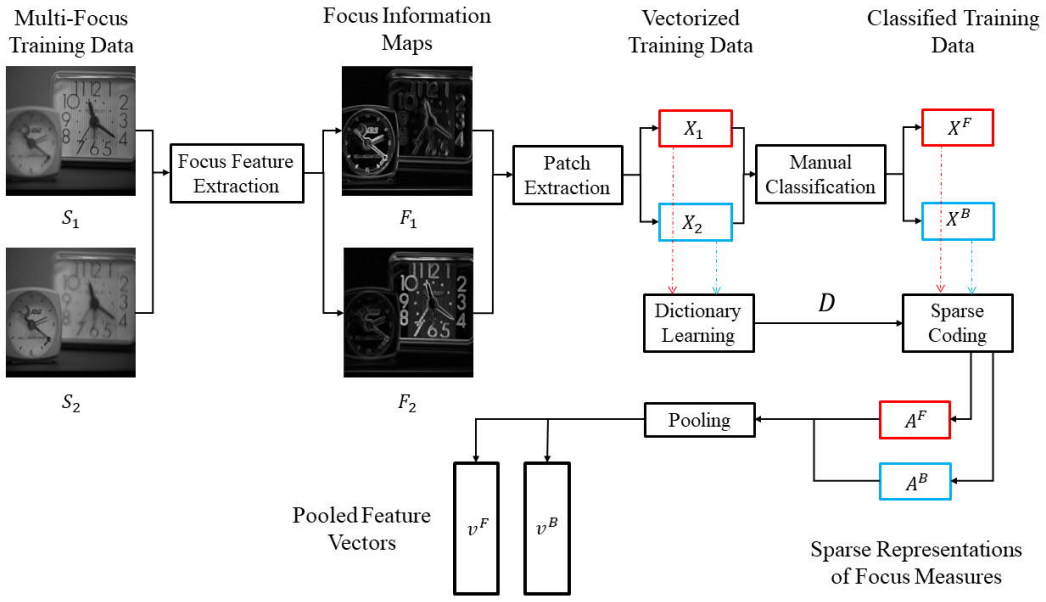


Figure 8. Multi-focus image fusion based on classification of sparse representations of focus measures, training phase.

In fusion phase (see Figure 9), first the focus information maps (F_1 and F_2) are extracted from the input multi-focus source images (S_1 and S_2), and then their sparse representations (A_1 and A_2) are found over the dictionary D. It should be noted that similar focus measures are used in the training and fusion phases.

Next, by comparing the correlations between the corresponding sparse vectors in sparse representation matrices (A_1 and A_2) and the pooled feature vectors v^F and v^B, the focused pixels are found and the decision map is formed. The correlation values are calculated using dot product of the vectors. Finally, by masking the multi-focus source images using the decision map, the all-in-focus fuse image is obtained (Nejati et al., 2015).

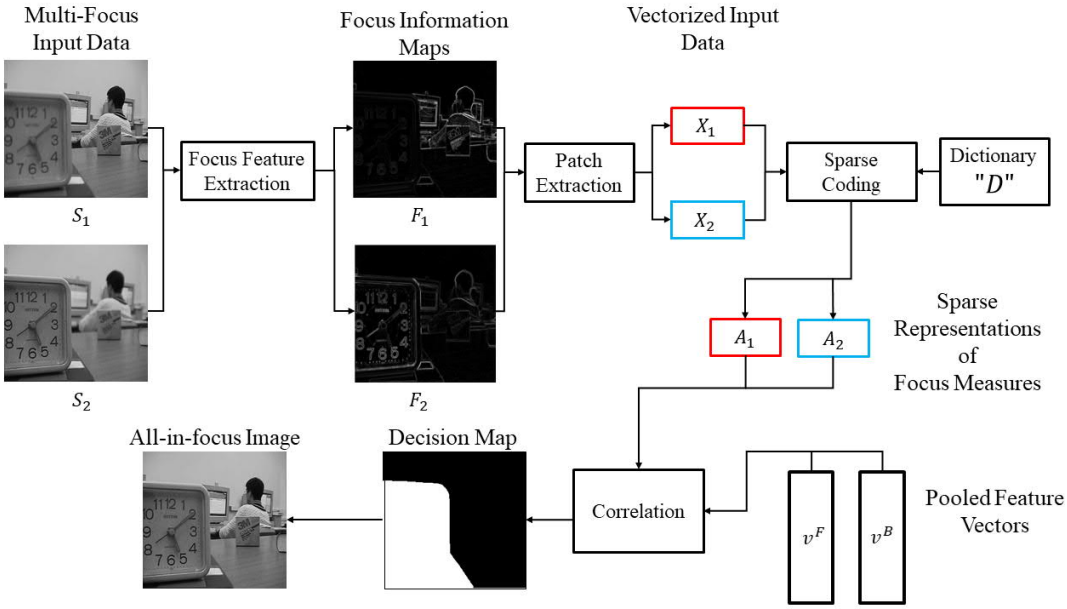


Figure 9. Multi-focus image fusion based on classification of sparse representations of focus measures, fusion phase.

2.5. Fusion quality measures

The importance of image fusion applications in unmanned systems (e.g., satellite imaging (Nichol & Wong, 2005)) and the ever increasing customer demand for convenience in all commercial sectors that utilize image fusion techniques (e.g., medical imagery, mobile phone industry, etc.) lead to recognition that the quantitative measurement of performance in image processing is an important problem. Moreover, a precise method for measuring the quality of fusion can be a part of the solution of the multi-focus image fusion problem itself, because it can be used as an objective function for designing a fusion method. Therefore, the problem of finding effective methods for the objective evaluation of image fusion quality remains an important topic in the image fusion research field.

In this work, the fusion quality is evaluated using four quality measurement indexes: *the objective image fusion measure* ($Q_{AB/F}$) (Xydeas & Petrovic, 2000), *normalized mutual information* (NMI) (Hossny, Nahavandi, & Creighton, 2008), *structural similarity index* (SSIM) (Wang & Bovik, 2004), and *mean squared error* (MSE). The first two indexes

are non-reference-based, meaning that they compare the fused image only to the source images, and the second two indexes are reference-based, which means they compare the produced all-in-focus image to a reference image known as the ideal perfect fusion of the source images.

2.5.1. Petrovic's index

The objective image fusion measure $Q_{AB/F}$, also known as Petrovic's index, evaluates how the important visual information of source images are transferred into the fused image. Here the important information in images is the *edge information*.

In this method, using the *edge strength* and *edge orientation* information of source images, obtained by applying the *Sobel edge operator*[2], the indexes Q_{α}^{AF} and Q_{g}^{AF}, which respectively show how well the *edge* and *strength* information of the pixels of source image *A* are persisted in the fused image *F*, are computed. Then Q^{AF} is calculated as

$$Q^{AF} = Q_{\alpha}^{AF} . Q_{g}^{AF}. \quad \textbf{(7)}$$

For example, for the case with two source images (*A* and *B*) of size $M \times N, Q_{AB/F}$ is calculated as

$$Q_{AB/F} = \frac{\sum_{n=1}^{N}\sum_{m=1}^{M} Q^{AF}(n,m) w^{A}(n,m) + Q^{BF}(n,m) w^{BF}(n,m)}{\sum_{n=1}^{M}\sum_{m=1}^{M} w^{AF}(n,m) + w^{BF}(n,m)} \quad \textbf{(8)}$$

where (*n*, *m*) is the pixel location in the image and $w^{AF}(n,m)$ and $w^{BF}(n,m)$ are the weights of $Q^{AF}(n,m)$ and $Q^{BF}(n,m)$. The formula (8) can be expanded for more than two source images by finding the weighted average of more Q^{XF} indexes, where *X* can be from any number of images (Xydeas et al., 2000).

---

[2] "The Sobel operator performs a 2-D spatial gradient measurement on an image and so emphasizes regions of high spatial frequency that correspond to edges" (Fisher, Perkins, Walker, & Wolfart, 2003).

2.5.2. Normalized mutual information

The *normalized mutual information* (NMI) is a modified version of the *mutual information* measuring index (MI) that is widely used in information processing field. The MI calculates the mutual information using the *marginal entropies* of source images, *H(X)* and *H(Y)*, and their *joint entropy*, *H(X, Y)*, as (Hossny et al., 2008)

$$MI(X,Y) = H(X) + H(Y) - H(X,Y)\ . \quad \textbf{(9)}$$

For the case of image fusion, MI is found as

$$MI(A,B,F) = MI(A,F) + MI(B,F) \quad \textbf{(10)}$$

where A and B are the source images and F is the fused image. The NMI is then the normalized version of (10), and it is found as

$$NMI(A,B,F) = 2\left[\frac{MI(A,F)}{H(A)+H(F)} + \frac{MI(B,F)}{H(B)+H(F)}\right]. \quad \textbf{(11)}$$

The NMI measures how precise the pixel information (intensities) from the source images is transferred to the fused image (Hossny et al., 2008).

