IMT School for Advanced Studies Lucca
Lucca, Italy

Joint Registration and Segmentation of CP-BOLD MRI

Ph.D. Program in Computer, Decision, and Systems Science
XXIX Cycle

By
ILKAY OKSUZ
2017

Table of Contents

List of Figures

List of Tables

Vita

January 29, 1987	Born, Istanbul, Turkey
2006–2010	B.sc. in Electronics Engineering Istanbul Technical University Istanbul, Turkey
2008–2009	Exchange Student Darmstadt Technical University Darmstadt, Germany
2010–2013	M.Sc. in Electrical and Electronics Engineering Bahcesehir University Istanbul, Turkey
2013–Date	Phd. in Computer, Decision and Systems Science Pattern Recognition and Image Analysis Research Unit IMT Institute for Advanced Studies Lucca, Italy
2015–2016	Visiting Researcher Magnetic Resonance Research Center Image Processing and Analysis Group Yale University New Haven, CT, USA
2017	Visiting Researcher Institute for Digital Communications University of Edinburgh Edinburgh, UK

Publications

[1] **Oksuz, I.**, Dharmakumar, R., Tsaftaris, S.A., "Joint Myocardial Registration and Segmentation of Cardiac BOLD MRI", Statistical Atlases and Computational Models of the Heart (STACOM), 2017. **PDF**

[2] **Oksuz, I.**, Mukhopadhyay A., Dharmakumar, R., Tsaftaris, S.A., "Unsupervised Myocardial Segmentation for Cardiac BOLD", IEEE Transactions on Medical Imaging (TMI) 2017. **PDF**

[3] Yang H.S.*, **Oksuz, I.***, Klein M., Sobczyk O., Dey D., Sykes J., Butler J., Bi X., Sharif B., Cokic I., Li, D., Slomka D., Prato F.S., Fisher J., Tsaftaris S.A., Dharmakumar, R., "Cardiac fMRI - A Novel Approach for Reliably Detecting Myocardial Oxygenation Changes with Precise Modulation of Arterial CO2", International Society of Magnetic Resonance in Medicine Meeting (ISMRM), 2017. **PDF**

[4] Suinesiaputra A., Ablin P., Alba X.,Alessandrini M.,Allen J., Bai W., Cimen S., Claes P., Cowan P., D'hooge J., Duchateau N., Ehrhardt J., Frangi A.F., Gooya A., Grau V., Lekadir K., Lu A., Mukhopadhyay A., **Oksuz I.**, Parajuli N., Pennec X., Pereanez M., Pinto C., Piras P., Rohe M., Rueckert D., Saring D., Sermesant M., Siddiqi K., Tabassian M., Teresi L., Tsaftaris S.A., Wilms m., Young A.A., Zhang X.,Gracia P.M., 'Statistical shape modeling of the left ventricle: myocardial infarct classification challenge", IEEE Journal of Biomedical and Health Informatics (JBHI) 2017 , to appear. **PDF**

[5] **Oksuz, I.**, Dharmakumar, R., Tsaftaris, S.A., "Fully automated myocardial segmentation of cardiac BOLD MRI", International Society of Magnetic Resonance in Medicine Meeting (SCMR), 2017. **PDF**

[6] Onofrey, J.A., **Oksuz, I.**, Sarkar, S., Venkatamaran, R., Staib, L.H., Papademetris X., "MRI-TRUS Image Synthesis with Application to Image-Guided Prostate Intervention", Simulation and Synthesis in Medical Imaging (SASHIMI), 2016. **PDF**

[7] **Oksuz, I.**, Dharmakumar, R., Tsaftaris, S.A., "Multi-Resolution Registration and Segmentation for cardiac BOLD MRI", International Society of Magnetic Resonance in Medicine Meeting (ISMRM), 2016.**PDF**

[8] **Oksuz, I.**, Bevilacqua, M., Mukhopadhyay, A., Dharmakumar, R., Tsaftaris, S.A., "BOLD contrast: A challenge for cardiac image analy-

sis", Society for Cardiovascular Magnetic Resonance Annual Meeting (SCMR), 2016. **PDF**

[9] **Oksuz, I.**, Dharmakumar, R., Tsaftaris, S.A., "Towards joint segmentation and registration of the myocardium in CP-BOLD MRI at rest", Society for Cardiovascular Magnetic Resonance Annual Meeting (SCMR), 2016. **PDF**

[10] Mukhopadhyay, A., **Oksuz, I.**, Tsaftaris, S.A., "Supervised Learning of Functional Maps for Infarct Classification", Statistical Atlases and Computational Models of the Heart (STACOM), 2015. **PDF**

