Dictionary Learning

Intensity Normalization For Generating Synthetic Tissue Contrast In Brain MRI Images

Y.Luktuke

Dictionary Learning

Intensity Normalization For Generating Synthetic Tissue Contrast In Brain MRI Images

by

Y. Luktuke

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on Thursday January 26, 2017 at 09:30 AM.

Student number:	4402952	
Thesis committee:	Prof. Dr. M. Loog,	TU Delft UD PRB, Supervisor
	Dr. E. A. Hendriks,	TU Delft UHD PRB-INSY-EWI
	Dr. Ir. R. Heusdens,	TU Delft UHD ME(CAS)

An electronic version of this thesis is available at `http://repository.tudelft.nl/`.

TU Delft

Preface

My interest in image processing and machine learning began in the final semester of my Third Year of Bachelor studies in India. I feel that everything that has happened since then has led me to this moment where I am preparing to defend my Master Thesis. Towards the end of my First Year at TU Delft I was presented with a rather interesting research topic, if machine learning methods could be used to learn a transformation between MRI images for the task of making more consistent datasets with these images. While the application is essentially from the field of medical image processing, it blends theory and models that are common to the areas of image processing and machine learning. This made the project interesting to me, while allowing me to stay connected with my engineering background as well. I hope that this report conveys to the reader how my thoughts and ideas progressed over the better part of the previous year. It is a great feeling when I reflect back how this project has progressed, from having a suggested research topic to this phase where I am making the final steps towards obtaining my degree as a Master of Science in Electrical Engineering.

Y. Luktuke
Delft, January 2017

Acknowledgements

There are quite a few people that I would like thank who have been instrumental in me reaching this point. First of all my family back home, my father Yoginish Luktuke, my mother Yamini and my brother Yadvindra who have provided financial and moral support through the good and challenging times. This has been possible only because of each one of you and your patience with me. My Supervisor Prof.Marco Loog for the various inputs, suggestions and comments, but particularly for keeping me focused on my research while still allowing me the freedom to study those areas of the project which I found interesting. Thanks are also in order for Prof.Jan van Gemert who suggested the software on which most of my project is based during a casual coffee break as well as to Wouter Kouw who initially suggested the research topic and who has been present to answer several of the little doubts that I've had during the project. One more important person to thank back in India is Prof.Alwin Anuse whose ideas, projects and lectures got me started in the area of machine learning and image processing in the first place. Last but certainly not the least, all of my friends and colleagues with whom I have interacted over the past two years at TU Delft. Your inputs and suggestions have kept me going, thank you to all of you.

Y. Luktuke
Delft, January 2017

Contents

List of Tables

List of Figures

1 Introduction

1.1. MRI Images And Image Segmentation

Magnetic Resonance Imaging (MRI) is a non-invasive procedure that is considered as the gold standard for obtaining images of the Brain [12]. Analysis methods such as image segmentation and registration have allowed us to study the Brain in great detail using such MRI images obtained. Such studies have been used to further our understanding or normal aging, brain diseases such as Multiple Sclerosis and Alzheimer's Disease [11]. Sometimes such analysis methods also enable us to detect tumours and other abnormalities present in the brain of a patient.
Image Segmentation is the procedure of highlighting areas within an image having common properties such as intensity. When applied to Brain MRI images it is often used to identify separate tissue classes present within each brain [11]. This often allows us to study brain images by locating regions of interest and observing the changes that occur in these regions over a period of time in better detail.

1.2. Problem Statement

Most image segmentation methods in practice use image intensities as their primary features [11]. However, the intensities present in MRI images do not have a specific numeric meaning and the values obtained differ as per the pulse sequence, its parameters and external effects such as the manufacturing of the scanner and its calibration [11]. As a result the intensities in MRI images and their distribution vary considerably when acquired with different scanners, even images of the same patient. Given a set of images of a patient obtained by pooling data from a multi-site multi-scanner study, this affects segmentation which cannot produce consistent results on the images considered [11], [12].
A fundamental property of MRI images that changes due to the differences in intensities is the tissue contrast in each image. The inconsistent results of image segmentation in Brain MRI images can be attributed in part to the differences in tissue contrast [12]. It has been reported earlier, that generating synthetic tissue contrast in MRI images can be used to obtain more consistent results with image segmentation [11], [12]. The MIMECS method developed in these works uses over-complete dictionaries obtained directly from a pair of atlas images for generating synthetic tissue contrast in test images.
However this limits the applicability of the method since a pair of atlas images needs to be known for every set of test images. Hence it is worthwhile to verify if the intensities in a pair of MRI images termed as the source and target images can be normalized without the need of an atlas pair. Dictionary Learning is chosen for this task of intensity-normalization since it models a corresponding pair of image patches as sparse combinations from a pair of over-complete dictionaries. The advantage of considering dictionary learning is that the pair of over-complete dictionaries can be learnt from the pair of images themselves without requiring an external input. It can be expected that normalizing the intensities within a source image with respect to the intensities present in the target image will reduce the difference in tissue contrast present between each pair of images. This is the question being researched in this Master Thesis; whether dictionary learning be used for intensity-normalization between a pair of source and target images such that the output image obtained has similar tissue contrast as that present in the target image.

1.3. Datasets Used

Before considering any real-life MRI image data it is helpful to evaluate the performance of any intensity-normalization method on synthetic images whose intensities can be controlled very accurately. These images are known in literature as ground truth images [11], [12]. In Brain MRI studies, such ground truth images can be obtained from the Brainweb: Simulated Brain Database (SBD) [5]. Each image is a phantom generated using a specific pulse sequence, T_1, T_2 or P_D. Other parameters such as slice thickness (t), noise percentage (n) and intensity non-uniformity (rf) can be controlled very accurately using the online tool on [5].
Using these SBD images two synthetic datasets were created for evaluating the performance achieved with the chosen intensity normalization method. This is consistent with experimental results in [11] and [12]. An added advantage of using these ground truth images is that each phantom image is generated using the same geometric system. Hence a pair of voxels from two images display the intensity of the same tissue obtained using the corresponding pulse sequence. In other words, a pair of synthetic images obtained from [5] do not need registration as a pre-processing stage.

1.4. Software Used

All the experiments reported for this project were performed in the MATLAB R2015b environment [10] with all available toolboxes. The dictionary learning algorithms were implemented with the freely available toolboxes, K-SVD Toolbox from [13] and the SPAMS Optimization Toolbox from [7]. Additional software was needed from the MATLAB Central File Exchange such as [3] and [1].

1.5. Overview

This report is organized in the following manner, Chapter 2 focuses on prior work related to improving the consistency of segmentation results on Brain MRI images obtained from different scanners. Following this Chapter 3 describes the sequence of reasoning followed in selecting the most appropriate dictionary learning algorithm for the task of intensity-normalization between a pair of source and target images. Chapter 4 details the experiments performed on the synthetic datasets mentioned above using the chosen dictionary learning method. It also analyses these results in the form of a discussion. Chapter 5 concludes this report by offering a few suggestions and improvements to consider in any future work related to this research topic.

2

Prior Work

2.1. Supervised Segmentation

In order to improve the consistency of image segmentation on Brain MRI images obtained from multiple scanners, several works have been presented previously in literature. One important class of methods is supervised segmentation of brain voxels. These methods use an appropriate set of features to train a classifier, which produces an output indicating the class (label) to which a single voxel/group of voxels belongs to. The following section describes two works presented by independent authors, which have been reported to achieve consistent segmentation on multi-site multi-scanner Brain MRI images.

2.1.1. Using Feature-Space Transformation for Supervised Segmentation

The work of Opbroek et al., [15], achieves consistent segmentation for a set of Brain MRI images by defining a transformation between the voxels of a pair of images considered as the source and target respectively. The main details of this method are explained below:

1. Source and Target Images are registered appropriately, in order to obtain correspondence between voxels and to eliminate geometric distortions which usually occur in images from different scanners.
2. Each source voxel is represented by features from a certain source distribution F_s, and each target voxel is represented similarly using features from a target distribution F_t.
3. The method of Feature-Space Transformation then takes place from the distribution F_s to F_t by locating the k-closest source voxels and replacing each feature value from F_s by the median of the difference between the corresponding target voxel and the k-closest neighbours.
4. Following the Feature-Space Transformation a supervised classifier can be trained on the transformed features and available labelled data.

2.1.2. Supervised Segmentation Based on Template Patch Library

In their work Tong et al.,[14] mention two methods of supervised classification based on a collection of atlas images. These are:

Sparse Representation Classification

1. For every patch from the training images, the closest patch from every image in a pre-defined atlas is located by means of a sum of squared intensity differences (SSD) [14].
2. A dictionary is constructed by arranging the atlas patches as its columns. This dictionary is fixed for every training voxel.
3. Then Elastic-Net Decomposition [14] is used to calculate the sparse representation coefficients for each patch centred around a training voxel using its pre-defined dictionary.
4. Using the available label for each atlas patch and the sparse coefficients obtained earlier, a classifier is trained on all the voxels from the training images.

5. In order to segment a test image, a similar strategy is followed by first locating the closest atlas patches. The sparse coefficients for each test patch are computed similar to the earlier method. These sparse coefficients are used as input to a classifier for estimating the label of the test voxel associated with the sparse coefficients.

Discriminative Dictionary Learning

1. Similar to the earlier method, the closest atlas patches are located for each patch in the training images.

2. However instead of using a pre-defined dictionary as before, a dictionary and supervised classifier are jointly learnt using the labels available from the atlas patches and the sparse coefficients of each training patch.

3. All patches from the test image are then labelled based on the trained classifier, similar to the method followed earlier.

One common problem with these methods and indeed with supervised segmentation is the need of accurately labelled atlas images which can be used for training the classifiers. In practice, a collection of atlas images is not available for every dataset of Brain MRI images. This makes the application of supervised segmentation difficult for general image datasets. In addition both methods presented in [14] require a dictionary for every voxel of each training image. This places a tremendous computational and memory requirement and makes the application of this method even more difficult. As a result, all the three methods presented above were not considered further in this project. Instead, attention was focused on methods which can be used for pre-processing Brain MRI images without the use of high quality atlas images.

2.2. Statistical Models and Registration Based Methods

As explained in [12], segmentation methods such as those which assume specific probability distribution functions (for the every tissue-class to be segmented) have been reported to have inconsistent results on data obtained from multiple scanners. This is because the estimates related to the various pdf parameters are intrinsically dependent on the tissue contrast present in MRI images. Since tissue-contrast varies between scanners due to the specific acquisition method used each time, so does the consistency of these methods.
In order to guarantee consistency in the segmentation results, one strategy is to obtain a correctly segmented atlas image and register each test image to this atlas using an appropriate registration method. However this method too relies on a high-quality segmented atlas image, and in addition depends upon the accuracy of the atlas-to-subject registration used. Hence such a strategy was also not considered any further.

2.3. The MIMECS Method [11],[12]

A common strategy to improve the consistency in segmentation is to normalize the intensities of the MRI images prior to segmentation using a fixed mapping. However this strategy also has a drawback, as it often requires manually chosen landmarks which are time-consuming to generate and lack robustness due to human-error. As a result of this and the previously discussed drawbacks, there is a need for an automatic pre-processing method for Brain MRI images. The goal of such a method would be to normalize the intensities of all images in a given dataset obtained from multi-site multi-scanner studies such that the tissue contrast is similar for all the images considered. The method of MR-Image Example Based Contrast Synthesis (MIMECS) presented in [11] and [12] achieves this by modelling image patches from an MRI image as the sparse linear combinations of the columns from fixed patch dictionaries. These are known as the source and target dictionaries and directly obtained from a fixed pair of atlas images (having known contrast) by decomposing the patches from each image as the columns of two dictionaries respectively. This strategy has been reported to normalize the intensities of images irrespective of the pulse sequence used during image acquisition as well as the different internal parameters used then. It has also been reported to achieve consistent segmentation on the processed images obtained after intensity-normalization. The main parts of this strategy have been described in the following sub-sections.

2.3.1. Compressed Sensing

The main idea behind compressed sensing is that most observed signals usually permit a sparse representation in a higher-dimensional space as compared to the observed signal. Hence it is advantageous to observe

only a part of the signal in this high dimensional space and reconstruct it using these measurements as opposed to observing the whole signal. Compressed sensing is thus a strategy that can exactly recover the sparse vectors associated with each observed signal based on its projections on randomly chosen vectors that may or may not form a basis [12].
Suppose $\mathbf{x} \in \mathbb{R}^N$ is a vector that is s-sparse. Hence the number of non-zero elements in $\mathbf{x}$ is at most s. Now suppose we can observe another vector $\mathbf{y} \in \mathbb{R}^d$ such that $s < d < N$. Then $\mathbf{x}$ can be reconstructed exactly from $\mathbf{y}$ by using $\mathbf{y} = \Phi\mathbf{x}$ with $\Phi \in \mathbb{R}^{d \times N}$ known as the over-complete dictionary. The approach to solve for $\mathbf{x}$ exactly is given by:

$$\hat{\mathbf{x}} = min\|\mathbf{x}\|_0 \text{ such that } \left\|\mathbf{y} - \Phi\mathbf{x}\right\|_2^2 < \epsilon_1 \tag{2.1}$$

Here ϵ_1 is the allowed error in the data fitting term on the right side of the equation. One major problem with this formulation is that it leads to a NP-hard problem [12]. Another approach is to solve:

$$\hat{\mathbf{x}} = min\|\mathbf{x}\|_1 \text{ such that } \left\|\mathbf{y} - \Phi\mathbf{x}\right\|_2^2 < \epsilon_2 \tag{2.2}$$

with ϵ_2 being the accepted tolerance in this formulation. The above convex problem can be cast into a linear program as follows (when ϵ_2 is unknown):

$$\hat{\mathbf{x}} = argmin_x \left\|\mathbf{y} - \Phi\mathbf{x}\right\|_2^2 + \lambda \left\|\mathbf{x}\right\|_1 \tag{2.3}$$

Here the regularization factor λ controls the trade-off between the data-fitting term and the sparsity on the vector $\hat{\mathbf{x}}$.

2.3.2. Building atlases and overcomplete dictionaries

An atlas $A = \{A_1, A_2\}$ is defined as a pair of images that can represent a set of images having tissue contrasts C_1 and C_2 respectively. It is assumed that each image within the atlas has the same resolution and both images have been appropriately co-registered. Further the authors assume that each atlas image is made up of $p \times q \times r$ 3D patches, and that each patch can be stacked as a $d \times 1$ vector where $d = p \times q \times r$. Then the over-complete dictionaries $\Phi_\mathbf{1}$ and $\Phi_\mathbf{2}$ are constructed by arranging each individual patch from A_1 and A_2 as the columns of source and target dictionaries respectively. In other words, each column of both fixed dictionaries is composed of the patches in their respective atlas images. Hence $\Phi_\mathbf{1}$, $\Phi_\mathbf{2} \in \mathbb{R}^{d \times N}$ where N is the number of patches in each of the component images.