2.5.3. Structural similarity

The *structural similarity* (SSIM) index compares the input signals (here the reference and fused images) through three features: *luminance*, *contrast* and *structure*. All three features are calculated locally using a sliding window, then the global measurement is obtained by averaging all local measures (Wang et al., 2004).

The luminance is estimated using local mean intensity as

$$l_{i,j} = \mu_{i,j} = \frac{1}{d^2}\sum_{n=1}^{d^2} x_n \quad \textbf{(12)}$$

where $l_{i,j}$ is the luminance of the sliding window with size d^2 and the right top corner at location (i, j), and x is the vectorized form of the patch that corresponds to the sliding window.

To calculate the local contrast c, first the mean intensities are removed from the signal ($x = x - l$), so both the reference and fused images have same local mean intensities. Then the contrast is calculated using standard deviation as

$$c_{i,j} = \sigma_{i,j} = \left(\frac{1}{d^2 - 1}\sum_{n=1}^{d^2}(x_n - \mu_{i,j})^2\right)^{1/2} \quad \textbf{(13)}$$

where $\mu_{i,j}$ is the mean intensity of $x_{i,j}$. The structure components of signals are obtained by removing their mean intensities and dividing them by their standard deviation as (Wang et al., 2004)

$$s_{i,j} = \frac{x_{i,j} - l_{i,j}}{c_{i,j}} = \frac{x_{i,j} - \mu_{i,j}}{\sigma_{i,j}}. \quad \textbf{(14)}$$

Thus, structure components have zero mean and unit standard deviation (the normalized reference and fused image patches).

The simplified formula for calculating local $SSIM_{i,j}$ is given as

$$SSIM_{i,j} = \frac{(2\mu_f\mu_{ref} + C_1)(2\sigma_{xy} + C_2)}{(\mu_f^2 + \mu_{ref}^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad \textbf{(15)}$$

Where the sub-scripts *f* and *ref* stand for the patches from the fused and reference images, respectively, with top right corner at location (i, j), and C_1 and C_2 are constants that

stabilize the function in the cases that $(\mu_f^2 + \mu_{ref}^2)$ and $(\sigma_x^2 + \sigma_y^2)$ are close to zero. The global value of SSIM is calculated by averaging all the local $SSIM_{i,j}$ values for all (i, j) (Wang et al., 2004).

2.5.4. Mean square error

The *mean square error* (MSE) is a simple statistical measure for calculation of the average squared difference between two signals, which here are the fused *(F)* and reference *(Ref)* images. For two dimensional (gray scale) images of size $N \times M$, the MSE is calculated as

$$MSE = \frac{1}{N \times M} \sum_{i=1}^{N} \sum_{j=1}^{M} (F_{i,j} - Ref_{i,j})^2 . \quad \textbf{(16)}$$

3. PROPOSED METHOD

This work proposes a fusion method based on the sparse representations of multi-focus source images over a couple of dictionaries, representing the focused and blurred feature spaces. The dictionaries are learned using two separate training datasets, one taken from *all-in-focus* and the other from *all-out-of-focus* images.

3.1. Problem formulation

The problem of finding a smooth all-in-focus image I^F by fusing all focused regions of a set of K multi-focus source images $\{I_k\}_{k=1}^K$ can be formulated as

$$I^F = \mathcal{F}\{I_k\}_{k=1}^K + V \quad \textbf{(17)}$$

where $\mathcal{F}\{\cdot\}$ is the fusion operator and V represents the noise. The problem of finding an optimal $\mathcal{F}\{I_k\}_{k=1}^K$ is an *ill-posed*[3] problem. Thus, I^F can be found as a *maximum a-posteriori probability* (MAP) estimation

$$I^F = \underset{I^F}{\operatorname{argmax}}\, p(\{I_k\}_{k=1}^K | I^F).\, p(I^F) \quad \textbf{(18)}$$

where the first term (likelihood) admits the similarity between I^F and the source images, and the prior guarantees the smoothness and sharpness of the fused image. Considering the fact that fusion is performed using image patches, the fusion problem can be rewritten as the following optimization problem

$$\min_{i^F} \| i^F - \mathcal{F}\{\, [D^F, D^B], \{i_k\}_{k=1}^K \,\} \|_F^2 + \eta p(I^F) \quad \textbf{(18)}$$

[3] See Appendix 1

where $\{i_k\}_{k=1}^K$ are K sets of patches extracted from multi-focus images $\{I_k\}_{k=1}^K$, i^F is the locally found focused patches, and $\eta p(I^F)$ is the penalty function enforcing the global smoothness and sharpness of the reconstructed image I^F, with η being the regularization parameter. The notation $\|\cdot\|_F^2$ stands for the *Frobenios norm* of a matrix.

The notation $[D^F, D^B]$ represents the fact that fusion is performed over a coupled dictionary, consisting D^F and D^B representing blurred and focused dictionaries, respectively. The first term in the optimization problem (18) concerns the local fusion, while the second term concerns the global reconstruction of I^F.

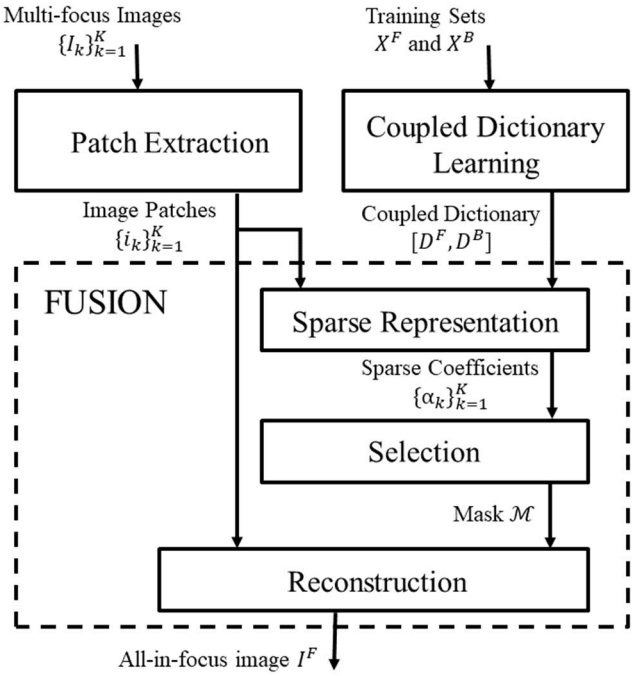


Figure 10. Diagram of the proposed fusion method.

In the block-diagram shown in Figure 10, the proposed fusion method is illustrated. First, according to the block *patch extraction*, the input image patches $\{i_k\}_{k=1}^K$ are extracted from $\{I_k\}_{k=1}^K$ using the same method that has been explained in Chapter 2, Subsection 2.1. In the block *coupled dictionary learning*, D^F and D^B are learned using the focus and blurred training data, respectively, which are denoted here by X^F and X^B.

The proposed fusion operator is divided into three sub-blocks. After the sparse representations of K sets of image patches, $\{\alpha_k\}_{k=1}^{K}$ are obtained, and then the most focused vectors are found using the proposed fusion rule and the mask $\mathcal{M}$ is formed. Finally, using $\mathcal{M}$ and $\{i_k\}_{k=1}^{K}$, the all-in-focus image is reconstructed.

To avoid blocking artifacts and uneven edges in the procedure of reconstruction of the all-in-focus image, overlapping patches are extracted from the source images. This helps to obtain a smoother image by averaging multiple patches at each pixel location, so the unwanted sudden changes of intensity in neighbor pixels are avoided.

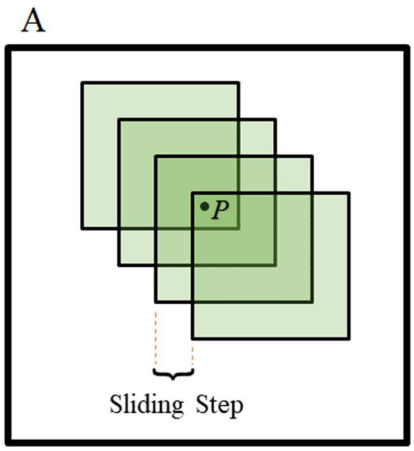


Figure 11. Extracting overlapping patches from images.

In Figure 11, a visual example of extracting overlapping patches from an image is shown. The point P appears in multiple patches. Thus, the reconstruction of this point is an average of all points at the same location from all patches that include this point. A bigger sliding window increases the number of these patches, while bigger sliding step decreases it.

In reconstruction phase, the image is obtained by, first, adding all patches at their corresponding positions, and then element-wise dividing the resulted matrix by a weight

matrix, where the entries of the weight matrix are the number of patches that include the entries.

3.2. Proposed fusion method

The proposed fusing operator can be formulated as

$$i^F \triangleq \mathcal{F}\{\{i_k\}_{k=1}^K\} = \mathcal{M}\left\{\mathcal{R}\{\mathcal{L}\{\alpha_k; i_k, D\}_{k=1}^K\}\right\}.\{i_k\}_{k=1}^K . \quad \textbf{(19)}$$

Equation (19) shows that the local fusion operation is performed by applying the mask operator $\mathcal{M}$ to the K sets of source image patches, $\{i_k\}_{k=1}^K$, where the mask operator is obtained using the selection operator $\mathcal{R}\{\cdot\}$.