[11] **Oksuz, I.**, Mukhopadhyay, A., Bevilacqua, M., Dharmakumar, R., Tsaftaris, S.A., "Dictionary Learning Based Image Descriptor for Myocardial Registration of CP-BOLD MR", Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI), 2015. **PDF**

[12] Mukhopadhyay, A., **Oksuz, I.**, Bevilacqua, M., Dharmakumar, R., Tsaftaris, S.A., "Unsupervised myocardial segmentation for cardiac MRI", Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI), 2015. **PDF**

[13] **Oksuz, I.***, Mukhopadhyay, A.*, Bevilacqua, M., Dharmakumar, R., Tsaftaris, S.A., "Data-Driven Feature Learning for Myocardial Segmentation of CP-BOLD MR", Functional Imaging and Modeling of Heart (FIMH), 2015. **PDF**

[14] **Oksuz, I.**, Mukhopadhyay, A., Bevilacqua, M., Yang H.J., Dharmakumar, R., Tsaftaris S.A., "Effect of BOLD Contrast on Myocardial Registration", International Society of Magnetic Resonance in Medicine Meeting (ISMRM), 2015. **PDF**

[15] Mukhopadhyay, A., Bevilacqua, M., **Oksuz, I.**, Dharmakumar, R., Tsaftaris, S.A., "Data Driven Feature Learning For Representation of Myocardial BOLD MR Images", International Society of Magnetic Resonance in Medicine Meeting (ISMRM), 2015. **PDF**

[16] Bevilacqua, M., Mukhopadhyay, A., **Oksuz, I.**, Rusu C., Dharmakumar, R., Tsaftaris S.A., "Dictionary-based Support Vector Machines for Unsupervised Ischemia Detection at Rest with CP-BOLD Cardiac MRI", International Society of Magnetic Resonance in Medicine Meeting (ISMRM), 2015. **PDF**

[17] Rudyanto, R.D., Kerkstra, S., van Rikxoort, E.M., Fetita, C., Brillet, P., Lefevre, C., Xue, W.,Zhu, X.,Liang, J., **Oksuz, I.**, Unay, D., Kadipasaoglu, K.,Estépar, R., Ross, J.C., Washko, G. R.,Prieto, J., Hoyos, Marcela H., Orkisz, M., Meine, H., Hüllebrand, M., Stöcker, C.,Mir, F.,Naranjo, V., Villanueva, E., Staring, M., Xiao, C., Stoel, B.C., Fabijanska, A., Smistad, E., Elster, Anne C., Lindseth, F., Foruzan, A., Kiros, R.,Popuri, K.,Cobzas, D., Jimenez-Carretero, D., Santos, A., Ledesma-Carbayo, M.J., Helmberger, M., Urschler, M., Pienn, M., Bosboom, D.G.H., Campo, A., Prokop, M.,de Jong, P.A.,Ortiz-de-Solorzano, C., Muñoz-Barrutia, A., van Ginneken, B., "Comparing algorithms for automated vessel segmentation in computed tomography scans of the lung: the VESSEL12 study", Medical Image Analysis (MedIA), 18 (7). pp. 1217-1232, 2014. **PDF**

[18] Kirisli, H.A., Schaap, M., Metz, C.T., Dharampal, A.S., Meijboom, W.B.,Papadopoulou, S. L.,Dedic, A., Nieman, K., de Graaf, M.A., Meijs, M.F.L.,Cramer, M.J.,Broersen, A., Cetin, S., Eslami, A.,Flórez-Valencia, L., Lor, K.L., Matuszewski, B.,Melki, I., Mohr, B., **Oksuz, I.**, Shahzad, R., Wang, C., Kitslaar, P.H., Unal, G., Katouzian, A., Orkisz, M.,Chen, C.M., Precioso, F., Najman, L., Masood, S., Unay, D., van Vliet, L.,Moreno, R.,Goldenberg, R., Vucini, E., Krestin, G.P., Niessen, W.J., van Walsum, T., "Standardized Evaluation Framework for Evaluating Coronary Artery Stenosis Detection, Stenosis Quantification and Lumen Segmentation Algorithms in Computed Tomography Angiography", Medical Image Analysis (MedIA), 17 (8). pp. 859-876, 2013. **PDF**

[19] **Oksuz, I.**, Unay, D., Kadipasaoglu, K., "Region Growing on Frangi Vesselness Values in 3-D CTA Data", Proceedings of the 21st Signal Processing and Communications Applications Conference (SIU). IEEE, pp. 1-4., 2013. **PDF**

[20] Unay, D., Harmankaya, I., **Oksuz, I.**, Kadipasaoglu, K., Cubuk, R., Celik, L., "Automated aortic supravalvular sinus detection in conventional computed tomography image", In: Proceedings of the 21st Signal Processing and Communications Applications Conference (SIU). IEEE, pp. 1-4, 2013. **PDF**