2.3.3. Problem with using reverse imaging equation

Consider two MRI images I_1 and I_2 (source and target) of the same patient obtained using two different scanners. These images can be T_1-w, T_2-w or P_D-w 2D images or 3D volumes generated using:

$$I_1 = W_1\left(T_1, T_2, P_D, T_2^\star, \Theta_1\right) + \eta_1 \tag{2.4}$$

$$I_2 = W_2\left(T_1, T_2, P_D, T_2^\star, \Theta_2\right) + \eta_2 \tag{2.5}$$

where W_1 and W_2 are the imaging equations with respect to the pulse sequence used, Θ_1 and Θ_2 are the intrinsic parameters for an acquisition, η_1 and η_2 are the noise in each imaging model, T_1 and T_2 are the relaxation times associated with each acquisition and P_D is the proton density map of the tissue as explained in [12].

Ideally if the intrinsic parameters Θ_1 and Θ_2 are known for each image pair and a large number of images for the scanner corresponding to I_1 are available, then the relaxation times and proton density maps can be estimated [12]. Plugging these values in the imaging equation for I_2, we obtain the required transformed image $\hat{I}_2$. However the intrinsic parameters are often not known, and obtaining multiple acquisitions is a costly and time-consuming task. Hence reconstructing a source-target pair of images based on the MRI imaging equation is a challenging task.

2.3.4. Finding Sparse Representation vectors using Compressed Sensing

As mentioned above it is not possible to use the imaging equation of each scanner to transform a source image into the target intensity-space. Hence we assume that each patch in a source image I_1 (having contrast similar to C_1) can approximately be represented as follows:

$$\mathbf{y}_1(j) \approx \Phi_1\mathbf{x}_1(j) \;\; \forall\, j = 1, 2, 3...N \tag{2.6}$$

where $\mathbf{y}_1(j)$ are the patches of image I_1, $\mathbf{x}_1(j) \in \mathbb{R}^N$ s.t $\|\mathbf{x}_1(j)\|_0 \ll N$ are the estimated sparse representation vectors for each patch and Φ_1 is the pre-defined atlas source dictionary as described earlier. Using compressed sensing we have:

$$\mathbf{x}_1(j) = argmin_\alpha [\ \|\mathbf{y}_1(j) - \Phi_1\alpha\|_2^2 + \lambda\|\alpha\|_1]\ ,\ \alpha \geq 0\ \forall\, j = 1,2,3...N \tag{2.7}$$

where the positivity constraint on the representation vector ensures a valid anatomical meaning to each vector [11],[12]. In this way any source image having contrast similar to that of the source atlas image (C_1) can be decomposed into patches and further into the sparse representation vector for each patch using compressed sensing.

2.3.5. Generating synthetic contrast

As per the authors if the test image and the source atlas image are identical the estimated sparse representation vector for each patch $\mathbf{y}_1(j)\ \forall\, j = 1,2,...N$ will be a vector with only one entry indicating the index of the same patch within the source dictionary itself. However when the test image is different from the source atlas image, the estimated sparse representation vector for each patch of source image I_1 indicates the indices and linear coefficients associated with columns from dictionary $\Phi_\mathbf{1}$ used to represent each patch. Since the atlas images A_1 and A_2 are co-registered prior to building the source and target dictionaries, the columns of both dictionaries correspond to co-registered patches. Hence the transformed image patch having the desired contrast C_2 is given by:

$$\hat{y_2}(j) = \Phi_2\mathbf{x}_1(j) \tag{2.8}$$

Once each estimated patch has been obtained, the estimated image $\hat{I_2}$ can be reconstructed by appropriately recombining all transformed image patches.

Hence each sparse representation vector $\mathbf{x_1}(j)$ is used to locate the most representative columns from the source dictionary $\Phi_\mathbf{1}$ which can be used to represent a patch from the source image $\mathbf{y_1}(j)$. These columns correspond to similarly indexed patches from the source and target atlas images due to the one-to-one correspondence established by co-registration. Hence the estimated sparse vectors can be used to transform each source image patch into the intensity-space of the target image such that the synthetic image $\hat{I_2}$ has similar contrast to the target atlas image A_2. As per the authors such a strategy eliminates the need of accurate atlas-to-subject registration and thus offers improved robustness while performing the required intensity normalization.

2.4. Drawbacks of the MIMECS Method

While the MIMECS method presented above avoids explicit atlas-to-subject registration in order to improve the consistency of segmentation, it is nonetheless dependent upon two atlas images A_1 and A_2 which must be known for every set of test images. Further these images must be sufficiently good to represent an entire dataset of source and target images having tissue-contrast similar to C_1 and C_2 respectively. Hence the MIMECS method requires a pair of high-quality atlas images during each test phase.
Distinguishing between the atlas images as source and target can be motivated based on the specific segmentation used as explained in [11] and [12]. The authors chose a Magnetization Prepared Rapid Gradient Echo (MPRAGE) image as the target atlas image in each case arguing that segmentation is usually more consistent on such images. Hence such a strategy requires prior knowledge of the segmentation performance on the images within each dataset.
In addition, as reported in [12] the MIMECS method requires a prior pre-processing stage to reduce the computational burden associated with the estimation of the sparse signals $\mathbf{x}_1$ for every patch of the source image. This strategy divides both source and target dictionaries into various sub-dictionaries based on tissue-class as obtained using the TOADS algorithm [12]. Hence the performance of the MIMECS method is dependent on the approximate segmentation used prior to building the dictionaries and sub-dictionaries. With all this said, it is clear that there is room for improvement over the MIMECS method in order to make it applicable to more general Brain MRI studies. This is especially true for studies in which atlas images cannot be obtained readily or the choice of the atlas images cannot be motivated easily.

2.5. Main Contributions of Compressed Sensing Model

While the MIMECS method presented in [11] and [12] has some drawbacks, it was considered as an important milestone in developing an intensity-normalization method for MRI images. Most important of all it proves that MRI image patches can be represented sparsely using an appropriate over-complete dictionary. Second the experimental work reported in [11] and [12] enables us to understand whether any intensity normalization method can be used to generate synthetic tissue contrast based on its performance on the Brainweb Simulated Brain Database phantom images [5]. Third it establishes the use of the Kullback-Liebler (KL) Divergence between a pair of normalized image histograms as a metric to compare the difference in tissue contrast between two images. The latter two points are explained in detail with the various experiments performed in Chapter 4. For images in which the KL-Divergence reduces after intensity normalization, we can expect more consistent segmentation performance as reported in these works.

2.6. Summary

This chapter describes the previous methods that were considered as part of the literature study for this Master Thesis. Amongst these the two most important methods considered were Supervised Segmentation as reported in [15], [14] and the MIMECS method reported in [11], [12]. Both these methods can be used when atlas images associated with the Brain MRI images are available. However in their absence designing labelled and unlabelled atlas images requires a lot of human effort and is time-consuming. Hence these methods are limited in their applicability to general datasets of Brain MRI images, especially for those datasets where it is not possible to design separate atlas images.

The method of dictionary learning is described in the following chapter and it is the main focus of this project. It has been proposed to verify whether dictionary learning can be used for normalizing the intensities of Brain MRI images obtained from different scanners such that the images have identical tissue contrast after its application. One additional advantage thought to be associated with dictionary learning is that computational cost can be kept within acceptable limits during the training and testing phase by appropriately controlling the parameters of the optimization.

It should be noted that like the MIMECS method [11], [12], representation learning based on over-complete dictionaries is intended to be a pre-processing step applicable when a pair of source and target MRI images of the same patient are available. If the need for further segmentation arises, the normalized intensities obtained by using the dictionaries may be used with or without modifications for segmenting the MRI images. However these methods are out of the scope for this project, and hence attention is focused on learning compact dictionaries from the MRI images and verifying whether this strategy can be used for intensity normalization to produce synthetic images that have similar tissue contrast as that of the target image each time.

3

Dictionary Learning

3.1. The Method of Learning Dictionaries

Dictionary Learning is a technique which aims to find a sparse representation of the input training signals at its input by means of an over-complete dictionary. Keeping notational consistency with the method of compressed sensing as explained in Chapter 2, let us assume that the observed signals $\mathbf{y}_i \in \mathbb{R}^d \ i = 1,2,3,.....N$ can be collected as the columns of a matrix $Y \in \mathbb{R}^{d \times N}$. These observed signals can be represented using an over-complete dictionary $D \in \mathbb{R}^{d \times K}$ $(d < K)$. Where each of the K columns is known as a dictionary prototype or atom in literature [2]. Each training signal is represented as a sparse linear combination of the columns of the same over-complete dictionary. In order to represent a set of signals in this manner we need a matrix of sparse representation vectors $X \in \mathbb{R}^{K \times N}$ and hence the equivalent formulation of the original dictionary learning problem is as follows:

$$min_{\mathbf{x}(i)} \|\mathbf{x}(i)\|_0 \text{ subject to } Y = DX \ \ \forall i = 1,2,...N \tag{3.1}$$

Here $\mathbf{x}(i)$ denotes one column from the matrix X. The goal thus is to find the sparsest solution possible such that each of the observed signals can be represented using the columns of the over-complete dictionary. One important point however is that the over-complete dictionary in each case has to be inferred from the input data itself. As reported in [8] such a technique often leads to an improvement over representations obtained from fixed dictionaries as in the former case, the dictionary matrix and in turn the representation associated with each training signal gets tuned to the available data itself by means of an appropriate cost function.

3.2. Representation Learning Using Over-Complete Dictionaries

As mentioned above each learnt dictionary is assumed to be over-complete since the number of columns in the dictionary is larger than the dimension of the original data $(d < K)$. This means that the system of equations given by:

$$Y = DX, d < K \ll N \tag{3.2}$$

is an under-determined system with greater columns than rows and hence there is no unique solution to the dictionary learning problem. In addition both D and X are unknown quantities that must be estimated using the available training data. Several algorithms have been devised to handle such an optimization function with the guarantee of convergence to a local minimum of the empirical cost function [2] and that to an almost sure convergence to the minimum of the expected cost function [8].

Owing to the higher number of columns as compared to the signal dimension, the representation associated with each signal is assumed to lie in a higher-dimensional space as compared to the original data. This is contrast to methods such as dimensionality reduction and principal component analysis (PCA) which assume that the training data can be represented using a lower-dimensional space as compared to the original data. Due to this property of the over-complete dictionary, its columns are no longer required to be orthogonal. This enables more flexible representations of the input data despite the fact that each training signal is represented using a linear combination of far fewer number of elements (dictionary atoms) [2].

3.3. Proposed Advantages of Dictionary Learning

Similar to the MIMECS method in [11] and [12] dictionary learning models each patch from a MRI image as the sparse combination of a few linear elements. However learning a pair of source-target dictionaries was considered to have a few advantages over using pre-defined dictionaries such as those used in the former approaches.

The MIMECS method decomposes a pair of atlas images as the columns of the pre-defined source and target dictionaries as described in Chapter 2. As mentioned previously obtaining such atlas images for every dataset is not an easy task. In their absence we need to choose appropriate images from the available set which can be treated as atlas images. In this case, the performance achieved with intensity normalization and the subsequent segmentation is then dependent on the choice of representative images. Since dictionary learning can learn a pair of source and target dictionaries without any external input, we can eliminate the dependency on atlas images. This makes the method more readily applicable for datasets containing Brain MRI images of the same patient.
In addition to this, each image patch centred about a voxel contains local information such as the variation of intensities within a neighbourhood of fixed size. Hence each image patch can be treated as a feature vector of the centre voxel of the patch [11]. As dictionary learning treats each image patch as a training signal, this local information can be very easily learnt using the associated dictionaries.
Another advantage of considering dictionary learning in this project is that the size of each learnt dictionary (K) can be controlled very easily during the training phase. By learning compact dictionaries every time, the computational cost associated with obtaining the sparse representation vectors for each image patch during the testing phase is reduced considerably. Doing so avoids the need of an approximate segmentation such as that followed in [12]. With these advantages stated, we can proceed to the dictionary learning algorithms studied as part of this project.

3.4. Dictionary Learning using K-SVD

In [2] the authors present a method of jointly learning an over-complete dictionary and representation vector for each patch in the training dataset. As per the authors this strategy was the first to offer improved speed of learning the dictionary over several other classical dictionary learning methods such as the Approximated Maximum-Likelihood method, the Union of Orthonormal Dictionaries method as well as the Method of Optimal Directions (MOD). As per the authors the K-SVD method is a generalization of the k-means clustering algorithm where each training signal is represented by the linear combination of a few dictionary elements acting as centroids like in the k-means algorithm. The cost function used in the K-SVD algorithm can be written as follows:

$$min_{D,X} \| Y - DX \|_F^2 \text{ subject to } \forall i \ \| \mathbf{x}_i \|_0 \leq T_0 \tag{3.3}$$

where the patch vectors and their sparse representation vectors are denoted compactly using the matrices Y and X respectively , T_0 is the chosen non-negative number indicating the maximum support of each representation vector and $\| \ \|_F$, $\| \ \|_0$ denote the Frobenius norm and L_0 pseudo-norm respectively. It is hence seen that the number of non-zero coefficients and in turn the number of linear combinations of the dictionary columns is kept fixed to T_0 in the K-SVD method. The algorithm was implemented using the freely available MATLAB Toolbox on [13] and its main steps are as follows:

1. **Initialization:** Set Dictionary Matrix $D \in \mathbb{R}^{d \times K}$ with l_2 normalized columns.
2. Repeat until convergence (or stoppping criterion)
   (a) **Sparse Coding Stage:** Using an appropriate pursuit method [2] compute the sparse representation vectors $\mathbf{x}(j)$ for the patches in the source dataset $\mathbf{y}(j) \ \forall \, j = 1, 2, 3, ...N$, by solving:

$$min_{\mathbf{x}(j)} \left\| \mathbf{y}(j) - D\mathbf{x}(j) \right\|_2^2 \text{ subject to } \left\| \mathbf{x}(j) \right\|_0 \leq T_0 \tag{3.4}$$

   (b) **Codebook Update Stage:** Denote the dictionary matrix from the previous iteration as D_1^{t-1}, then we observe that the cost function can be written as:

$$\left\| \left(Y - \sum_{j \neq k} \mathbf{d}_j \mathbf{x}_T^j \right) - \mathbf{d}_k \mathbf{x}_T^k \right\|_F^2$$

Hence we can also write:

$$\left\| E_k - \mathbf{d}_k \mathbf{x}_T^k \right\|_F^2 \tag{3.5}$$

Where E_k denotes the error obtained when the k^{th} dictionary column is inactive.

3. Observe that the cost function has been represented as the inner-product of a particular dictionary column and the sparse-coefficients using that particular column for representing the data points. Minimizing such a cost function for all k columns is hence equivalent to solving K Rank-1 Singular Value Decompositions, from which the algorithm derives its name [9].

4. Solving the cost function expressed in Equation 3.5 by following the strategy in [2] we obtain the updated dictionary to be used in the next iteration of the learning algorithm.

3.5. Proposed Algorithm: K-SVD

With the K-SVD method explained above, we can proceed to the actual contrast synthesis algorithm proposed using K-SVD as follows:

1. **Training Phase**

   (a) Given the images I_1 and I_2 considered as the source and target respectively, obtain a one-to-one correspondence between the voxels in each image using an appropriate registration method if required.

   (b) Decompose each image into N overlapping patches containing $n \times n \times n$ voxels each, such that $d = n^3$ is the dimension of each signal patch.

   $$y_1(i), y_2(i) \in \mathbb{R}^d \quad \forall\, i = 1, 2, ...N \tag{3.6}$$

   Following this collect all training patches obtained as the columns on matrices Y_1 and Y_2 respectively.