The operator $\mathcal{L}\{\cdot\}$ yields the sparse coefficient vectors $\{\alpha_k\}_{k=1}^K$ by sparse approximation of image patches $\{i_k\}_{k=1}^K$ over the coupled dictionary $D = [D^F, D^B]$. This operator is found by solving the following problem

$$\mathcal{L}\{\alpha_k; i_k, D\} = \underset{\alpha_k}{\operatorname{argmin}} \|\alpha_k\|_1 \quad s.t. \ \|D.\alpha_k - i_k\|_2^2 \leq \epsilon \quad \textbf{(20)}$$

where ϵ is the maximum tolerance parameter. Here problem (20) is solved using the OMP (Aharon et al., 2006). To reduce the computational cost and ignore the effect of luminance in the selection operator the mean values of extracted source image patches are removed $\left(i_k = i_k - mean(i_k)\right)$ before sparse approximation. The zero mean image patches are only used for the sparse approximation, not in the reconstruction phase.

According to second line of (19), after obtaining $\{\alpha_k\}_{k=1}^K$, the selection operator $\mathcal{R}\{\cdot\}$ is applied to find the sparse coefficient vectors that represent the most focused image patch i^F, and their corresponding index k^F.

Based on the max-l^1-norm rule, the sparse vectors with larger l^1-norm are approximated with higher activity level and represent the image patches with the most focus level (Yang et al., 2010). The proposed selection operator calculates the weighted l^1-norm. The weight for the entries that corresponds to atoms from the focused subspace is greater than the weight that corresponds to the entries that are related to the blurred subspace. It leads to a significantly more accurate decision map. The formulation of such a selection operator is as follows

$$k^F = \mathcal{R}\{\{\alpha_k\}_{k=1}^K\} = \underset{k}{\text{argmax}}\left\{(1-\gamma).\left\|\mathrm{P}_{\alpha_k}(D^F)\right\|_1 + \gamma.\left\|\mathrm{P}_{\alpha_k}(D^B)\right\|_1\right\} \quad \textbf{(21)}$$

where $(1-\gamma)$ and γ are the weights corresponding to the blurred and focused subspaces, respectively, $0 \le \gamma \le 0.5$, and $\mathrm{P}_{\alpha_k}(D^F)$ and $\mathrm{P}_{\alpha_k}(D^B)$ are the projection of the vector α_k onto the subspaces D^F and D^B, respectively. Thus, this operator, in addition to the l^1-norm, also utilizes the information about the distribution of sparse vectors entries over the two subspaces as a comparison tool. In this way, the sparse vectors with higher activity levels and larger coefficient in the focused subspace are found as the most focused vectors.

Then, the local all-in-focus image patches are found by applying the mask operator $\boldsymbol{\mathcal{M}}$ given as

$$i^F = \mathcal{M}\{m, \{i_k\}_{k=1}^K\} = \sum_{i=1}^{K} m_k.\, i_k \quad \textbf{(22)}$$

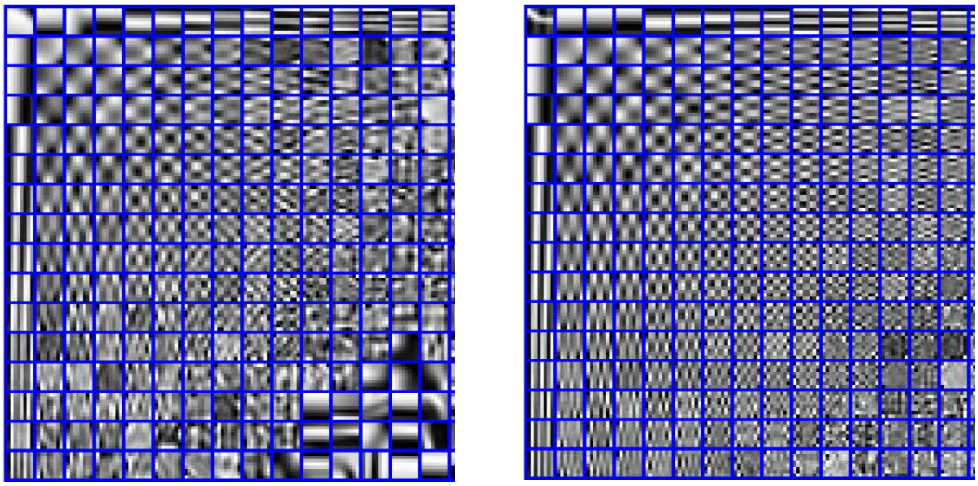
where $m = [m_1, \dots, m_k]^T$ is the mask vector with elements defined as

$$m_k = \begin{cases} 1, & if \quad k = k^F \\ 0, & otherwise. \end{cases} \quad \textbf{(23)}$$

Repeating the same procedure for all sets of K corresponding source image patches, across the whole source images, all local all-in-focus patches are obtained. Then, the all-in-focus image I^F can be formed by patch-wise reconstruction of them.

3.3. Coupled Dictionary Learning

The dictionaries D^F and D^B can be learned separately using a dictionary learning method such as K-SVD (Aharon et al., 2006). The separately learned dictionaries, as it is shown in Figure 12, contain either repeated or uncorrelated atoms. Thus, the proposed dictionary learning method aims to produce a couple of dictionaries, so that by providing the correlation between their atoms, each atom in D^B represents a blurred version of its corresponding atom in D^F.



(a) (b)

Figure 12. Separately learned dictionaries: (a) blurred dictionary D^B and (b) focused dictionary D^F.

In the next chapter, it is practically shown that the coupled dictionary learned by the proposed method significantly improves the fusion results while reducing the computational cost.

The correlation between atoms is ensured by enforcing a similar sparse representations of both training datasets (focused and blurred) during dictionary learning. The corresponding optimization problem can be formulated as

$$\min_{D^F, D^B, \Gamma} \|X^F - D^F\Gamma\|_F^2 + \|X^B - D^B\Gamma\|_F^2 \quad s.t. \quad \|\Gamma_i\|_0 < T_0 \,, \forall i \tag{24}$$

where Γ is the sparse coding matrix (the same for both dictionaries) and T_0 is the sparsity constraint. Thus, updates of the dictionaries can be performed separately using the same sparse coding matrix. In this work, K-SVD is used for the dictionary update phase. The diagram of the proposed coupled dictionary learning method is given in Figure 13.

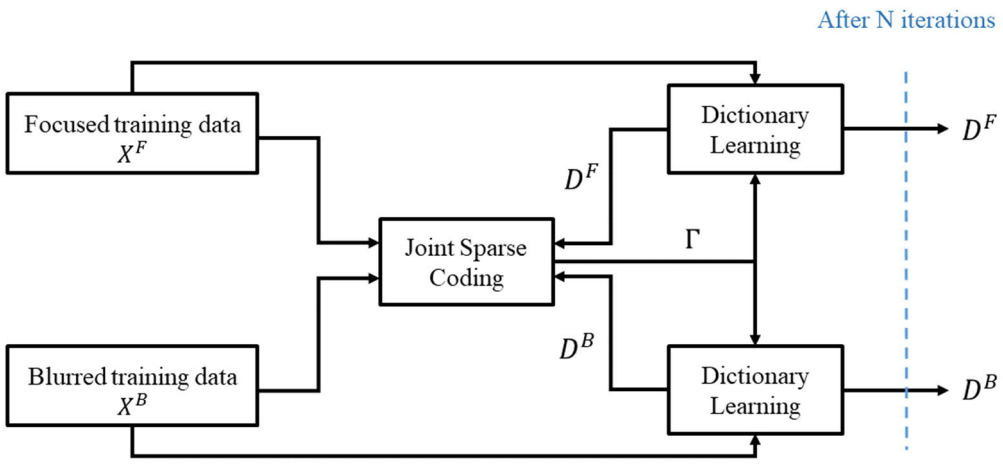


Figure 13. The diagram of the proposed coupled dictionary learning method. The algorithm is initialized using DCT dictionary.

The joint sparse coding problem is addressed by modifying the OMP method to solve

$$\min_{\Gamma} \|X^F - D^F\Gamma\|_F^2 + \|X^B - D^B\Gamma\|_F^2 \quad s.t. \quad \|\Gamma_i\|_0 < T_0 \,, \forall i \tag{25}$$

that can be rewritten as

$$\min_{\Gamma} \|X - D\Gamma\|_F^2 \quad s.t. \quad \|\Gamma_i\|_0 < T_0 \,, \forall i \tag{26}$$

where

$$X = \begin{bmatrix} X^F \\ X^B \end{bmatrix} \text{ and } D = \begin{bmatrix} D^F \\ D^B \end{bmatrix}. \quad \textbf{(27)}$$

The modified OMP, in each iteration, finds the atom d_m from D that matches the best the residue vector $x_i = \left[x_i^{F^T}, x_i^{B^T}\right]^T$, where x_i^F and x_i^B are first initialized by X_i^F and X_i^B. It is performed by solving the following optimization problem

$$d_m = \underset{d_m}{\text{argmax}} \|d_m^T . x_i\|_F^2 \quad , \forall m \quad \textbf{(28)}$$

where d_m^T stands for the transpose of vector d_m. After each iteration (adding a coefficient to Γ_i), x_i is updated using

$$x_i = X_i - D\Gamma_i \, . \quad \textbf{(29)}$$

When d_m is selected, the coefficient is found by solving

$$\min_{\Gamma_i^m} \left\|x_i^F - d_m^F \Gamma_i^m\right\|_F^2 + \left\|x_i^B - d_m^B \Gamma_i^m\right\|_F^2 \quad \textbf{(30)}$$

which is equivalent to solving

$$\min_{\Gamma_i^m} \|x_i - d_m \Gamma_i^m\|_F^2 \quad \textbf{(31)}$$

where $d_m = [d_m^F, d_m^B]^T$ and Γ_i^m is the m-th entry of Γ_i. Problem (30) can be solved using SVD. The algorithm iterates and adds new coefficients until either the number of coefficients reaches its maximum number T_0 or the error becomes less than ϵ. The error is computed here as

$$e = \|x_i\|_F^2 \, . \quad \textbf{(32)}$$

Alternating between K-SVD and the proposed modified OMP, the correlated coupled dictionary is found. The overall algorithm is summarized in Algorithm 1.

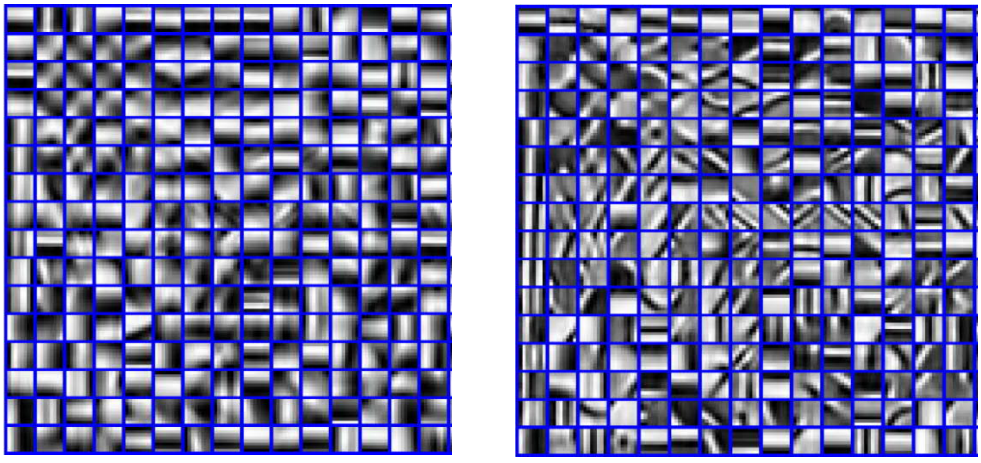
Algorithm 1. Joint Sparse Coding

Input: Focused and blurred training data set X_i^F and X_i^B, and $D = [D^F\ D^B]^T$ (D^F and D^B are initialized by DCT dictionary).

for $i = 1$ ... number of patches in X (Equation. (27))
  Set $x_i = X_i$;
  while $e > \epsilon$ and $\|\Gamma_i\|_0 < T_0$
    Find d_m using Equation. (28);
    Find Γ_i^m solving Equation. (31);
    Update x_i using Equation. (29) and the error using Equation. (32);
  end while
end for

Output: The common sparse representation matrix Γ

An example of dictionaries learned by the proposed method is shown in Figure 14, where the correlation between corresponding atoms can be clearly seen. Each two corresponding atoms represent the two (focused and blurred) versions of the same spectrum.



(a) (b)

Figure 14. Dictionaries learned by proposed method: (a) blurred dictionary D^B and (b) focused dictionary D^F.

4. SIMULATION RESULTS

In this chapter, the performance of the proposed fusion method is shown both visually and quantitatively, and compared to that of the state-of-the-art fusion methods. Also, the effects of the parameters such as the *maximum tolerance error* ϵ, *weighting parameter* γ in the proposed fusion rule (21), and *patch size d* on the fusion performance are assessed.

For quantitative evaluation, the four fusion quality measuring indexes that have been introduced in Subsection 2.5, namely $Q_{AB/F}$, NMI, SSIM and MSE, are used.

The results are compared to the following five well-known fusion methods.

- Discrete wavelet transform based image fusion (DWT) (Tian et al., 2012).
- Multi-focus image fusion using principal component analysis (PCA) (Raol & Naidu, 2008).
- Sparse representation choose-max based image fusion (SR-CM) (Yang et al., 2010).
- Sparse representation choose-max based image fusion via learned over-complete dictionary using K-SVD (SR-KSVD) (Aharon et al., 2006).
- Image fusion using sparse representation of focus measures (SR-FM) (Nejati et al., 2015).

4.1. Experimental setup

The implementation parameters of the mentioned fusion methods are as follows. In DWT method, the decomposition level of source images is 3 and the wavelet basis *"db1"* is used.

In all of the sparse representation based methods, the patch size is 8^2 and the dictionary size is 64×256 (64×512 for coupled dictionary), also the sliding step of 1 is used. The dictionaries are learned over 10 cycles.
For the proposed method, the weighting parameter γ is set between 0.4 and 0.45, and the maximum tolerance is $\epsilon = 4$. For the other methods ϵ is 0.1. In the SR-FM, the Laplacian energy (as focus measure) and max-pooling (for aggregation of sparse representations) are used. Also, for clearness of the presentation of the algorithms´ performances based on fair comparison, in the SR-FM, the reconstruction is performed using the same method as in the other sparsity based methods, as the segmentation is off the focus of this master's thesis.

All implemented algorithms are run on a PC using Inter(R) Xenon(R) 3.40GHz CPU.

For the performance comparison of the methods, six pairs of multi-focus images from the standard gray-scale dataset (Nejati et al., 2015) are used. The focused and blurred training data are produced by randomly selecting 20,000 image patches from all-in-focus and all-out-of-focus reference images that are manually found using the images from the same dataset. The dataset is shown in Figure 15.

The sizes of the source images Clocks, Doll and Pepsi are of the size 256^2. The source images Lab and Disk are of the size 480×640 and Jug has the size 256×384.

In Figure 15, the images in the first column (from the right) are the front focused images. In the second column, the images are back focused. In the third column, they are manually classified as all-in-focus, and in the fourth column, they are manually classified as all-out-of-focus. The images in the third and fourth columns are used as the focused and blurred training data. The manually found all-in-focus images are also used as reference for computing the reference-based fusion quality measurements.