   (c) Use the K-SVD algorithm to get dictionaries $D_1, D_2 \in \mathbb{R}^{d \times K}$ and sparse matrices $X_1, X_2 \in \mathbb{R}^{K \times N}$ for each image respectively, by optimizing the cost function in Equation 3.3 for each image.

2. **Testing Phase**

   (a) Assuming that the sparse representation vectors of the source image patches can be used as indices to the columns of the over-complete target dictionary, the transformed image patches can be compactly denoted as:

   $$\hat{Y_2} = D_2 X_1 \tag{3.7}$$

   The final output image $\hat{I_2}$ is then obtained by an appropriate recombination of the columns of $\hat{Y_2}$, where each patch of $\hat{I_2}$ is the corresponding column of $\hat{Y_2}$.

   (b) Repeat this procedure for all pairs of source and target images considered.

3.6. Problems with K-SVD

Initially the K-SVD method was considered as a good starting point for learning a pair of fixed size dictionaries from source and target images. However analysing the method a little deeper, revealed a few drawbacks associated with the optimization strategy followed in the K-SVD method. These were as follows:

1. **Fixed Sparsity of Representation Vectors:** The number of support coefficients (non-zero coefficients) in each representation vector is assumed to be same and known prior to the learning of any over-complete dictionary.

2. **Increased Learning Time:** As the K-SVD method considers the entire batch of training examples for updating each column of the dictionary in every iteration it typically takes a longer time to achieve convergence.

3. **The Use of a Pseudo-Norm:** Owing to the L_0 pseudo-norm in the objective function, it can no longer be considered as a convex optimization problem. Moreover as explained by the authors in [2] in each iteration, during the update of the k^{th} dictionary column $\mathbf{d}_k$ further reduction or no reduction of the mean-square error (MSE) is guaranteed in order to maintain the sparsity constraint. Hence the algorithm can only be guaranteed to converge to a local-minimum of the objectve function in each iteration.

3.7. Online Dictionary Learning

Owing to the drawbacks of the K-SVD algorithm as discussed above, there was a need for a better dictionary learning method which could be implemented in MATLAB [10] for learning a pair of dictionaries. Fortunately, the SPAMS Software available on [7] provides functions for learning dictionaries from a given dataset and obtaining the sparse representation coefficients through a decomposition based on the dictionary columns by using the method developed in [8].

The main dictionary learning function mexTrainDL solves the following objective function:

$$min_{D\in C} \lim_{N\to+\infty} \frac{1}{N}\sum_{i=1}^{N} min_{\mathbf{x}_i\in\mathbb{R}^K}\left(\frac{1}{2}\left\|\mathbf{y}_i - D\mathbf{x}_i\right\|_2^2 + \lambda\left\|\mathbf{x}_i\right\|_1\right) \tag{3.8}$$

Where λ is the regularization parameter and $\|\ \|_1$ denotes the L_1 norm of the sparse representation vector. The convex set of matrices that solves this objective function is given by:

$$C = \left\{D \in \mathbb{R}^{d\times K} \text{ s.t } \forall\, j = 1,2,..K, \mathbf{d}_j^T\mathbf{d}_j \leq 1\right\} \tag{3.9}$$

The constraint on the inner products of the dictionary columns is to ensure that the coefficients in each column do not get tuned to high values. This would in turn lead to the coefficients of the sparse vectors to get tuned to arbitrarily small values [8]. As explained in [8] the cost function given by Equation 3.8 is the expected value of the empirical cost function given below:

$$min_{D\in C,\mathbf{x}\in\mathbb{R}^K} \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{2}\left\|\mathbf{y}_i - D\mathbf{x}_i\right\|_2^2 + \lambda\left\|\mathbf{x}_i\right\|_1\right) \tag{3.10}$$

The authors further explain that one is usually interested in obtaining a solution to the expected cost as opposed to the empirical cost in settings such as dictionary learning. In addition stochastic techniques can be considered in order to achieve faster and almost sure convergence to the minimum of the expected cost function under certain conditions as considered in [8]. Thus the dictionaries obtained through Online Dictionary Learning are more mathematically tractable as compared to dictionaries of the same size obtained using K-SVD and do not require the high training time associated with the former method as well.

3.8. Proposed Algorithm: Online Dictionary Learning

With the background developed for the method of Online Dictionary Learning, we can proceed to the proposed algorithm for Contrast Normalization using Online Dictionary Learning.

1. **Training Phase**

   (a) Assume we have images I_1 and I_2 termed as the source and target image respectively. Further assume that a one-to-one correspondence has been established between each pair of voxels in these images.

   (b) Each image can be decomposed as overlapping patches containing $n \times n \times n$ voxels each, such that $d = n^3$ is the dimension of each signal patch.

   $$y_1(i), y_2(i) \in \mathbb{R}^d \quad \forall\, i = 1,2,....N \tag{3.11}$$

   (c) Let N be the total number of patches in each image, then we proceed to learn the source-target pair of dictionaries $D_1, D_2 \in \mathbb{R}^{d\times K}$ such that:

   $$min_{D_1\in C} \lim_{N\to+\infty} \frac{1}{N}\sum_{i=1}^{N} min_{\boldsymbol{\alpha}_1(i)}\left(\frac{1}{2}\left\|y_1(i) - D_1\boldsymbol{\alpha}_1(i)\right\|_2^2 + \lambda\left\|\boldsymbol{\alpha}_1(i)\right\|_1\right) \tag{3.12}$$

   $$min_{D_2\in C} \lim_{N\to+\infty} \frac{1}{N}\sum_{i=1}^{N} min_{\boldsymbol{\alpha}_2(i)}\left(\frac{1}{2}\left\|y_2(i) - D_2\boldsymbol{\alpha}_2(i)\right\|_2^2 + \lambda\left\|\boldsymbol{\alpha}_2(i)\right\|_1\right) \tag{3.13}$$

   (d) Here $\boldsymbol{\alpha}_1(i)$ and $\boldsymbol{\alpha}_2(i)$ $\forall i = 1,2,3,...N$ denote the sparse representation vectors associated with each corresponding pair of source and target image patches during training respectively.

2. **Testing Phase**

   (a) Obtain the sparse representation vectors for the source image patches using the LARS algorithm [8] by solving the following cost function:

$$min_{\boldsymbol{\alpha}_1 \in \mathbb{R}^K} \frac{1}{2} \left\| \mathbf{y}_1 - D_1 \boldsymbol{\alpha}_1 \right\|_2^2 + \lambda \left\| \boldsymbol{\alpha}_1 \right\|_1 \forall i = 1,2,3...N \tag{3.14}$$

   (b) We assume that the coefficients of each sparse vector $\boldsymbol{\alpha}_1(i)$ can be used as indexes into the target dictionary for each image patch $y_1(i)$ $i = 1,2,...N$ from the source image. Hence we can write the transformed patch from a source image as:

$$\hat{y_2}(i) = D_2 \boldsymbol{\alpha}_1(i) \quad \forall i = 1,2,3...N \tag{3.15}$$

   (c) With each transformed image patch obtained thus, the reconstructed target image $\hat{I_2}$ can be built by recombining the image patches obtained as mentioned earlier with the K-SVD method.

3.9. Common Problem with Earlier Algorithms

The earlier contrast synthesis algorithms using K-SVD and Online Dictionary Learning allow us to represent both source and target images using a source-target pair of data adaptive over-complete dictionaries. Using an appropriate registration method one-to-one correspondence can be established between the voxels of the source image and those of the target image. This correspondence is maintained while decomposing each image into patches and hence during the learning of each dictionary pair as well. Hence the relative position of each pair of source-target image patches is fixed with respect to one another.
However no relationship is established between the pair of dictionaries themselves. This will lead to each source image patch having entirely different sparse representation vectors based on the source and target dictionary pair. This would in turn mean that the sparse representation vectors obtained from a decomposition on the source dictionary, do not allow transformation into the target intensity-space by indexing the appropriate columns of the target dictionary. Hence one strategy is to ensure that each pair of patches from the source and target images share a common sparse representation based on two different dictionaries $\hat{D_1}$ (source) and $\hat{D_2}$ (target) respectively. The following method describes an algorithm, which can possibly lead to such a solution.

3.10. Contrast Synthesis Algorithm: Joint Dictionary Learning

The work of Yang et al., [16] focuses on designing a pair of dictionaries for image super-resolution between a pair of low-resolution and high-resolution images. Interestingly this pair of dictionaries is learnt jointly for a single image and a synthetically generated low resolution counterpart of the same image. In addition, during the training phase each dictionary is constrained to have a common sparse representation for a corresponding pair of image patches used as input.

It was considered that such a approach could be used for intensity-normalization between a pair of images by replacing the low resolution image with the source and the high resolution image with target respectively. One added advantage of considering such an approach is that it allows both dictionaries to be learnt jointly by enforcing simple constraints on the cost function used in joint dictionary learning. Hence using the SPAMS Toolbox [7] Joint Dictionary Learning could be implemented very easily in the MATLAB [10] environment. The main steps of the considered contrast synthesis algorithm can be explained as follows:

1. **Training Phase**

   (a) Assume we have source image I_1 and target image I_2 where there exists a one-to-one correspondence between each pair of voxels from both images.

   (b) Each image can be decomposed as overlapping patches containing $n \times n \times n$ voxels each, such that $d = n^3$ is the dimension of each signal patch.

$$y_1(i), y_2(i) \in \mathbb{R}^d \quad \forall i = 1,2,....N \tag{3.16}$$

(c) Let N be the total number of patches in each image, then the cost function associated with Joint Dictionary Learning can be written as:

$$min_{\hat{D}_1 \in C} \lim_{N \to +\infty} \frac{1}{N} \sum_{i=1}^{N} min_{\boldsymbol{\alpha}(i)} \left(\frac{1}{2} \left\| y_1(i) - \hat{D}_1 \boldsymbol{\alpha}(i) \right\|_2^2 + \lambda \left\| \boldsymbol{\alpha}(i) \right\|_1 \right) \tag{3.17}$$

$$min_{\hat{D}_2 \in C} \lim_{N \to +\infty} \frac{1}{N} \sum_{i=1}^{N} min_{\boldsymbol{\alpha}(i)} \left(\frac{1}{2} \left\| y_2(i) - \hat{D}_2 \boldsymbol{\alpha}(i) \right\|_2^2 + \lambda \left\| \boldsymbol{\alpha}(i) \right\|_1 \right) \tag{3.18}$$

$$s.t. \left\| \hat{D}_1(:,k) \right\|_2 \leq 1, \left\| \hat{D}_2(:,k) \right\|_2 \leq 1 \quad \forall k = 1,2...K$$

(d) This is equivalent to solving:

$$min_{\hat{D}_1, \hat{D}_2 \in C} \lim_{N \to +\infty} \frac{1}{N} \sum_{i=1}^{N} min_{\boldsymbol{\alpha}(i)} \left(\frac{1}{2} \left\| y_1(i) - \hat{D}_1 \boldsymbol{\alpha}(i) \right\|_2^2 + \frac{1}{2} \left\| y_2(i) - \hat{D}_2 \boldsymbol{\alpha}(i) \right\|_2^2 + \lambda \left\| \boldsymbol{\alpha}(i) \right\|_1 \right) \tag{3.19}$$

(e) As suggested in [16] the reconstruction error terms can be grouped together, and hence denoting

$$\bar{y}(i) = \begin{bmatrix} y_2(i) \\ \\ y_1(i) \end{bmatrix}, \bar{D} = \begin{bmatrix} \hat{D}_2 \\ \\ \hat{D}_1 \end{bmatrix} \tag{3.20}$$

we can solve the joint optimization problem as:

$$min_{\bar{D} \in C} \lim_{N \to +\infty} \frac{1}{N} \sum_{i=1}^{N} min_{\boldsymbol{\alpha}(i)} \left(\frac{1}{2} \left\| \bar{y}(i) - \bar{D} \boldsymbol{\alpha}(i) \right\|_2^2 + \lambda \left\| \boldsymbol{\alpha}(i) \right\|_1 \right) \tag{3.21}$$

$$s.t. \left\| \bar{D}(:,k) \right\|_2 \leq 1 \quad \forall k = 1,2...K$$

to obtain the source and target dictionaries $\hat{D}_1$ and $\hat{D}_2$ that allow a common sparse representation based on the columns of each over-complete dictionary.

2. **Testing Phase**

(a) Since the pair of source and target dictionaries allow a common sparse decomposition based on their respective columns, the sparse representation for each source image patch can be obtained by solving the following equation:

$$min_{\boldsymbol{\alpha}(i) \in \mathbb{R}^K} \frac{1}{2} \left\| y_1(i) - \hat{D}_1 \boldsymbol{\alpha}(i) \right\|_2^2 + \lambda \left\| \boldsymbol{\alpha}(i) \right\|_1 \tag{3.22}$$

(b) Then the transformed image patches can be written as:

$$\hat{y}_2(i) = \hat{D}_2 \boldsymbol{\alpha}(i) \quad \forall i = 1,2,3...N \tag{3.23}$$

and the final transformed image $\hat{I}_2$ is obtained by recombining the image patches obtained appropriately.

3.11. Summary

This chapter explains the main algorithms of dictionary learning considered for this project. These methods were considered to be advantageous over the MIMECS method presented in [11] and [12] since they are dependent on only the source and target image considered for intensity normalization. This in turn makes the method more readily applicable when a pair of MRI images of the same patient is available. An added advantage of learning dictionaries over using pre-defined dictionaries is that local information present in a joint pair of image patches is included in the training phase and hence incorporated in the dictionaries that are learnt from the data.

The first method considered was the K-SVD algorithm presented in [2], however due to the reasons of mathematical intractability and increased computational burden associated with training dictionaries using K-SVD, this method was not considered any further during the experimental phase. Following the failure of

the K-SVD, Online Dictionary Learning [8] was considered for learning the pair of source-target dictionaries. However a closer look revealed that this method also cannot be directly used for a task such as intensity normalization such as the one considered in this project. This is mainly because these methods do not enforce a relationship between the pair of source and target dictionaries. Each dictionary learnt using Online Dictionary Learning will hence be adapted to the individual image that is used during the training phase and cannot be used to transform the patches from a source image into the intensity space of the target image.

This brings us to the method of super-resolution between image patches developed by Yang et al., [16]. The most important feature of this approach is that the source-target dictionary pair learnt during the training phase allow for a common sparse representation with respect to one another. This relationship was used to develop the third and final contrast synthesis algorithm considered in this project. During training both images are used to jointly learn a pair of dictionaries which encourage a common sparse representation vector for a given pair of source and target image patches. However during testing, only the patches from the source image and the source dictionary are used to recover the associated sparse representation vectors. Using the linear combination of the columns from the target dictionary, as indicated by the coefficients of each sparse representation vector would then possibly allow each source image patch to be transformed into the intensity-space of the target image. The final transformed image would then have intensities and tissue contrast similar to that in the target image.

The next chapter describes the various experiments conducted to verify if the proposed method can actually be used for normalizing the intensities within a source image with respect to the associated target image. It also describes the choice of the different synthetic image datasets used to evaluate the performance of the method as well as the different metrics used. Particularly important is the use of KL-Divergence between a pair of normalized image histograms to quantify the difference in tissue contrast between a pair of source and target images as reported previously in [11] and [12].