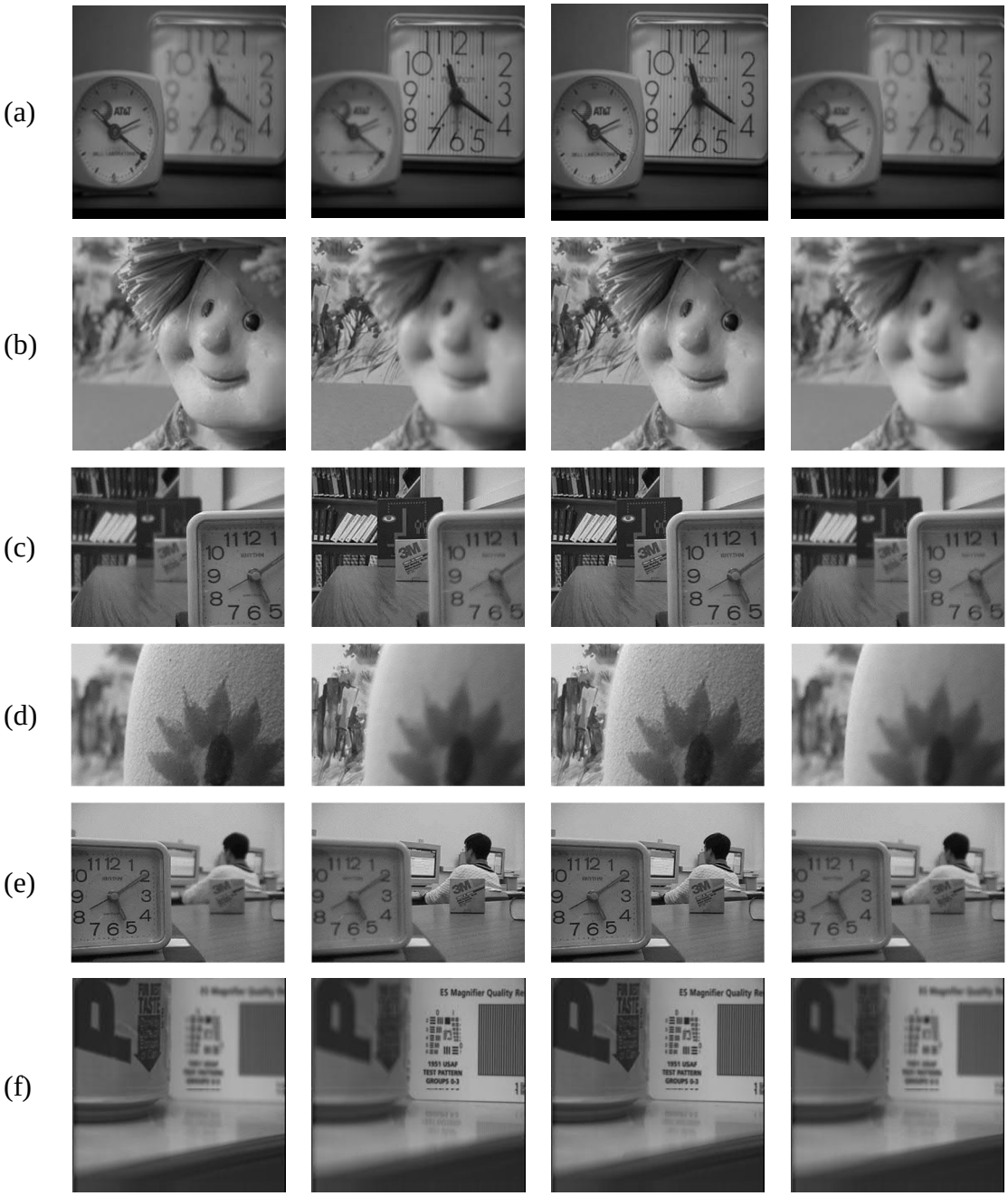


Figure 15. Source images: (a) Clocks, (b) Doll, (c) Disk, (d) Jug, (e) Lab, (f) Pepsi

4.2. Visual comparison

In Figures 16, 17 and 18, the visual fusion results for three pairs of images, respectively, Clocks, Doll and Pepsi are given.

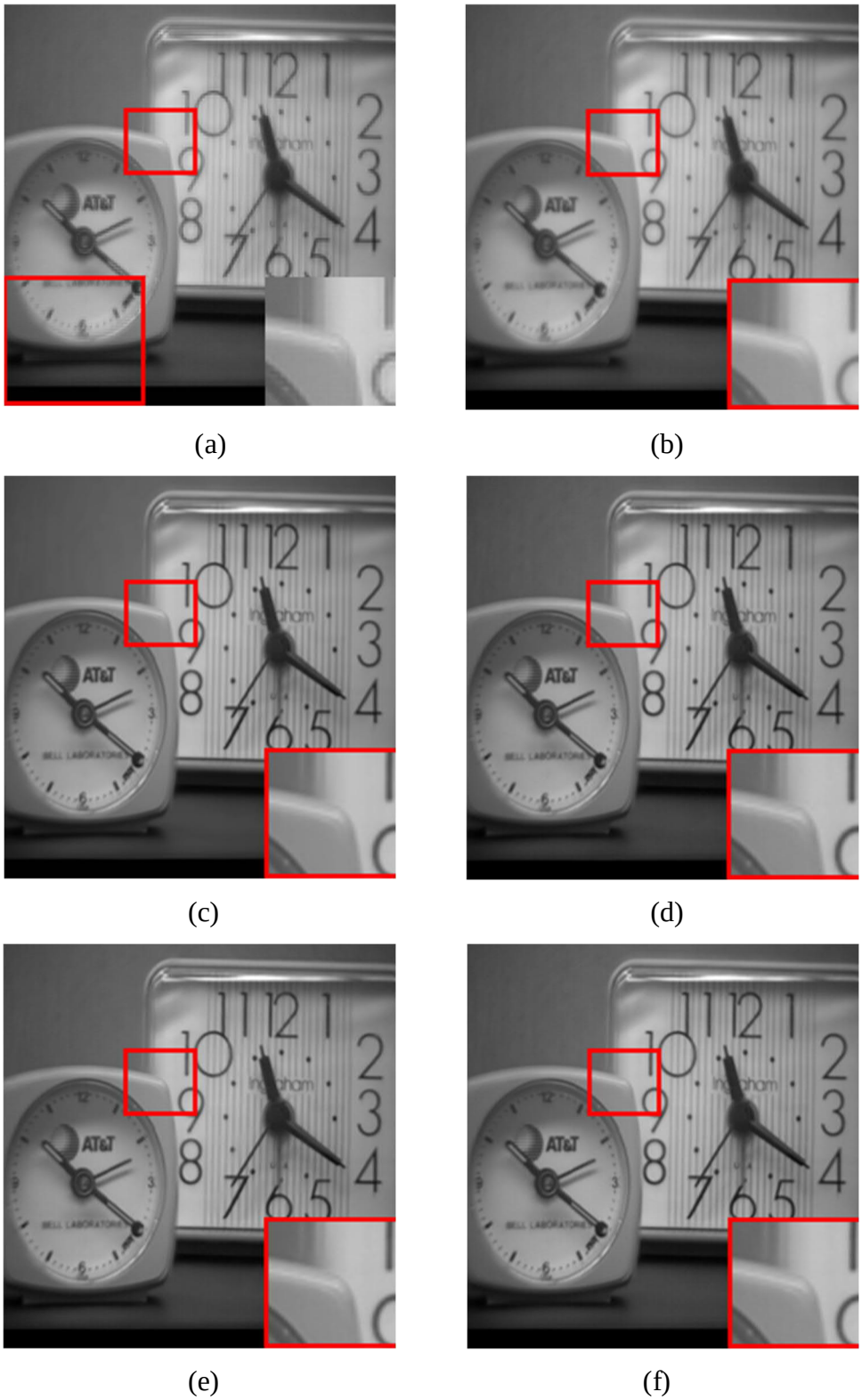
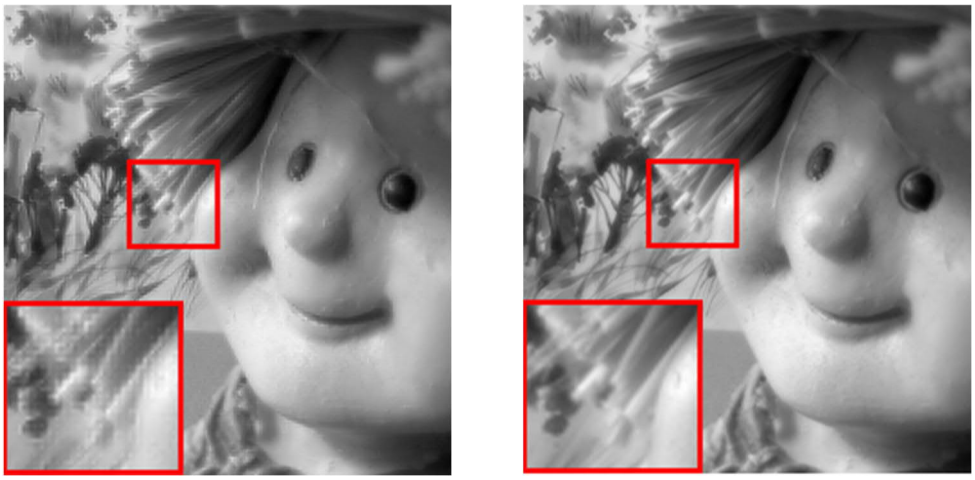
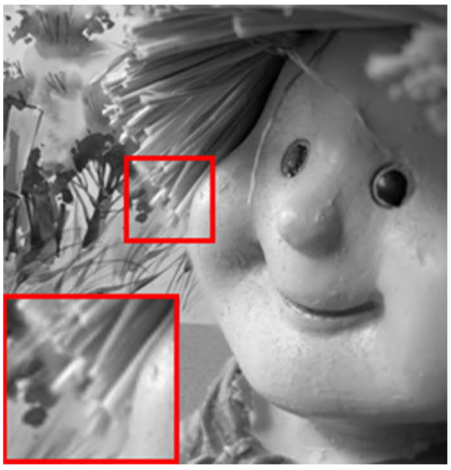


Figure 16. Fusion results for the source images "Clocks". (a) DWT, (b) PCA, (c) SR-CM, (d) SR-KSVD, (e) SR-FM, and (f) proposed

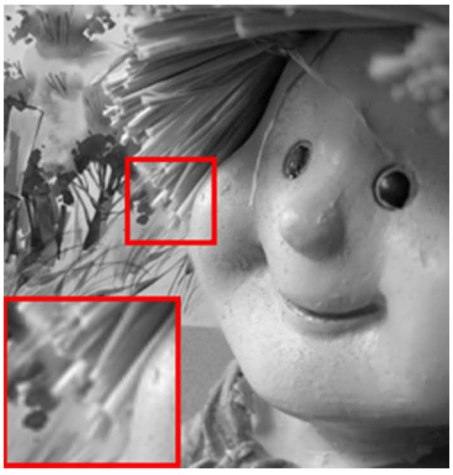


(a)

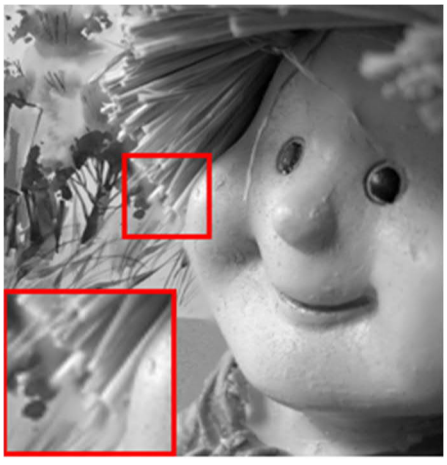
(b)



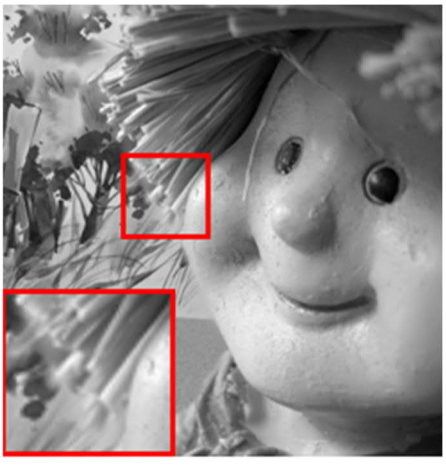
(c)



(d)

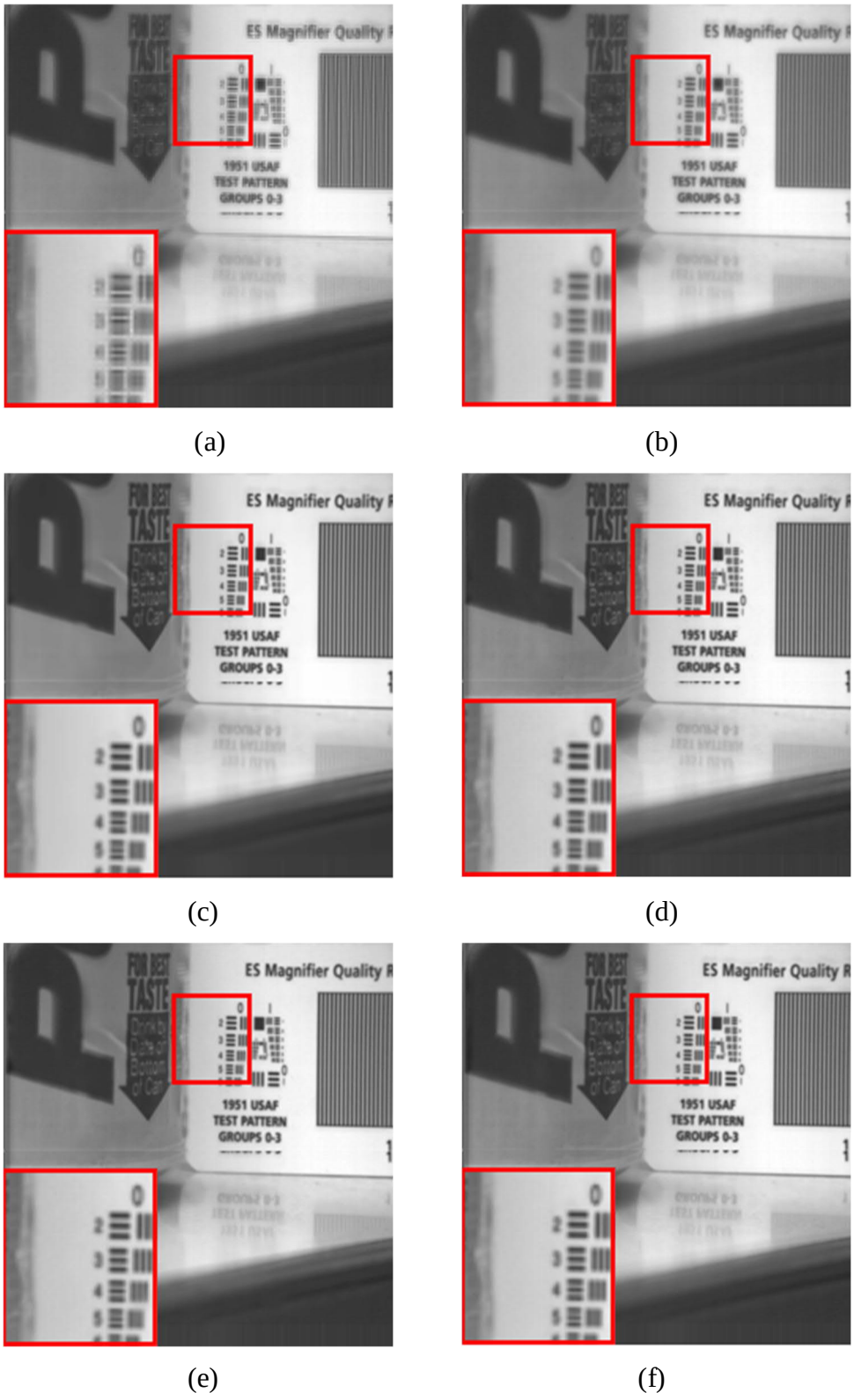


(e)



(f)

Figure 17. Fusion results for the source images "Doll", in the same order as Figure 16.



(a) (b) (c) (d) (e) (f)

Figure 18. Fusion results for the source images "Pepsi", in the same order as Figure 16.

Looking at the representative fusion results that are shown above, especially the details that are magnified at the bottom corners, it can be seen that the fused images obtained using the DWT contain blocking artifacts and are of a very low quality in terms of sharpness. The results produced by the PCA are also excessively blurred, although they are relatively smooth. Comparatively, these two methods have the weakest performances.

By carefully inspecting the results of the methods, the SR-CM, SR-KSVD, SR-FM and the proposed method, and considering the sharpness of edges, visibility of details and smoothness of the fused images, it can be concluded that the results obtained using the proposed method have the highest quality, while they are produced with less computational cost and require shorter running time compared to the rest of the competitive methods (see Table 1).

Table 1. Comparison of the running times of all methods tested for fusion of "Clocks", using parameters given in Subsection 4.1.

Methods	Feature Extraction (s)	Sparse Coding (s)	Fusion (s)	Total (s)
DWT	-	-	0.7219	0.7219
PCA	-	-	0.2571	0.2571
SR-FM	31.50	14.44	1.8798	47.8198
SR-CM	-	354.4512	1.4312	355.8824
SR-KSVD	-	324.9160	1.2708	326.1868
Proposed	-	12.8243	1.4859	14.3102

4.3. Quantitative comparison

To ensure that the proposed method yields the best results, a quantitative comparison among all methods tested over all six pair of source images is given in Table 2. The results show that regarding the index $Q_{AB/F}$, the proposed method has the best performance in all cases, which means that the edge information from the source images is nicely transferred

into the fused image. The best NMI results for the proposed method also show the high fidelity of the fusion process in terms of pixel by pixel intensities, which is admitted by the best MSE results in five out of six cases.

The PCA shows the best SSIM results in three cases, which admits the smoothness of its results, despite of having low performances in other measures because of the blurred outputs.

In addition, as mentioned before, to show the efficiency of the proposed method in terms of its computational cost, the run times of all methods tested for fusion of the source images "Clocks" are compared in Table 1.

Finally , to show the effectiveness of using a coupled dictionary instead of separate dictionaries learned over focused and blurred training data, the indexes $Q_{AB/F}$ and NMI for the two cases are compared in Figure 19, noting that the runtimes of dictionary learning in the coupled and separately cases are 775 and 1129 seconds, respectively. It can be seen that a coupled dictionary learning leads to a better results in terms of both $Q_{AB/F}$ and NMI.

4.4. Effects of parameters

In this subsection, the effects of three main parameters, the weighting parameter γ, the maximum tolerance error ϵ, and the patch size d, on the fusion performance are investigated. For this purpose, the proposed algorithm is run over all source images and the results are averaged for finding the quality measures NMI and $Q_{AB/F}$.

In Figure 20, the average results obtained over different values of γ are presented. As the figure shows, the best results are achieved for γ between 0.4 and 0.45.

Table 2. Objective evaluation of fusion results. Results are ranked by colors: **Red**: Best, **blue**: Second best, and **green**: third best.