4

Experiments With Synthetic Data

4.1. Proposed Outcome of Experiments

The goal of the contrast synthesis algorithm developed in Section 3.10 is to transform the source image from a source-target pair into the intensity-space of the target image by first learning a pair of over-complete dictionaries for both images. These dictionaries are constrained to have the same sparse representation with respect to each other for a given pair of input signals in the training phase. During the testing phase this allows us to obtain the sparse representation for each patch from the source image based on the learnt source dictionary. The coefficients of the sparse vector obtained for each source image patch indicate the linear combination of the columns from the target dictionary to be used in representing the source patch in the target intensity-space. This can thus be considered as a transformation of each source image patch based on the over-complete dictionaries and its underlying sparse representation. Such a strategy falls under the broad heading of intensity normalization methods similar to the MIMECS method developed in [11] and [12].

It was reported in these works that normalizing the intensities into a common target intensity-space can be used to generate synthetic tissue contrast in test images such that the output image has tissue contrast similar to that present in the target atlas image considered. Similarly, it is expected that the image at the output of the contrast synthesis method based on Joint Dictionary Learning will have tissue intensities and a intensity distribution similar to those present in the target image. In turn, this is expected to reduce the difference in tissue contrast between each corresponding pair of images. The experiments reported in this chapter evaluate the performance of the algorithm presented in Section 3.10 with regard to intensity normalization, while the differences in tissue contrast are measured using an appropriate metric that is described later.

4.2. Simulated Brain Database

Before considering any dataset of real-life images it is important to evaluate the performance of any intensity normalization method on synthetic images whose intensities can be controlled very accurately. Such images are known in literature as ground truth images [11], [12]. For studies related to Brain MRI images, the Brainweb: Simulated Brain Database (SBD) provides ground truth images in the form of phantom images generated using one of the pulse sequences from T_1, T_2 or P_D. Other parameters such as slice thickness (t), noise percentage (n) and intensity non-uniformity (rf) can be controlled using the software tools available on the Brainweb website. Using these two distinct synthetic datasets were created for observing the performance of the contrast synthesis method chosen. These are described below:

4.2.1. Dataset 1

Consider a scenario in which a single patient is imaged at different times with the same scanner using the same pulse sequence and internal parameters. Further assume that the external parameters such as software settings are also identical. In this case the differences in tissue contrast between the images will be due to the differences in noise present in each image. This scenario can be simulated using Brainweb images of the same modality (pulse sequence) but having different noise percentages (n). For simplicity the intensity non-uniformity (rf) is kept constant and fixed to zero for all images.

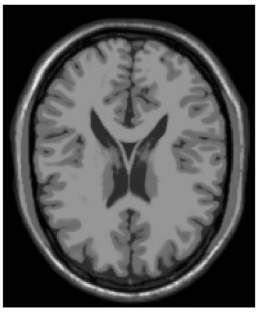

Figure 4.1: Target Image for Dataset 1

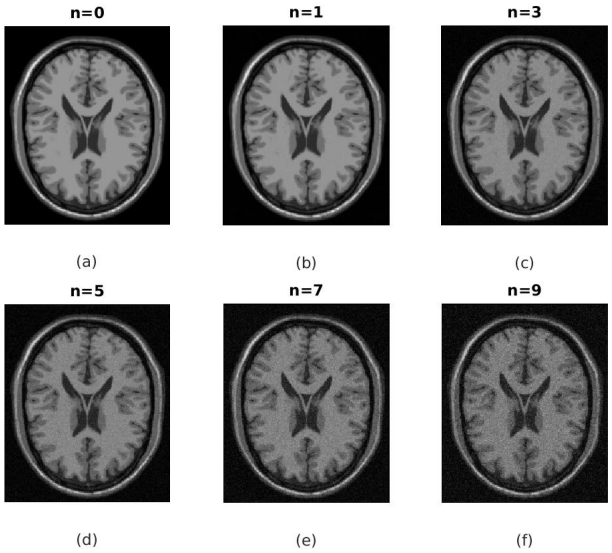

Figure 4.2: Source Images for Dataset 1

This was termed as Dataset 1 in this project, with source images chosen as the T_1-w phantoms with increasing noise percentages $n \in \{0,1,3,5,7,9\}$. The target image for this dataset was chosen as the T_1-w phantom with $n = 0$ which is the same as the first source image. Figures 4.1 and 4.2 display the target and source images of Dataset 1 respectively.

4.2.2. Dataset 2

MRI images of the same patient can generally be obtained by collecting data from different scanners. These images are usually obtained by using different pulse sequences and different internal and external parameters. In this case the dataset considered above will not be sufficient to evaluate any contrast synthesis method since the differences in tissue contrast now arise because of the differences in modalities as well as due to the differences in noise between images. Hence a different dataset was simulated using the Brainweb phantoms to evaluate the performance of the Joint Dictionary Learning method.
Dataset 2 uses the same source images as before, these are the T_1-w phantoms with $n \in \{0,1,3,5,7,9\}$. However the target image is changed to the T_2-w phantom with $n = 0$. These images are displayed in Figures 4.3 and 4.4 respectively.

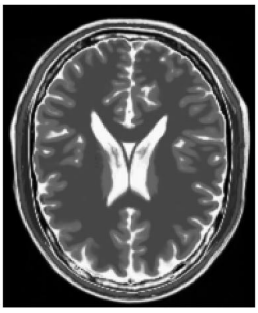

Figure 4.3: Target Image for Dataset 2

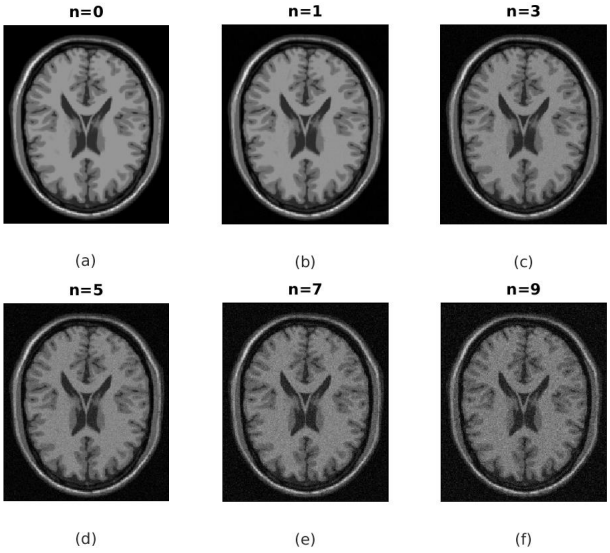

Figure 4.4: Source Images for Dataset 2

4.2.3. Normalizing the Peak Intensities of MRI Images

In synthetic images such as those from the Simulated Brain Database [5], image intensities are dependent on the pulse sequence used to generate each phantom image. Hence as a result the images especially those obtained using the pulse sequences T_1 and T_2 were observed to have widely different intensities. Even amongst themselves the T_1-w phantoms from Dataset 1 were observed to have different intensities at identical locations due to the presence of noise. Hence it is important to pre-normalize the intensities within each Brain MRI image such that the peak intensity associated with white matter corresponds to unity while the lowest intensity corresponding to zero is associated with the background. For a given MRI image Y, this was achieved by using the following two-step method as described in [6].

$$Y_m(x) = Y(x) - min_x Y(x) \tag{4.1}$$

$$Y_n(x) = \frac{Y_m(x)}{max_x Y_m(x)} \tag{4.2}$$

Here x denotes a voxel in an MRI image Y and min_x and max_x denote the minimum and maximum voxel intensity in an image respectively. This method ensures that the range of intensities present in the output MRI image Y_n varies from zero (background voxels) to unity (white matter). This procedure was repeated for

all source and target images considered from both datasets. It should be noted that the range of intensities has been normalized in each image, however the difference in tissue contrast between each pair of source and target image still exists and needs to be normalized in a later stage.

4.2.4. Observing the Tissue Contrast

Like in 2D images, the contrast of 3D MRI images is a qualitative measure which needs to quantized. However the amount of brain tissue acquired during imaging varies from slice to slice in 3D images. Very often, few slices in each image contain little or no tissue at all. Hence the central slice in each MRI image can be treated as a single 2D image containing most of the observable brain mass. Assuming that the central slice in an MRI image is the most representative slice, its contrast can be observed using its histogram expressed as follows [6]:

$$h(k) = n_k \quad \forall\, k = 1, 2, ...L \tag{4.3}$$

Here k is a discrete intensity value, n_k is the number of pixels having intensity k and L is the maximum value of the set of discrete intensities present in the central slice image.
The normalized histogram of the image can then be expressed as:

$$h(k) = \frac{1}{N} n_k \quad \forall\, k = 1, 2, ...L \tag{4.4}$$

where N is the total number of pixels present in the image. As explained in Section 3.3 in [6], the normalized histogram of an image can be treated as an estimate of the underlying probability distribution function of the intensities in an image. Hence the tissue contrast of an MRI image can be observed through the histogram of the central slice in each image [11], [12]. A histogram that tends to concentrate at either low intensities or high intensities is indicative of an overall dull and saturated image respectively having poor tissue contrasts.

4.2.5. Comparing Tissue Contrasts

As explained above, the normalized histogram of the central slice in each MRI image can be treated as an estimate of the probability distribution function of the intensities in an MRI image. Once this estimate is available to us, a quantitative measure for comparing the contrasts in two images can be developed. A natural quantitative metric is to compare the estimated probability distribution functions of two central slice images. The use of a such a metric is reported in [11] and [12], in which the authors observe the asymmetric Kullback-Liebler (KL) Divergence between the pair of normalized image histograms from each source and target image to compare the differences in tissue contrast between each pair of images. The same metric was adopted in this project as well, and can be expressed as follows:

$$d_{KL} = \sum_{k=1}^{L} h_1(k) \ln\left(\frac{h_1(k)}{h_2(k)}\right) \tag{4.5}$$

Where h_1 and h_2 are the normalized histograms of the central slice within the target image and source image respectively. Since we are using discrete intensities of images to compute each histogram, the KL-Divergence between the probability distribution functions was computed using the heuristic method suggested in [4].
Tables 1 and 2 display the original difference in tissue contrast between the target image and each source image from Dataset 1 and 2 respectively. We can observe that the presence of noise also contributes significantly to the difference in tissue contrast between a pair of MRI images. Also observe that the first column of Table 1 displays a zero since the source and target image considered in this case are identical to each other.

Noise in Source Image	n=0	n=1	n=3	n=5	n=7	n=9
Calculated KL Divergence	0	0.6512	1.3306	1.3730	1.3064	1.4386

Table 4.1: Original Difference in Tissue Contrast Between Target and Source Images: Dataset 1

Noise in Source Image	n=0	n=1	n=3	n=5	n=7	n=9
Calculated KL Divergence	1.2008	1.1569	1.2737	1.3501	1.3637	1.3321

Table 4.2: Original Difference in Tissue Contrast Between Target and Source Images: Dataset 2

4.3. The Need for Sub-Sampling Images

The size of each phantom image considered in this project was $217 \times 181 \times 181$. For all experiments described in this chapter the image patches were chosen as overlapping cubical blocks. This resulted in a dataset of a few million signals per image. Despite the improved speed of learning dictionaries, processing a dataset this large places a huge computational burden and requires tremendous system memory. This can be handled by reducing the size of the dataset and hence each slice per image was sub-sampled twice by removing alternate pixels. The advantages of such a strategy are:

1. The method of sub-sampling operates on each slice within a 3D image. Hence a major portion of the imaged content is still retained in the volume of each MRI image.
2. Since the sequence of discarding pixels from each slice is the same for source and target images, the voxel-to-voxel correspondence between images is still preserved.
3. Sub-sampling produces a smaller dataset that is handled very easily by the dictionary learning algorithm.

Following sub-sampling twice, each image can again be divided as overlapping patches of $p \times q \times r$ voxels each. However now the size of the dataset per image is reduced to a few hundred thousand training signals which can be very easily handled using the functions implemented in the SPAMS Toolbox [7].

One additional pre-processing stage was added before the images could be divided into overlapping patches which is known in literature as zero padding [6]. Following a similar strategy as in the 2D case, each image was padded with $\frac{w-1}{2}$ rows, columns and slices of zeros on all six sides of the 3D image considered as a cube. Here w is the size of the neighbourhood considered around each centre voxel. For all experiments in this project cubical blocks corresponding to odd sized neighbourhoods were used and hence $w = p = q = r$ for all experiments reported in this project.

4.4. Experimental Evaluation

This section describes the various experiments performed with the contrast synthesis method of Section 3.10 on the images from the two datasets. The line of reasoning followed in these experiments is consistent with the Validation Experiments on the T_1-w and T_2-w phantoms reported in Chapter 4 of [11]. These can be used to understand the performance of any contrast synthesis method when the source image is varied from the available set of T_1-w phantoms for different values of noise. Through the following subsections, each part of the experimental evaluation is described in detail.

4.4.1. Evaluation Metric

Consider a pair of sub-sampled images I_{1s} and I_{2s} known as the source and target image respectively. Using the algorithm presented at the end of Chapter 3 we can transform each patch from the source image into the intensity space of the target image. This results in the transformation from the intensity space of the source image to that of the target image given by:

$$y_{1s}(i) \rightarrow \hat{y_{2s}}(i) \quad \forall i = 1,2,3....N_s \tag{4.6}$$

where y_{1s} is a patch from the source image, $\hat{y_{2s}}(i)$ denotes each output image patch and N_s is the total number of patches in the sub-sampled image obtained as described earlier. Recombining all the transformed patches obtained gives us the transformed source image $\hat{I_{2s}}$ which is a result of the transformation:

$$I_{1s} \rightarrow \hat{I_{2s}} \tag{4.7}$$

Hence a straightforward way to compare each output image with the target is to measure the Mean Square Error (MSE) between the original target image and the transformed source image obtained. This is given by:

$$MSE = \frac{1}{N_v} \sum_{i=1}^{N_v} \left(I_{2s}(i) - \hat{I_{2s}}(i) \right)^2 \tag{4.8}$$

Here N_v denotes the number of voxels in each sub-sampled target image and the transformed image as well. This metric is consistent with the experiments followed in [11] and [12] to evaluate the MIMECS method and was hence adopted without much modification.

4.4.2. Validating Performance: Dataset 1

As mentioned earlier, Dataset 1 allows us to study the performance of an intensity normalization method when source and target images belong to the same modality but differences in tissue contrast arise due to the noise added to the source image itself.

In each trial of the experiment, the source image I_1 was varied from the set of T_1-w phantoms with $n \in \{0,1,3,5,7,9\}$ while the target image I_2 was fixed as the T_1-w phantom with $n = 0$. Following sub-sampling, zero padding and the subsequent decomposition into d-dimensional patches a pair of source-target dictionaries were learnt for each source-target pair of images. Using the target dictionary each source image was transformed by using the testing phase of the algorithm described in Section 3.10 and the Mean Square Error between the target image and the transformed source image was computed. This procedure was repeated for a range of values of λ chosen as $\lambda \in \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$.

It is clear that the number of non-zero coefficients in each representation vector decrease as λ gets set to higher values as observed through the cost function in Equation 3.19. Further, in order to get a reliable estimate of the Mean Square Error for each pair of source and target images considered each trail of the experiment was repeated ten times (for all values of λ considered) and the average Mean Square Error over ten trials was computed.

1. **Results with $3 \times 3 \times 3$ patches** : For this size of cubical neighbourhood considered around each voxel of an MRI image, the resulting dimension of each training signal is $d = 27$. Hence the number of dictionary columns was fixed to $K = 100$. It was observed that the performance of the Joint Dictionary Learning method depends more on the choice of λ as opposed to the choice of K. Hence K was selected to satisfy the requirement of over-completeness of the dictionary matrix ($d < K$) while keeping the computational cost associated with training dictionaries and obtaining sparse representation vectors during the testing phase within reasonable limits. Through the experiments it was observed that for this particular size of the cubical neighbourhood intensity normalization using Joint Dictionary Learning had three regions of operation based on the Average Mean Square Error between the target image and each output image. These were as follows:

   - Low Sparsity Region of Operation ($\lambda = 0.1, 0.2, 0.3$): Characterized by low Mean Square Error between the target image and output images corresponding to source images having low noise levels. The performance of the Joint Dictionary Learning degrades very quickly as the noise in the source image increases. This effect was most pronounced at $\lambda = 0.1$ while at $\lambda = 0.3$ it was observed that a single output image results in a very high Mean Square Error. This effect is displayed in Figure 4.5 and Figure 4.8 displays the output images obtained for $\lambda = 0.3$.
   - Mid Sparsity Region of Operation ($\lambda = 0.4, 0.5, 0.6$): The Mean Square Error in this region is higher than before and increases as the noise in the source image increases. This effect and the corresponding output images for $\lambda = 0.4$ are displayed in Figures 4.6 and 4.9 respectively.
   - High Sparsity Region of Operation ($\lambda = 0.7, 0.8, 0.9$): The Mean Square Error between each output image and fixed target image was highest in this region and increases with the noise in each source image. Figures 4.7 and 4.10 display the results obtained by setting $\lambda = 0.9$.

   In each of the Figures from 4.5 to 4.7, the Baseline Error denotes the error between the target image and each source image considered. As explained in [11] this is a characteristic of the Brainweb phantoms themselves [5], while the Residual Error denotes the Mean Square Error between the target image and each output image obtained by using our contrast synthesis method based on Joint Dictionary Learning. Table 4.3 display the obtained tissue contrast differences post transformation of the source image.

Noise in Source Image	n=0	n=1	n=3	n=5	n=7	n=9
KL Divergence ($\lambda = 0.3$)	1.3438	1.3829	1.4373	1.4420	1.2794	1.3024
KL Divergence ($\lambda = 0.4$)	1.2678	1.2288	1.2286	1.2363	1.1964	0.2892
KL Divergence ($\lambda = 0.9$)	1.1604	2.1987	0.9017	0.8704	0.8621	0.7174

Table 4.3: Difference in Tissue Contrasts After Transformation: Dataset 1 with $d = 27$

2. **Results with $5 \times 5 \times 5$ patches** : As the size of the cubical neighbourhood around a voxel increases, the

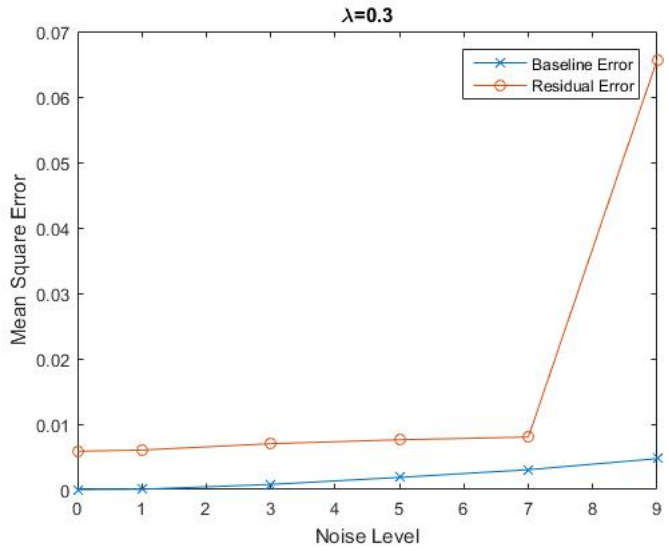

Figure 4.5: Dataset 1: Joint Dictionary Learning with $d = 27$ and $\lambda = 0.3$

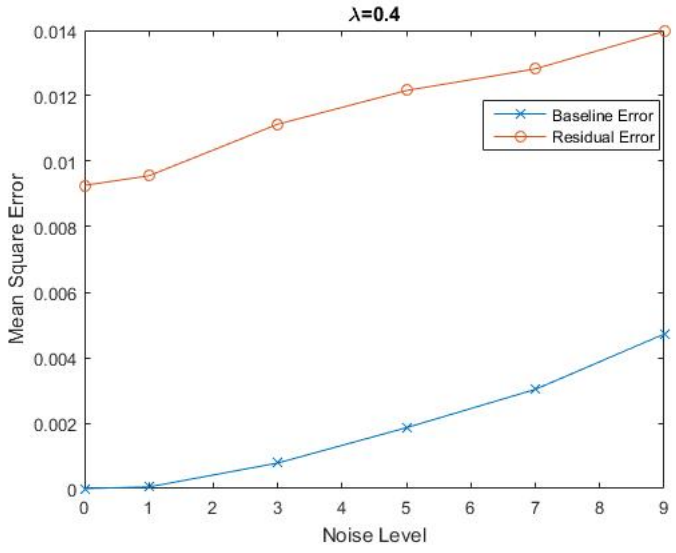

Figure 4.6: Dataset 1: Joint Dictionary Learning with $d = 27$ and $\lambda = 0.4$

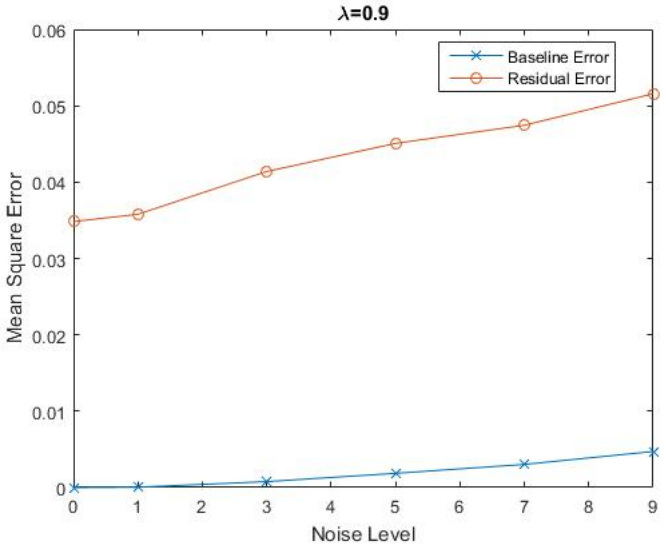

Figure 4.7: Dataset 1: Joint Dictionary Learning with $d = 27$ and $\lambda = 0.9$

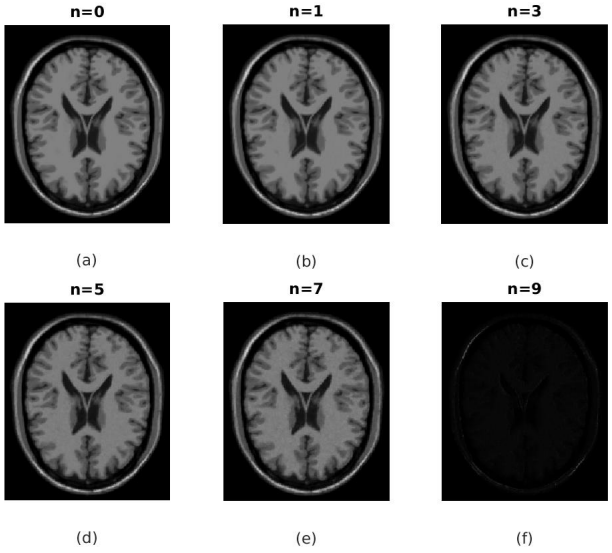

Figure 4.8: Dataset 1: Output Images $d = 27$ and $\lambda = 0.3$

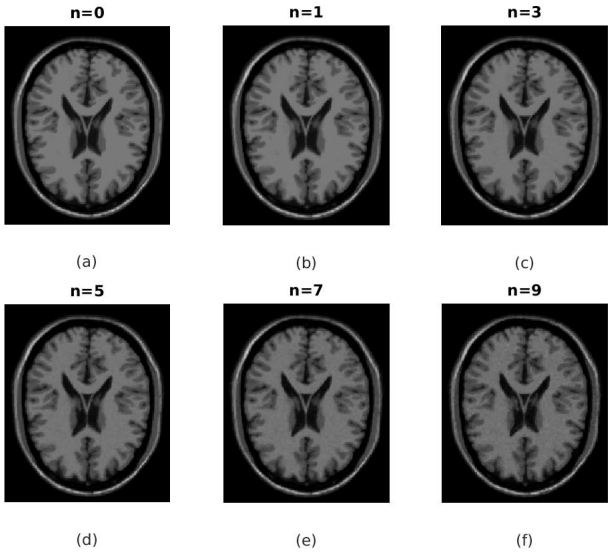

Figure 4.9: Dataset 1: Output Images $d = 27$ and $\lambda = 0.4$

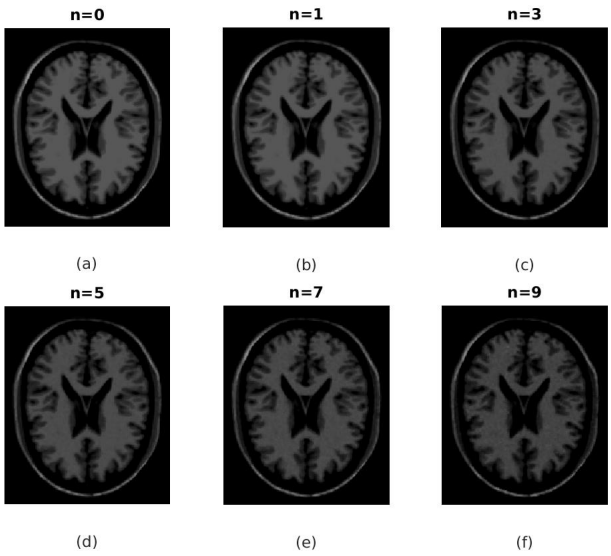

Figure 4.10: Dataset 1: Output Images $d = 27$ and $\lambda = 0.9$

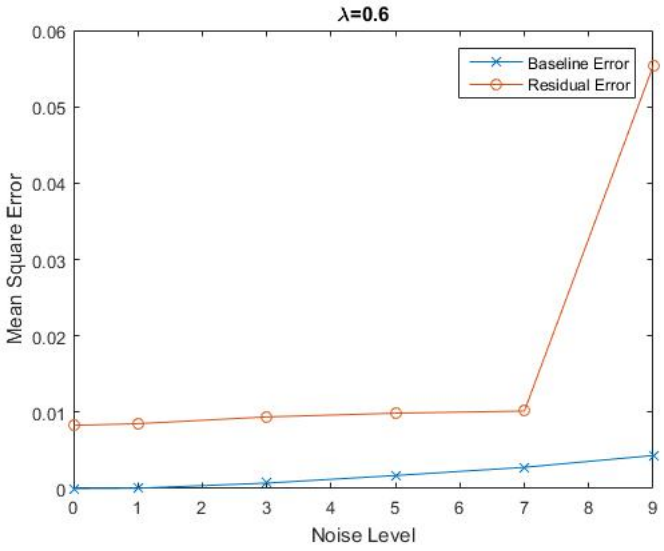

Figure 4.11: Dataset 1: Joint Dictionary Learning with $d = 125$ and $\lambda = 0.6$

amount of local image information contained in a training signal increases as well. This effect was considered to improve the performance of the Joint Dictionary Learning approach since the dictionaries obtained through the training phase will be more representative of the source and target images respectively. However this performance comes at the price of an increase in the dimension of each training signal. Specifically the increased dimension due to the larger neighbourhood is $d = 125$. Since we are considering over-complete dictionaries, the number of dictionary columns needs to be increased to satisfy the over-completeness criteria as explained before. This number was arbitrarily chosen as $K = 250$ for this experiment.
Two regions of operation could be identified for this experiment based on the Average Mean Square Error between the target image and each output image.

- Low-to-Mid Sparsity Region of Operation ($\lambda = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6$) Similar to the experiment with the smaller patch size the Average Mean Square Error between the target image and output images corresponding to low-noise source images is small. This performance however degrades quickly as the noise in the source image increases. This effect was observed for even higher values of λ till $\lambda = 0.6$ where only the error between the target image and the output image corresponding to the T_1-w phantom with $n = 9$ is substantially high as compared to the other output images. Figures 4.11 and 4.13 display the comparison between the Baseline Error and the Residual Error and the output images obtained for $\lambda = 0.6$ in this case respectively.
- High Sparsity Region of Operation ($\lambda = 0.7, 0.8, 0.9$) The performance in terms of the Average Mean Square Error worsens for every output image as the noise in the source image is increased in this region. However the difference between the Residual Error for different levels of noise is small and lower than that obtained for the High Sparsity Region when $d = 27$. Figures 4.12 and 4.14 display the obtained results and output images for $\lambda = 0.8$ respectively.

Similar to the earlier experiment, the Baseline Error in each corresponding figure denotes the difference between the T_1-w phantoms considered as source and target respectively while the Residual Error denotes the difference between the target image and the transformed source image in each case. Table 4.4 displays the differences in tissue contrast post transformation in this case for $\lambda = 0.6, 0.8$.