Methods	Measures	Clocks	Lab	Pepsi	Disk	Jug	Doll
DWT	*NMI*	0.9847	1.0027	1.0079	0.8129	0.8497	0.8553
	$Q_{AB/F}$	0.6600	0.5487	0.6587	0.5102	0.5048	0.6184
	SSIM	0.9403	**0.9372**	**0.9362**	**0.9068**	0.8871	0.9211
	MSE	32.2172	60.1514	43.2703	94.3113	51.6737	46.3669
PCA	*NMI*	1.0276	1.0270	1.0610	0.8372	0.8854	0.8965
	$Q_{AB/F}$	0.6939	0.5651	0.6752	0.5352	0.5083	0.6335
	SSIM	**0.9572**	**0.9468**	**0.9351**	**0.9226**	0.9048	0.9418
	MSE	24.8221	54.4139	29.4576	80.9688	45.1700	36.2968
SR-FM	*NMI*	1.1100	1.0573	**1.1764**	0.8878	0.9490	**1.0935**
	$Q_{AB/F}$	**0.7462**	0.6900	**0.7577**	0.6380	0.7174	0.7380
	SSIM	**0.9451**	0.8153	0.9296	0.8325	0.9490	**0.9862**
	MSE	5.5989	12.0835	5.8016	30.7394	19.3786	7.4314
SR-CM	*NMI*	**1.1118**	**1.1079**	1.1063	**0.9460**	**1.0630**	**1.0547**
	$Q_{AB/F}$	0.7301	**0.7058**	0.7290	**0.7653**	**0.7656**	**0.7402**
	SSIM	0.8813	0.7843	0.8229	0.8367	**0.9609**	0.9817
	MSE	**1.8879**	**7.4700**	**3.9962**	**11.0090**	**3.3700**	**5.3617**
SR-KSVD	*NMI*	**1.1658**	**1.1235**	**1.1685**	**0.9821**	**1.1417**	1.0517
	$Q_{AB/F}$	**0.7557**	**0.7295**	**0.7613**	**0.7206**	**0.7766**	**0.7454**
	SSIM	**0.9527**	**0.8400**	0.9258	0.8667	**0.9925**	**0.9888**
	MSE	**1.6457**	**7.7026**	**3.6903**	**11.0777**	**3.2798**	**3.4843**
Proposed	*NMI*	**1.1742**	**1.1652**	**1.1802**	**1.0067**	**1.1653**	**1.1496**
	$Q_{AB/F}$	**0.7604**	**0.7334**	**0.7651**	**0.7229**	**0.7811**	**0.7498**
	SSIM	0.9340	0.8337	**0.9413**	**0.8773**	**0.9965**	**0.9942**
	MSE	**2.1808**	**5.1499**	**3.0203**	**7.1657**	**1.0052**	**2.5981**

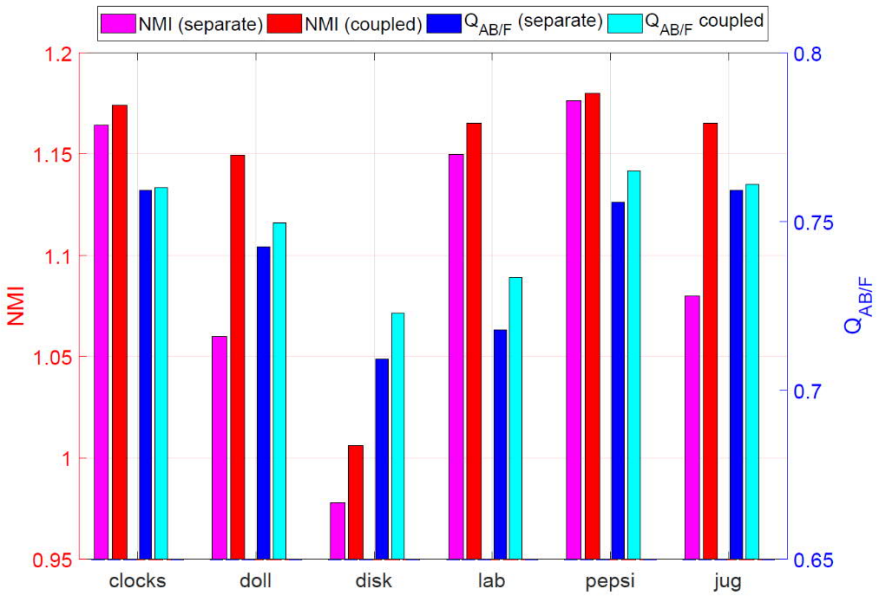


Figure 19. Comparison of fusion results using proposed method over a coupled dictionary and separately learned dictionaries.

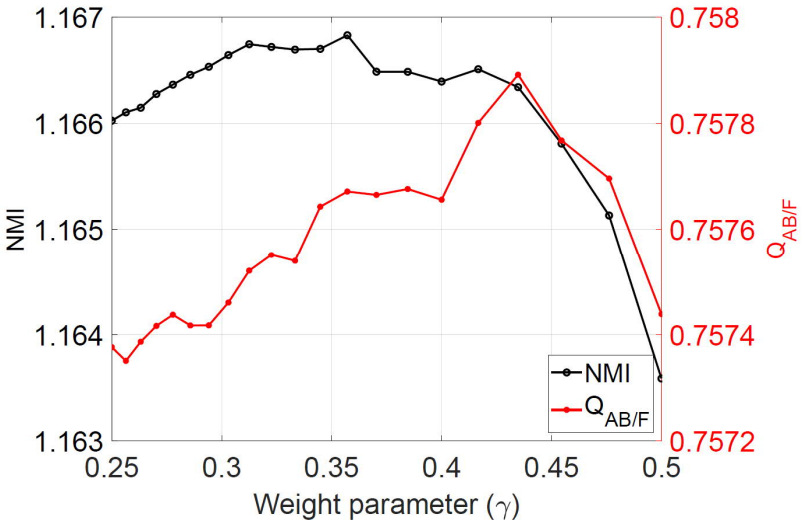


Figure 20. Effect of weighting parameter (γ) on fusion performance

The effect of maximum tolerance error ϵ on the fusion performance is shown in Figure 21. It can be seen that the optimal performance is obtained for values of ϵ between

4 and 5. These values of ϵ lead to the best compromise between the sparsity and accuracy of the sparse approximations, which is essential in order to perform precise fusion.

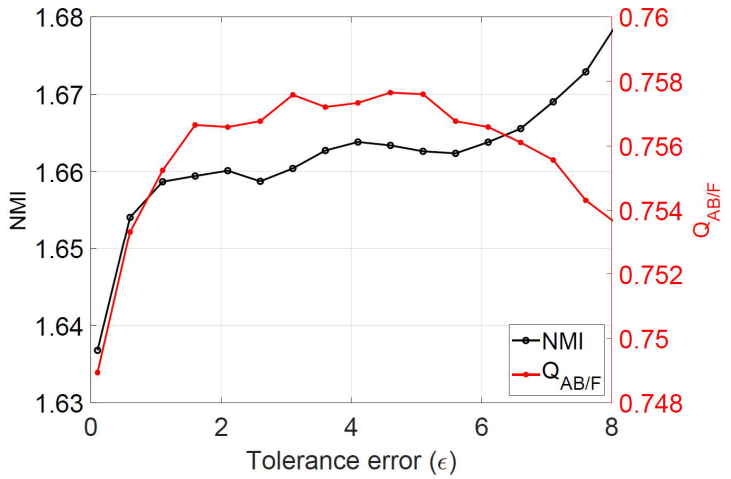


Figure 21. Effect of tolerance error (ϵ) on fusion performance.

The third studied parameter is the patch size *d*. The results are presented in Figure 22 and show that by increasing the patch size, the performance of the proposed method is improved. However, as it has been discussed before, processing large patches needs extremely high computational power and time. Thus, our earlier experiments have been run using *d=8*.

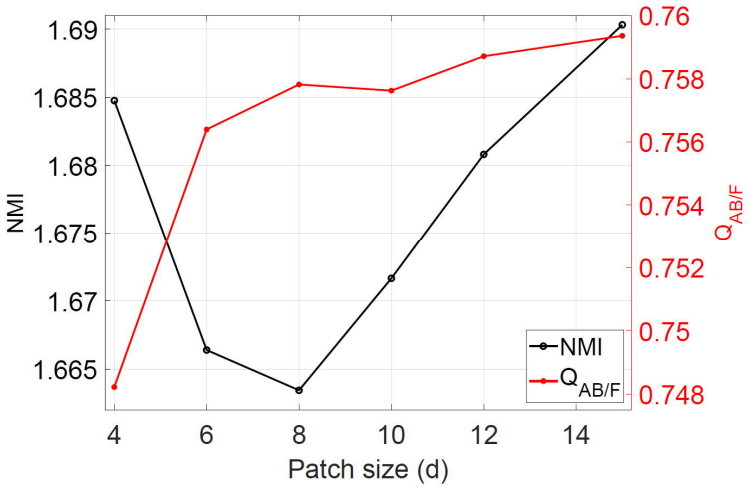


Figure 22. Effect of patch size (d) on fusion performance.

Another observation that can be made based on Figures 21 and 22 is that the NMI increases for larger ϵ values and smaller patch sizes, while $Q_{AB/F}$ is decreasing. It is because of the failure in the selection operation and as a consequence returning a fused image consisting mostly of only one of the source images.

5. CONCLUSION AND FUTURE WORK

5.1. Conclusion

In this master's thesis, a fusion method to fuse multiple multi-focus images into one smooth all-in-focus image has been developed. The proposed method works based on the sparse representations of multi-focus inputs using a pair of over-complete dictionaries, representing the focused and blurred feature spaces.