Noise in Source Image	n=0	n=1	n=3	n=5	n=7	n=9
KL Divergence ($\lambda = 0.6$)	1.2079	1.2081	1.2474	1.2059	1.1530	1.6186
KL Divergence ($\lambda = 0.8$)	1.2511	1.2542	1.2550	1.3908	1.3708	1.3177

Table 4.4: Difference in Tissue Contrasts After Transformation: Dataset 1 with $d = 125$

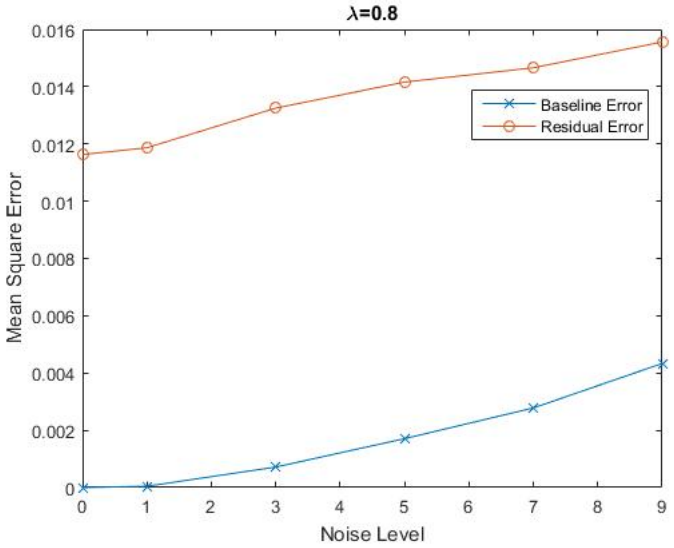

Figure 4.12: Dataset 1: Joint Dictionary Learning with $d = 125$ and $\lambda = 0.8$

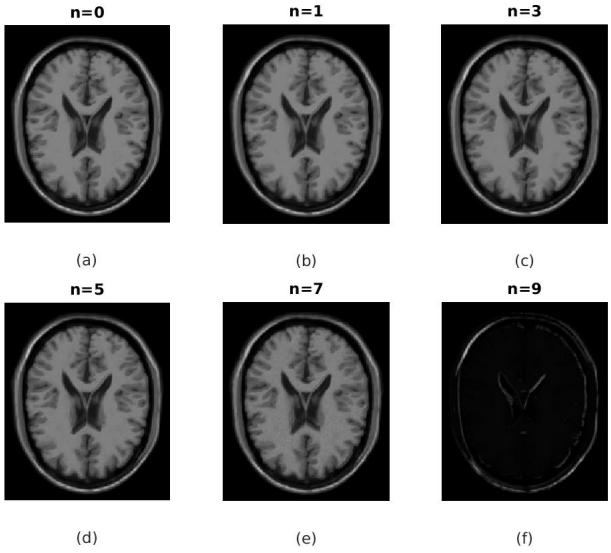

Figure 4.13: Dataset 1: Output Images $d = 125$ and $\lambda = 0.6$

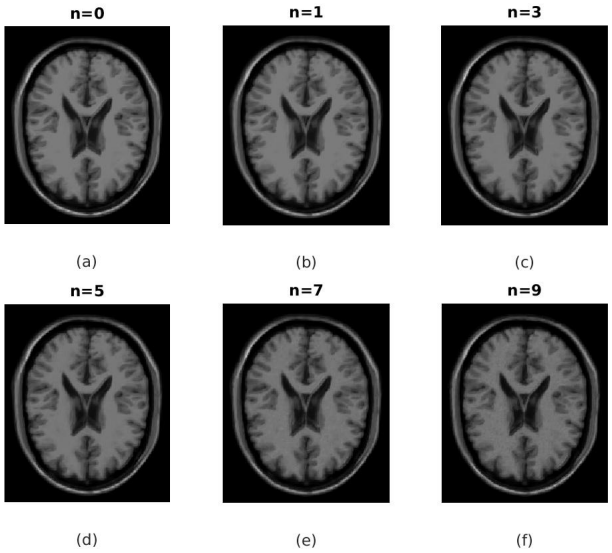

Figure 4.14: Dataset 1: Output Images $d = 125$ and $\lambda = 0.8$

4.4.3. Validating Performance: Dataset 2

In [12] the authors have used the MIMECS method to transform the noise-free T_1-w phantom into the noise-free T_2-w phantom. The Mean Square Error between the target image and the transformed source image is used to tune the parameter λ to be used for obtaining the sparse representation vectors using compressed sensing. Through the results reported in [12] it is clearly seen that the value of λ obtained using such a strategy is robust enough to be used on real-life images as well.

However, in order to keep consistency with the experiments performed earlier and those reported in Section 4.4.2 of [11] it was decided to not observe the performance of Joint Dictionary Learning on a single source-target pair. Instead the same source images used in the earlier experiment were retained, these were the T_1-w phantoms with $n \in \{0,1,3,5,7,9\}$ obtained from [5]. The target image used was however changed to the noise-free T_2-w phantom.
It is thought that intensity normalization between a pair of source and target images considered in this way can be used to understand the performance of a intensity-normalization method when the source image is noisy and there exist intensity differences between the source and target image due to the change in modality. The choice of using the T_1-w phantoms and T_2-w phantom as source and target respectively seems to be motivated by the fact that most real-life datasets of MRI images contain images that are either T_1-w or T_2-w. This is clearly a harder validation test for any intensity normalization method since the source and target image are made up of vastly different intensities and intensity distributions in addition to the presence of noise in each source image.

1. **Results with $3 \times 3 \times 3$ patches:** Owing to the same size of the cubical neighbourhood considered around each voxel, the dimension of each training signal was same as the earlier experiment, i.e $d = 27$. The dictionary size was kept the same as before, $K = 100$ as this size was found to be very convenient in the testing phase as mentioned earlier.
   In this experiment no regions of operation could be identified based on the observed Mean Square Error between the target image and each output image. Instead, the Mean Square Error obtained remains high for all values of λ with the error increasing slightly as the noise in the source image increases with higher values of λ having slightly higher error as well. Figures 4.15 to 4.17 display the result of observing the Mean Square Error between the target image and the transformed source image for different values of noise while Figures 4.18 to 4.20 display the output images obtained using these values of λ.

   From the output image in the lowermost right corner of Figure 4.18 we observe the same effect at low sparsity ($\lambda = 0.3$) as that observed with the images from Dataset 1. This result is discussed later in this chapter. One important distinction from the earlier experiment performed on Dataset 1 is that

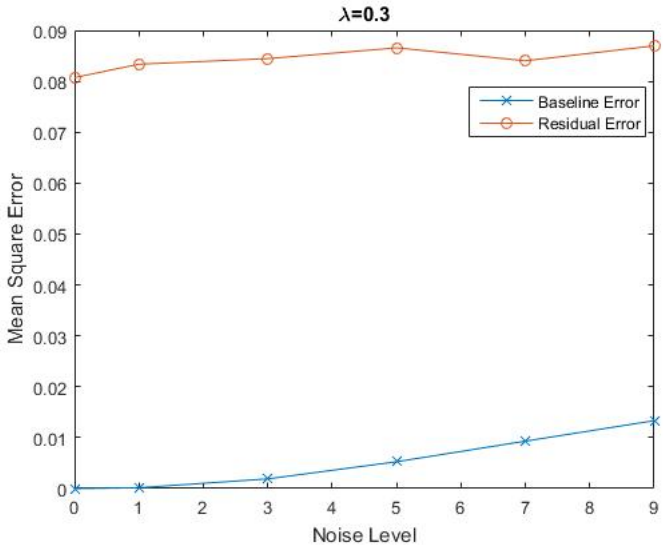

Figure 4.15: Dataset 2: Joint Dictionary Learning with $d = 27$ and $\lambda = 0.3$

the Baseline Error in this experiment measures the Mean Square Error between the target image, i.e. the noise-free T_2-w phantom and a similar T_2-w phantom but one having the same noise level as that of the source image similar to the experiment reported in Section 4.4.2 in [11]. Following this Table 5 displays the tissue contrast difference between the target image and each transformed source image for the values of λ reported above.

Noise in Source Image	n=0	n=1	n=3	n=5	n=7	n=9
KL Divergence ($\lambda = 0.3$)	1.5720	1.5095	1.5393	2.0254	1.5569	1.5023
KL Divergence ($\lambda = 0.4$)	1.4789	1.4057	1.5462	1.6082	1.4929	2.0024
KL Divergence ($\lambda = 0.9$)	1.8316	1.4160	0.2756	0.2551	0.2354	0.2231

Table 4.5: Difference in Tissue Contrasts After Transformation: Dataset 2 with $d = 27$

2. **Results with $5 \times 5 \times 5$ patches**: As mentioned earlier, the dimension of each training signal increases in cubical proportion to the size of the neighbourhood considered around a centre voxel. Hence the size of the training signal in this case was the same as before, i.e. $d = 125$. The dictionary size was retained as $K = 250$ as well, while λ was varied in the range $\lambda \in \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$. For this experiment too no regions could be observed based on the Average Mean Square Error. The error stays high and only varies slightly as the source image becomes more noisy. These observations are plotted in Figures 4.21 and 4.22 with Figures 4.23 and 4.24 displaying the output images for $\lambda = 0.6$ and $\lambda = 0.8$ respectively. Observe once more that the output image corresponding to the T_1-w phantom with $n = 9$ as source contains almost no observable tissue mass as seen from the lowermost right corner in Figure 4.23. Table 6 displays the difference in tissue contrast measured between the target image and each output image for the values of λ considered above.

Noise in Source Image	n=0	n=1	n=3	n=5	n=7	n=9
KL Divergence ($\lambda = 0.6$)	2.0378	2.0658	2.0588	2.0964	1.8450	1.9647
KL Divergence ($\lambda = 0.8$)	2.1240	2.3081	2.0784	1.6499	1.6519	1.6269

Table 4.6: Difference in Tissue Contrasts After Transformation: Dataset 2 with $d = 125$

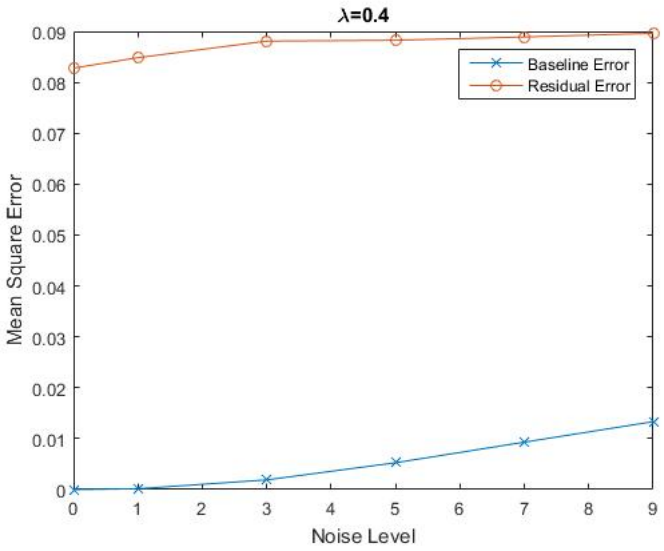

Figure 4.16: Dataset 2: Joint Dictionary Learning with $d = 27$ and $\lambda = 0.4$

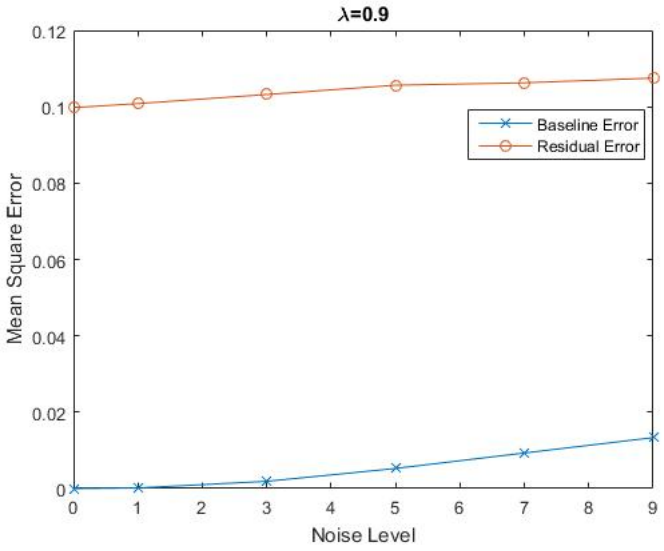

Figure 4.17: Dataset 2: Joint Dictionary Learning with $d = 27$ and $\lambda = 0.9$

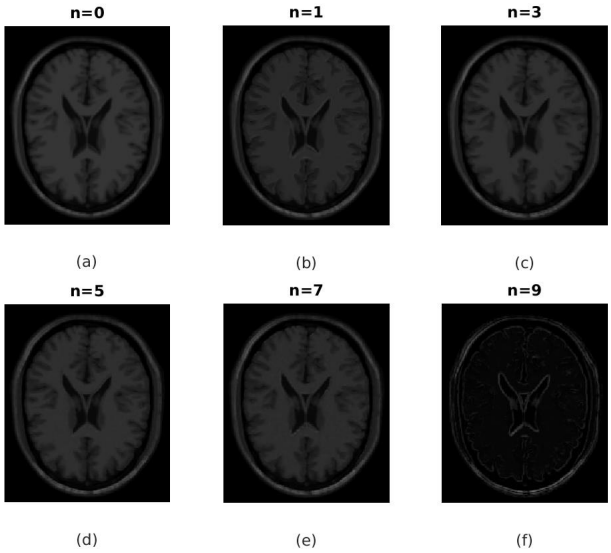

Figure 4.18: Dataset 2: Output Images $d = 27$ and $\lambda = 0.3$

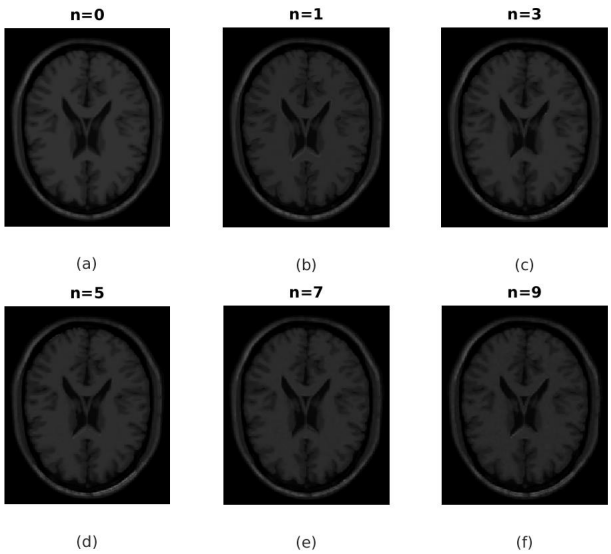

Figure 4.19: Dataset 2: Output Images $d = 27$ and $\lambda = 0.4$

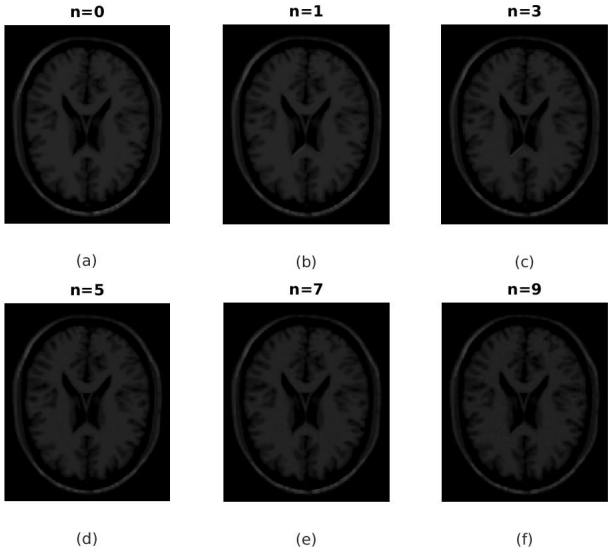

Figure 4.20: Dataset 2: Output Images $d = 27$ and $\lambda = 0.9$

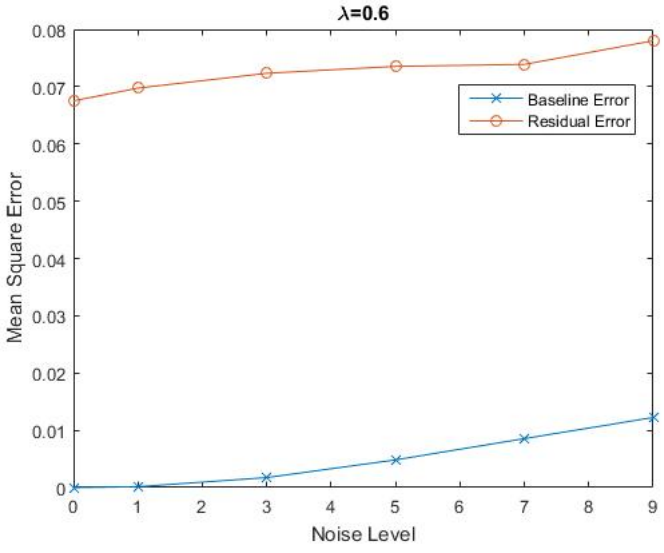

Figure 4.21: Dataset 2: Joint Dictionary Learning with $d = 125$ and $\lambda = 0.6$

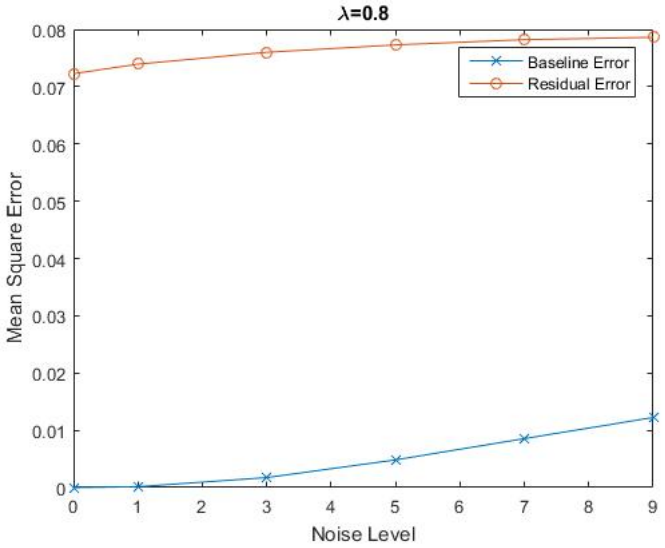

Figure 4.22: Dataset 2: Joint Dictionary Learning with $d = 125$ and $\lambda = 0.8$

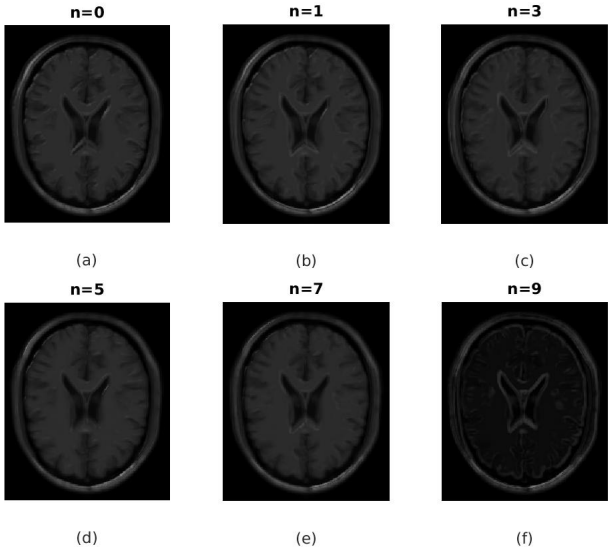

Figure 4.23: Dataset 2: Output Images $d = 125$ and $\lambda = 0.6$

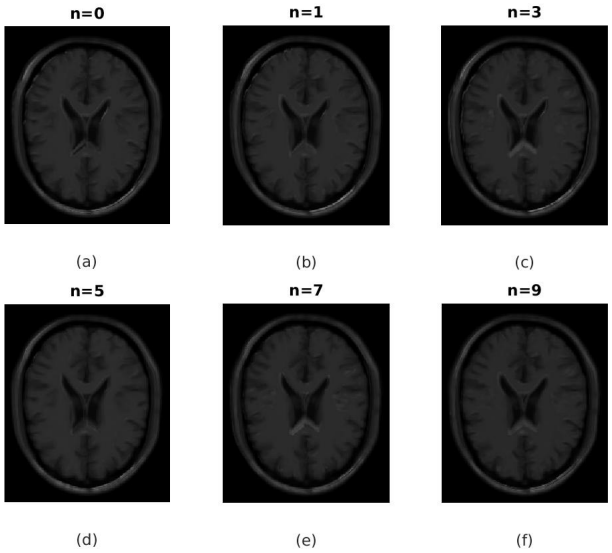

Figure 4.24: Dataset 2: Output Images $d = 125$ and $\lambda = 0.8$

4.5. Discussion of Experimental Results

For each experiment reported above we observe that the performance of Contrast Synthesis method presented in Section 3.10 is not satisfactory in terms of the Mean Square Error between the target image and the transformed source image. As reported in [11] achieving lower Residual Error as compared with the Baseline Error indicates a robustness of an intensity normalization method in producing synthetic images similar to the target image. This is clearly not the case with any of the experimental results reported earlier. Further, based on the output images and obtained KL-Divergences after transformation, we can conclude that the contrast synthesis method considered will not be useful with the simulated phantom images from either datasets.

Real-life images would require registration to produce a one-to-one correspondence between the voxels of the source and target images. This is particularly true for images obtained from different scanners, since these images are often misaligned with respect to each other. The performance of any contrast synthesis method considered will therefore be dependent on the registration accuracy as well. Hence due to the poor performance on ground truth images as well as the possibility of a further degradation in the performance as a result of improper registration, this method was not applied on any real-life images.

This section analyses the failure of the Joint Dictionary Learning method in all of the experiments reported above.

4.5.1. Results With Dataset 1

By comparing Figures 4.2, 4.9 and 4.14 we observe that the images at the output of the contrast synthesis method appear as de-noised versions of the source image in each case. However the high Residual Error as well as the KL-Divergence obtained after transformation clearly indicate that the method has failed to produce the correct output. We observe that the method fails even when the source and target image are identical, that is the T_1-w phantom $n = 0$ is considered as both source and target. The difference between this target image and the image obtained at the output of the contrast synthesis algorithm can be clearly seen through the increase in the KL-Divergence as reported in the first column of both Tables 4.3 and 4.4.

This can possibly be explained by considering the difference images displayed in Figures 4.25 and 4.26 for the $3 \times 3 \times 3$ neighbourhood corresponding to $d = 27$ and $\lambda = 0.4$ and $\lambda = 0.9$ respectively and Figure 4.27 for the $5 \times 5 \times 5$ neighbourhood corresponding to $d = 125$ and $\lambda = 0.8$. Observe that each difference image contains a significant percentage of gray intensities. This is especially true in the areas where the output image intensity is constant or varies very little as observed from Figures 4.9 to 4.10 and 4.14. Based on this observation, we speculate that each output image from Figures 4.9 and 4.14 contains a small positive offset (indicated by the gray intensities) with respect to the target image, the T_1-w phantom with $n = 0$. The presence of such

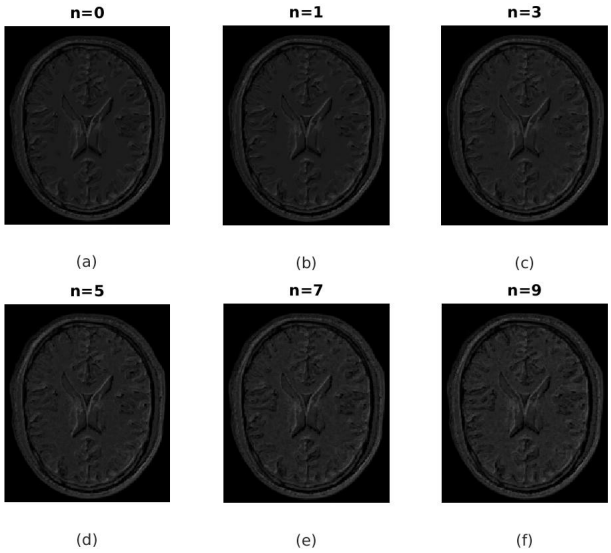

Figure 4.25: Dataset 1: Difference Images $d = 27$ and $\lambda = 0.4$

an offset in each voxel of the output image would result in the high Mean Square Error between the target image and each output image as reported earlier.
This speculation seems to be further supported by observing the intensities in Figure 4.26 in comparison to Figures 4.25 and 4.27. The gray level intensities in each output image in Figure 4.26 appear brighter, indicating a higher offset between the target image and each transformed source image. This increase in the offset would lead to the higher Mean Square Error when $d = 27$ and $\lambda = 0.9$ as reported earlier.
It is considered that the presence of such an offset results in the higher difference in tissue contrast post transformation between an output image at the fixed target image. If we observe Table 4.3 the values of a few columns seem to suggest that the difference in tissue contrast has reduced as a result of the contrast synthesis method, however the higher Mean Square Error between the target and these output images clearly proves otherwise. This could be a result of the heuristic estimate of KL-Divergence between the pair of histograms of discrete images described in Section 4.2.5. Hence this suggests that the metric used to observe tissue contrast differences needs to further considered.

4.5.2. Results With Dataset 2

Similar to the earlier discussion we consider the difference images shown in Figures 4.28, 4.29 for Dataset 2 images with $d = 27$ and $\lambda = 0.4$, $\lambda = 0.9$ respectively and Figure 4.30 for the $5 \times 5 \times 5$ neighbourhood corresponding to $d = 125$ and $\lambda = 0.8$.

Following this we observe that each difference image displays a high density of black intensities especially in areas where the intensity within each output image is constant. We can compare the target image displayed in Figure 4.3 and the output images shown in Figures 4.19, 4.20 and 4.24. By observing the locations containing black intensities in the difference images and the corresponding locations from the target and output images, we can conclude that these are negative intensities. These appear to be black (indicated by zero) as they have been scaled to zero for displaying each difference image. Hence the offset between the target and an output image is speculated to have positive and negative values in this case. In addition, we observe the presence of higher intensities displayed by white (corresponding to unity) in each difference image. Hence the absolute value of the offset between corresponding voxels from the target image and each output image is thought to have increased as well in this case.

Earlier it was reported that the Mean Square Error between each pair of output images and the target image of Dataset 2, was higher than that reported for Dataset 1. It is speculated that this increase is due to the increased offset between the target image and each corresponding output image. Such an offset could also explain the higher values of KL-Divergence after transformation similar to that considered with the out-

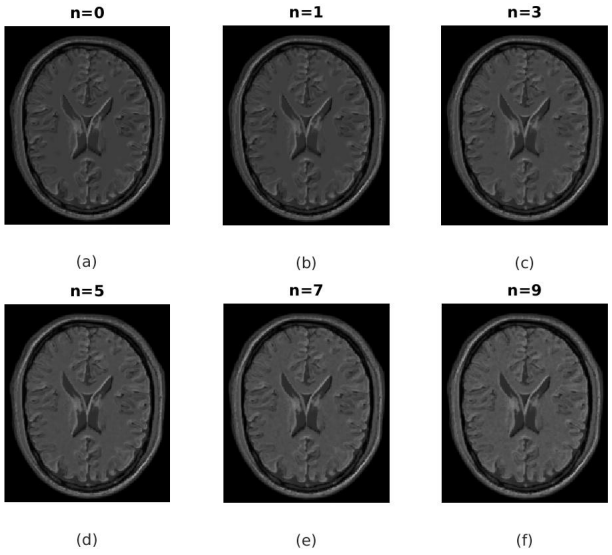

Figure 4.26: Dataset 1: Difference Images $d = 27$ and $\lambda = 0.9$

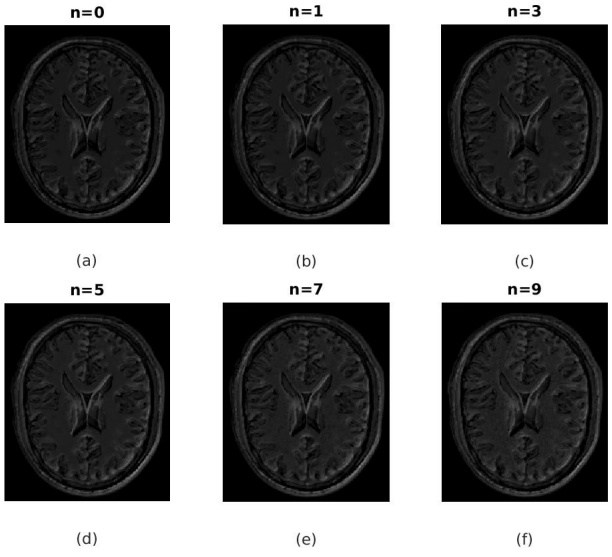

Figure 4.27: Dataset 1: Difference Images $d = 125$ and $\lambda = 0.8$

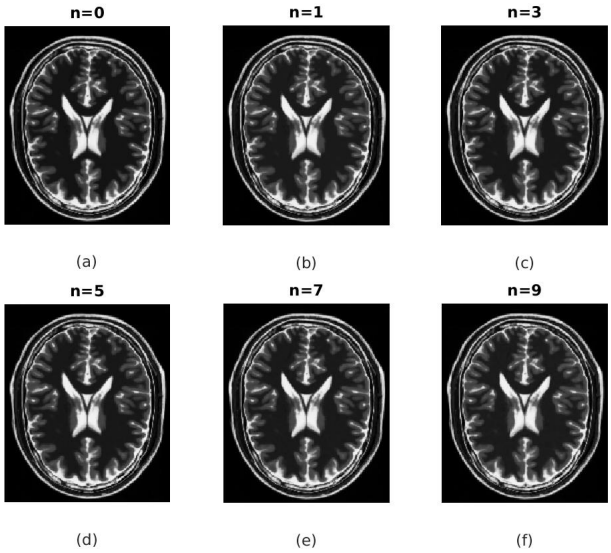

Figure 4.28: Dataset 2: Difference Images $d = 27$ and $\lambda = 0.4$

put images of Dataset 1. Once again we observe that the values of KL-Divergence post transformation have reduced for some columns in Table 4.5. The high error between the target image and each output image however clearly indicate that the intensities in the output image are different from those in the target. It is hence clear that these images do not have similar tissue contrast, an effect also observed in each output image from Figures 4.19, 4.20 and 4.24. Hence it is clear that the metric used to observe the differences in tissue contrast needs adjustment.

4.5.3. Sub-Optimal Estimate of Sparse Vectors

The main steps of the contrast synthesis method from Section 3.10 are as follows. Given a pair of input images considered as source and target respectively, we divide each image into d-dimensional training signals. Using each pair of source and target training signals we jointly-learn a pair of dictionaries using Equation 3.21. These are termed as the source dictionary $\hat{D}_1$ and target dictionary $\hat{D}_2$ respectively. Each dictionary is constrained to have the same sparse representation vector for a corresponding pair of training signals.