In addition, a coupled dictionary learning algorithm has been developed that improves the fusion performance significantly by ensuring correlation between atoms of two dictionaries. It also improves the performance of the fusion process in terms of computational requirements.

It has been shown that the proposed fusion method using a couple of dictionaries (focused and blurred) learned by the proposed coupled dictionary learning algorithm produces smooth images that contain the highest amount of focused information from the source multi-focus images, as compared to the state of the art fusion methods including the sparse representation based methods that use a single over-complete dictionary learned to represent only the focused features.

In addition to the algorithm development and results comparison, the concept of focus in lens optics has been briefly explained and the most well-known multi-focus image fusion methods have been introduced. Also, the essential key topics in the research field of image sparse representation have been reviewed.

5.2. Future work

The most important information in an image is around the edges, where different objects need to be distinguished from each other. Thus, using the edge information of the source

images, a fusion method can improve its efficiency by exploiting more accurate schemes near edges (e.g. larger window size) and using less computationally costly techniques for the other areas in that image.

The idea to overcome the constraint of the patch size is to extract the focus information using the inter-patch information in a neighborhood (of size, e.g., 3×3). In other words, instead of only using pixel intensities, some more measures (e.g. the Laplacian energy of the patches or the l_1-norm of their sparse representations) from all patches in a neighborhood can be used for calculating the focus measure of that neighborhood. Thus, the outliers (the scattered error in the decision map) can be diminished and the blurredness and shadowing effects (around focus borders) can be avoided. Alternatively, using a down sized version of the source images can help to address the same problem. However, in both cases the precision of the fusion needs to be at the pixel level around the edges. This can be addressed by refining the obtained mask so that its borders are consistent with the edges of source images.

REFERENCES

Aharon, M., Elad, M., & Bruckstein, A. (2006). K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation. *IEEE Transactions on Signal Processing, Vol. 54*, 4311-4322.

Bertero, M., Poggio, T. A., & Torre, V. (1988). Ill-Posed Problems in Early Vision. *Proceedings of the IEEE, Vol. 76* (pp. 869 - 889). IEEE.

Calhou, V. D., & Adali, T. (2009). Feature-Based Fusion of Medical Imaging Data. *IEEE Transaction on Information Technology in Biomedicine, Vol. 13*, 711-720.

Chartrand, R., & Yin, W. (2008). Iteratively Reweighted algorithms For Compressive Sensing. *IEEE International Conference on Acoustics, Speech and Signal Processing*, 3869–3872.

Chen, S. S., Donoho, D. L., & Saunders, M. A. (2001). Atomic Decomposition by Basis Pursuit. *SIAM Review, Vol. 43*, 129-159.

Elad, M. (2007). Optimized Projections for Compressed Sensing. *IEEE Transactions on Signal Processing, Vol. 55*, 5695-5702.

Elad, M., & Yavneh, I. (2009). A Plurality of Sparse Representations Is Better Than the Sparsest One Alone. *IEEE Transactions on Information Theory, Vol. 55*, 4701 - 4714.

Engan, K., Aase, S. O., & Husoy, J. H. (1999). Method of Optimal Directions for Frame Design. *International Conference on Acoustics, Speech, and Signal Processing, Vol. 5* (pp. 2443–2446). Phoenix: IEEE .

Fisher, R., Perkins, S., Walker, A., & Wolfart, E. (2003). *Sobel Edge Detector*. Retrieved from Image Processing Learning Resources: http://homepages.inf.ed.ac.uk/rbf/HIPR2/sobel.htm

Hossny, M., Nahavandi, S., & Creighton, D. (2008). Comments on 'Information Measure for Performance of Image Fusion'. *Electronics Letters , Vol. 44*, 1066-1067.

Li, S., & Yang, B. (2008). Multifocus Image Fusion Using Region Segmentation and Spatial Frequency. *Image and Vision Computing, Vol. 26*, 971-979.

Liu, Y., Liu, S., & Zengfu, W. (2015). Multi-Focus Image Fusion with Dense SIFT. *Information Fusion, Vol. 23*, 139-155.

Mairal, J., Bach, F., Ponce, J., & Sapiro, G. (2010). Online Learning for Matrix Factorization and Sparse Coding. *Journal of Machine Learning Research, Vol. 11*, 19-60.

Mallat, S. G., & Zhang, Z. (1993). Matching Pursuits With Time-Frequency Dictionarie. *IEEE Transaction on Signal Processing, Vol. 41*, 3397-3415.

Nejati, M., Samavi, S., & Shirani, S. (2015). Multi-Focus Image Fusion Using Dictionary-Based Sparse Representation. *Information Fusion, Vol. 25*, 72-84.

Nencini, F., Garzelli, A., Baronti, S., & Alparone, L. (2007). Remote Sensing Image Fusion Using The Curvelet Transform. *Information Fusion, Vol. 8*, 143-156.

Nichol, J., & Wong, M. S. (2005). Satellite Remote Sensing for Detailed Landslide Inventories Using Change Detection and Image Fusion. *International Journal of Remote Sensing, Vol. 26*, 1913–1926 .

Pajares, G., & Cruz, J. M. (2004). A Wavelet-Based Image Fusion Tutorial. *Pattern Recognition, Vol. 37*, 1855-1872.

Pati, Y., Rezaiifar, R., & Krishnaprasad, P. (1993). Orthogonal Matching Pursuit: Recursive Function Approximation with Applications to Wavelet Decomposition. *Asilomar Conference on Signals, Systems and Computers* (pp. 40-44). Pacific Grove: IEEE .

Pradnya P., M., & Ruikar, S. D. (2013). Image Fusion Based on Stationary Wavelet Transform. *International Journal of Emerging Technology and Advanced Engineering*, 99-101.

Raol, J., & Naidu, V. (2008). Pixel-level Image Fusion using Wavelets and Principal Component Analysis. *Defence Science Journal, Vol. 58*, 338-352.

Tian, J., & Chen, L. (2012). Adaptive Multi-Focus Image Fusion Using a Wavelet-Based Statistical Sharpness Measure. *Signal Processing, Vol. 92*, 2137-2146.

Tibshirani, R. (2011). Regression Shrinkage and Selection via The LASSO: A Retrospective. *Journal of Royal Statistical Society, Vol. 58*, 273-282.

Tropp, J. A. (2006a). Algorithms for Simultaneous Sparse Approximation. Part II: Convex Relaxation. *Signal Processing*, 589-602.

Tropp, J. A., Gilbert, A. C., & Strauss, M. J. (2006b). Algorithms for Simultaneous Sparse Approximation. Part I: Greedy pursuit. *Signal Processing, Vol. 86*, 572-588.

Wan, T., Canagarajah, N., & Achim, A. (2008). Compressive Image Fusion. *IEEE International Conference on Image Processing* , 1308-1311.

Wan, T., Qin, Z., Zhu, C., & Liao, R. (2013). A Robust Fusion Scheme for Multifocus Images Using Sparse Features. *International Conference on Acoustics, Speech and Signal Processing* (pp. 1957-1961). Vancouver: IEEE.

Wang, Z., & Bovik, A. S. (2004). Image Quality Assessment: From Error Visibility to Structural Similarity . *IEEE Transactions on Image Processing, Vol. 13*, 600 - 612.

Wright, J., Ganesh, A., Sastry, S. S., & Ma, Y. (2009). Robust Face Recognition via Sparse Representation . *IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 31*, 210-227.

Xydeas, C., & Petrovic, V. (2000). Objective Image Fusion Performance Measure . *Electronics Letters, Vol. 36*, 308 - 309 .

Yang, B., & Li, S. (2010). Multifocus Image Fusion and Restoration With Sparse Representation. *IEEE Transactions on Instrumentation and Measurement, Vol. 59*, 884-892.

Yuan, X. T., Liu, X., & Yan, S. (2012). Visual Classification With Multitask Joint Sparse Representation . *IEEE Transactions on Image Processing, Vol. 21*, 4349 - 4360 .

Zhang, H., Li, J., & Huang, Y. (2014). A Nonlocal Weighted Joint Sparse Representation Classification Method for Hyperspectral Imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing , Vol. 7*, 2057-2066.

Zhang, Q., Liu, Y., Blum, R. S., Han, J., & Tao, D. (2018). Sparse Representation Based Multi-Sensor Image Fusion for Multi-Focus and Multi-Modality Images: A Review. *Information Fusion, Vol. 40*, 57-75.

Zhou, Z., Li, S., & Wang, B. (2014). Multi-Scale Weighted Gradient-Based Fusion for Multi-Focus Images. *Information Fusion, Vol. 20*, 60-72.

APPENDIX 1. WELL/ILL POSED PROBLEM

The mathematical term *well-posed*, defined by *Jacques Hadamard*, is applied to a linear problem whose solution has three characteristics:

a) Existence: the problem should have a solution.
b) Uniqueness: the solution should be unique.
c) Continuity: the small changes in input should be continuously followed by changes in the solution.

A problem that does not fulfil requirements above is known as *ill-posed* (Bertero, Poggio, & Torre, 1988).