Since we are interested in transforming each source image patch into the corresponding target intensity-space, we consider that only the source image is available during the testing phase. Further we assume:

$$y_1(i) \approx \hat{D}_1 \boldsymbol{\alpha}(i) \quad \forall i = 1,2,3,...N \tag{4.9}$$

where $y_1(i)$ denotes the i^{th} patch of the source image and $\boldsymbol{\alpha}(i)$ its sparse representation vector. The latter can be estimated by using the LARS algorithm [8] implemented by the function mexLasso of the SPAMS Toolbox [7] with the following formulation:

$$min_{\boldsymbol{\alpha}(i)\in\mathbb{R}^K} \frac{1}{2} \left\| y_1(i) - \hat{D}_1 \boldsymbol{\alpha}(i) \right\|_2^2 + \lambda \left\| \boldsymbol{\alpha}(i) \right\|_1 \quad \forall i = 1,2,3,...N \tag{4.10}$$

Once the estimated sparse vector is obtained, we can calculate the transformed image patch as follows:

$$\hat{y_2}(i) = \hat{D}_2 \boldsymbol{\alpha}(i) \quad \forall i = 1,2,3,...N \tag{4.11}$$

As reported in Section 3.7 of [9] estimating the sparse representation vectors using Equation 4.10 is not fully satisfactory of the original dictionary learning cost function given by:

$$min_{\hat{D}_1,\hat{D}_2\in C} \lim_{N\to+\infty} \frac{1}{N} \sum_{i=1}^{N} min_{\boldsymbol{\alpha}(i)} \left(\frac{1}{2} \left\| y_1(i) - \hat{D}_1 \boldsymbol{\alpha}(i) \right\|_2^2 + \frac{1}{2} \left\| y_2(i) - \hat{D}_2 \boldsymbol{\alpha}(i) \right\|_2^2 + \lambda \left\| \boldsymbol{\alpha}(i) \right\|_1 \right) \tag{4.12}$$

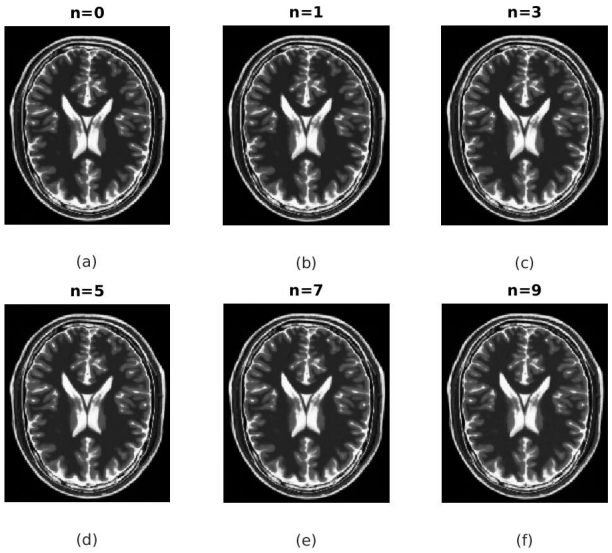

Figure 4.29: Dataset 2: Difference Images $d = 27$ and $\lambda = 0.9$

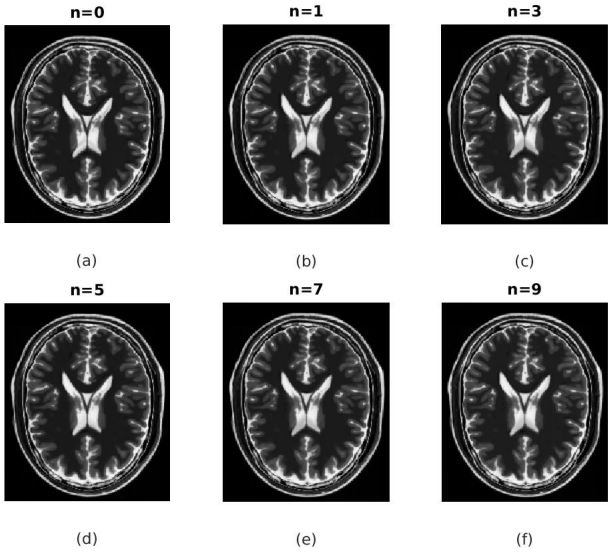

Figure 4.30: Dataset 2: Difference Images $d = 125$ and $\lambda = 0.8$

Instead each representation vector should ideally be obtained by using the following formulation:

$$min_{\boldsymbol{\alpha}^*(i)\in\mathbb{R}^K} \frac{1}{2}\left\|y_1(i)-\hat{D_1}\boldsymbol{\alpha}(i)\right\|_2^2+\frac{1}{2}\left\|y_2(i)-\hat{D_2}\boldsymbol{\alpha}(i)\right\|_2^2+\lambda\left\|\boldsymbol{\alpha}(i)\right\|_1 \quad \forall i=1,2,3,...N \tag{4.13}$$

The transformed image patches would then be obtained using:

$$\hat{y_2}(i)=\hat{D_2}\boldsymbol{\alpha}^*(i) \quad \forall i=1,2,3,...N \tag{4.14}$$

This suggests that the estimate of the sparse vectors obtained during the testing phase of our contrast synthesis algorithm is sub-optimal with respect to the target dictionary $\hat{D_2}$. This is considered as the main reason behind the poor performance of the Joint Dictionary Learning approach in producing the required output image for all pairs of source and target images considered. The strange results observed at low values of λ with both datasets of images as reported earlier could also be possibly explained by this observation.

On observing Equation 4.10, we notice that during the testing phase, the estimate of the sparse vectors depends on only the patches from the source image and the chosen value of λ. Hence it is reasonable to expect that the estimate becomes poor as the noise in each source image patch increases, which leads to a poor estimate of the transformed image patches and in turn the output image. This corresponds to the observed results of increasing error as the noise level in the source increases.
As mentioned previously, at low values of λ the output images are especially poor and contain almost no observable brain mass. This effect is observed for all source images containing noise, that is $n \in \{1,3,5,7,9\}$ at $\lambda = 0.1$. In addition, this effect is observed for all output images irrespective of patch size and modalities of the source and target image. As λ increases, this effect stops being prominent and finally at a certain value of λ that is dependent on the choice of d, all images can be observed but are thought to contain a offset as mentioned previously. Hence it is speculated that the sub-optimal estimate becomes particularly poor at low values of λ, however this effect is not fully understood yet.

4.6. Summary

This chapter describes the different experiments performed using ground truth images for Brain MRI data obtained from the Brainweb Simulated Brain Database [5]. By considering two different datasets, tissue contrast differences in phantom images can be simulated similar to those which occur in real-life Brain MRI images. The two scenarios considered are when the MRI images for a patient are obtained from a single scanner and when the imaged data from a multi-site multi-scanner study is pooled together. Using these datasets the performance of the contrast synthesis method developed at the end of Chapter 3 was evaluated similar to the strategy reported in [11].

Through each of the experiments reported above, it was observed that the performance of the method developed in Section 3.10 was not satisfactory for both datasets of simulated images. When the source and target image are of the same modality, each output image appears similar to the target image. However the Mean Square Error between transformed images and the original target image is higher than the Baseline Error threshold. This indicates that the method is not able to transform a source image as required. The poor performance is also reflected by the values of KL-Divergence between the normalized histograms of the target image and each transformed image. For source and target images belonging to different modalities, both visual output and Mean Square Error obtained are indicative of an erroneous result for each output image.

An observation from [9] suggests that during the test phase our employed strategy for intensity-normalization uses an estimator that is sub-optimal with respect to the target dictionary $\hat{D_2}$. This is speculated to be the main reason behind the poor performance and hence the contrast synthesis method developed needs further consideration. The next chapter concludes this report and suggests a few modifications which could be used to solve these problems.

5

Conclusion

This chapter concludes the report for this Master Thesis by briefly revisiting the research question considered at the start of this project. Following this each of the previous chapters is quickly summarized once again before offering some suggestions and areas of improvements over the main method of Joint Dictionary Learning described previously.

The intensities present in MRI images vary considerably with the different choice of pulse sequences, their parameters and external effects such as scanner make and calibration [11]. A fundamental property of MRI images that changes due to the differences in intensities is the tissue contrast in each image [11], [12]. As reported in these works inconsistency in segmentation performance achieved on MRI images obtained from different scanners can be attributed in part to this difference in tissue contrast. The research goal of this project was to verify if Joint Dictionary Learning can be used for intensity-normalization in ground truth Brain MRI images obtained from [5] such that the tissue contrasts in each output image obtained thus were normalized with respect to a fixed target image. It was expected that the tissue contrast of each output image would be similar to that of the fixed target image.

Chapter 2 provides the literature review related to prior work done in the areas of supervised segmentation, registering test images to a known labelled atlas and intensity normalization. The common goal of the former two methods is to generate consistent segmentation on MRI images obtained from different scanners. The latter focuses on pre-processing strategies that can generate the required tissue intensities in MRI images such that the images can be used with any post-processing method for consistent results. The original work behind the MIMECS method [11], [12] is described in detail in this chapter and inspirations are drawn from it for further use.

One common problem with the methods from Chapter 2 is the applicability of each method to general datasets of MRI images without the use of atlas images. To address this issue we consider algorithms from the area of dictionary learning. This is a category of representation learning techniques within machine learning that allow us to learn a sparse representation for each training signal based on an over-complete dictionary. Chapter 3 explains the choices made with respect to selecting the most appropriate dictionary learning method for the task of intensity normalization in MRI images to generate synthetic tissue contrast given only a pair of images from the same patient.

In Chapter 4 the different experiments performed to evaluate the performance of the chosen dictionary learning method are described. For performing these experiments ground truth MRI images were obtained from the Brainweb Simulated Brain Database [5]. Two different datasets of simulated MRI images were created to evaluate the performance of the Joint Dictionary Learning method. It is thought that these datasets simulate tissue contrast differences similar to what might occur in real-life images under two different scenarios. Through the experiments it is established that the considered method cannot be used for generating the required output images given any pair of source and target images. These results are discussed at the end of the chapter and the proposed solutions are mentioned below.

The first reason considered for the poor performance of the Joint Dictionary Learning is the generation of an offset in each output image with respect to the fixed target image. It was observed that difference images for the dataset of mono-modal images indicate the presence of relatively small positive offsets, while those for the dataset of multi-modal images indicate the presence of positive and negative offsets having higher absolute values. Due to such an offset, the Mean Square Error between the target image and the output images is higher than the baseline threshold between the source and target images. It also results in higher differences in the tissue contrast between images than those reported previously for each source-target pair.
This could possibly be solved by considering a better pre-normalizing strategy than the one considered in subsection 4.2.3. Commonly known strategies such as centering and scaling each image patch to unit variance resulted in poorer output images and were hence not considered further. For multi-modal images such as those from Dataset 2, the intensity distributions are vastly different for the source and target image. Hence any pre-normalization strategy will need to take this difference into account in its implementation for such datasets.

Another observation made at this point is that the method of observing the tissue contrast by means of the heuristic estimate of the KL-Divergence between a pair of normalized image histograms as suggested in [4] is not good in practice. Observed results indicate that such an estimate falsely points to a reduction in tissue contrast differences, even in the presence of high error between the target image and a particular output image. This suggests that a better estimate is needed to quantify the difference in tissue contrast between MRI images, particularly for images which contain discrete intensities.

It is thought that the main reason of the failure behind the Joint Dictionary Learning method is the sub-optimal testing phase of the method as explained in subsection 4.5.3. This is considered to be the reason behind the strange results observed with the output images at low values of λ. Any estimate of the sparse vectors using Equation 4.10 is not fully satisfactory of the original cost function considered in the Joint Dictionary Learning in Equation 4.12.
One strategy considered was to scale the cost function to give higher importance to the reconstruction term associated with the source image patches in Equation 3.19. Then the estimate used in the testing phase was thought to improve, since the sparse vectors associated with the dictionary pair now rely more on the source image patches and the source dictionary $\hat{D_1}$. However this strategy was not found to work well in practice. The estimate of sparse vectors given by Equation 4.13 also cannot be directly used since the target patches $y_2(i)\ \forall i = 1,2,3..N$ are assumed to be unknown in the testing phase. Hence there is a need for another method of learning the source-target dictionary pair which allows the sparse vector to be estimated using only the source image patches in the testing phase.

As it turns out, there is a small body of previous work devoted to addressing the problem of the sub-optimal estimate as reported above. These methods are described in detail in Section 3.7 of [9]. We feel that the strategy proposed using the regression formulation is particularly important in this context. The main advantage of such a strategy is that in the training phase each sparse representation vector is always computed by using the source image patches only. While such a strategy loses the ability to jointly learn a pair of dictionaries from the source-target pair of images, it allows us to estimate the sparse vectors using only the source image patches in the testing phase. This makes the testing phase of the suggested method less sub-optimal than the one considered in this project.

Unfortunately this strategy could not be studied further in the time provided for this Thesis. Hence it should be considered as the staring point in any further research for the task of intensity normalization using a pair of over-complete dictionaries. Such a strategy would retain the advantages of dictionary learning as proposed at the start of Chapter 3, and would be more readily applicable to general datasets of Brain MRI images without needing high quality atlas images in its implementation.

Bibliography

[1] Maysam Shahedi(updated Sep 2016). *imshow3D*. MATLAB Central File Exchange, Retrieved Oct 17, 2016. URL https://nl.mathworks.com/matlabcentral/fileexchange/41334-imshow3d--3d-imshow--new-version-released--see--imshow3dfull-.

[2] Michal Aharon, Michael Elad, and Alfred Bruckstein. K-SVD: An Algorithm For Designing Overcomplete Dictionaries For Sparse Representation. *IEEE Transactions On Signal Processing*, 54(11):4311–4322, 2006.

[3] Laszlo Balkay(2011). *Read minc format medical image file*. MATLAB Central File Exchange, Retrieved Sep 1, 2016. URL https://nl.mathworks.com/matlabcentral/fileexchange/32644-loadminc.

[4] Sung-Hyuk Cha. Comprehensive Survey On Distance/Similarity Measures Between Probability Density Functions. *International Journal Of Mathematical Models And Methods In Applied Sciences*, 1(2):1, 2007.

[5] Chris A Cocosco, Vasken Kollokian, Remi K-S Kwan, G Bruce Pike, and Alan C Evans. Brainweb: Online Interface to a 3D MRI Simulated Brain Database. In *NeuroImage*, 1997. URL http://www.bic.mni.mcgill.ca/brainweb/.

[6] Rafael Gonzalez and Richard Woods. *Digital Image Processing*. Pearson Education, 2009.

[7] Julien Mairal, Francis Bach, Jean Ponce, Guillermo Sapiro, Rodolphe Jenatton, and Guillaume Obozinski. SPAMS-Optimization Toolbox,v.2.5. URL http://spams-devel.gforge.inria.fr/index.html.

[8] Julien Mairal, Francis Bach, Jean Ponce, and Guillermo Sapiro. Online Dictionary Learning For Sparse Coding. In *Proceedings of the 26th Annual International Conference On Machine Learning*, pages 689–696. ACM, 2009.

[9] Julien Mairal, Francis Bach, and Jean Ponce. Sparse Modeling For Image And Vision Processing. *arXiv preprint arXiv:1411.3230*, 2014.

[10] MATLAB. *version 8.6.0.267246 (R2015b)*. The MathWorks Inc., Natick, Massachusetts, 2015.

[11] Snehashis Roy. *MR Image Contrast Synthesis For Consistent Segmentation*. PhD thesis, The Johns Hopkins University, 2013.

[12] Snehashis Roy, Aaron Carass, and Jerry Prince. A Compressed Sensing Approach For MR Tissue Contrast Synthesis. In *Biennial International Conference on Information Processing in Medical Imaging*, pages 371–383. Springer, 2011.

[13] Ron Rubinstein. K-SVD Toolbox, v.13. URL http://www.cs.technion.ac.il/~ronrubin/software.html.

[14] Tong Tong, Robin Wolz, Pierrick Coupé, Joseph V Hajnal, Daniel Rueckert, Alzheimer's Disease Neuroimaging Initiative, et al. Segmentation of MR images via discriminative dictionary learning and sparse coding: Application to hippocampus labeling. *NeuroImage*, 76:11–23, 2013.

[15] Annegreet van Opbroek, Hakim C Achterberg, and Marleen de Bruijne. Feature-Space Transformation Improves Supervised Segmentation Across Scanners. In *Medical Learning Meets Medical Imaging*, pages 85–93. Springer, 2015.

[16] Jianchao Yang, John Wright, Thomas S Huang, and Yi Ma. Image Super-Resolution via Sparse Representation. *IEEE Transactions On Image Processing*, 19(11):2861–2873, 2010.