[21] **Oksuz, I.**, Unay, D., Kadıpasaoglu, K., "A Hybrid Method for Coronary Artery Stenosis Detection and Quantification in CTA Images", Workshop on 3D Cardiovascular Imaging: A MICCAI Segmentation

Chalenge, Proc. 15th Int. Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI), Nice - France, 2012. **PDF**

[22] **Oksuz, I.**, Unay, D., Kadipasaoglu, K., "Multi-scale Hessian Based Approach of Lung Vessel Tree in 3-D CTA Data" : A ISBI Segmentation Challenge, Proc. of International Symposium on Biomedical Imaging: From Nano to Macro (ISBI), 2012. **PDF**

[23] **Oksuz, I.**, Unay, D., Kadipasaoglu, K., "Segmentation of lung vessel tree in 3-D CTA data", Proc. (MASFOR), Istanbul-Turkey, 2012. **PDF**

Abstract

Joint registration and segmentation of varying contrast images is a fundamental task in the field of image analysis, despite yet open. In this thesis, novel techniques for the tasks of segmentation and registration are discussed separately and jointly. Cardiac Phase-resolved Blood Oxygen-Level-Dependent (CP-BOLD) MRI is a new contrast agent- and stress-free imaging technique for the assessment of myocardial ischemia at rest. However, it introduces varying contrast in medical image analysis applications. Therefore, establishing voxel to voxel correspondences throughout the cardiac sequence, an inevitable component of statistical analysis of these images remains challenging. Furthermore, medical background and specific segmentation difficulties associated to these images are present. Alongside with the inconsistency in myocardial intensity patterns, the changes in myocardial shape due to the heart's motion lead to low registration performance for state-of-the-art methods.

The problem of low accuracy can be explained by the lack of distinguishable features in CP-BOLD and inappropriate metric definitions in current intensity-based registration and segmentation frameworks. In this thesis, sparse representations, which are defined by a discriminative dictionary learning approach, are used to improve myocardial segmentation and registration. Initially appearance information is combined with Gabor and HOG features in a dictionary learning framework to sparsely represent features in a low dimensional space. Moreover, the motion is incorporated as additional feature to establish an unsupervised segmentation framework. For registering the cardiac sequence a new similarity metric is proposed utilizing the sparse representations. Also a joint optimization scheme for dictionary learning based feature representations is proposed using the sparse coefficients and dictionary residuals. The superior performance of the dictionary-based descriptors are showcased with several experimental results.

Part I

Introduction and background

1 Introduction

Coronary Artery Disease (CAD) is number one cause of death in the world [16]. The most common form of CAD leads to coronary artery stenosis, which results in reduced blood oxygen supply to the myocardium (myocardial ischemia). Early detection of ischemia will help clinicians to intervene and re-establish the blood flow in the necessary regions of coronary arteries. Unfortunately, most methods for detecting ischemia are invasive, require ionizing radiation and contrast agent [48].

Cardiac Phase resolved Blood Oxygen Level Dependent Magnetic Resonance Imaging (CP-BOLD MRI) is a reliable, non-invasive, and repeatable imaging method for ischemia detection [27]. However, it relies heavily on post-processing for accurate analysis of the data and disease diagnosis. Appropriate techniques should be developed for the analysis of this modality [145]. Our major motivation is to fill this gap and open the path for statistical analysis of this data. The possible algorithms that could be used to solve the challenge of segmentation and registration jointly, are promising to overcome the major difficulties of CP-BOLD MRI. The development of such algorithms will not only be a breakthrough in CP-BOLD, but can also be used in varying contrast images such as natural images

with illumination issues and varying contrast images.

1.1 Objective and Challenge

Our primary goal is to come up with a rapid and robust algorithm to segment and register the myocardium in a cardiac sequence to achieve voxel to voxel correspondences. These correspondences will give birth to accurate time-series. The acquired time series can enable feature extraction and better statistical analysis of the data. This contribution has the potential to be a breakthrough in cardiac ischemia detection. Design of a rapid and accurate image analysis algorithm for this modality will enable this imaging technique enter clinical practice.

The theoretical evaluation of the current state of the art in cardiac MRI registration, segmentation and joint registration segmentation techniques is compulsory for developing an approach on this challenge. Firstly, with a theoretical study of the algorithms, we aim to understand possible developments in different contexts. The fundamental research focus is evaluating and studying the methods in details, so that we can apply the necessary changes in some parameters and regularization terms. It could be possible to develop methods suitable and applicable in the specific challenge of registering and segmenting the myocardium region. In this thesis, we summarize the methods that have been developed for both challenges using dictionary learning.

The main idea is to analyze images in order to detect and extract the features in cardiac studies. We begin with the methods we developed for segmentation. Then, we report algorithms for registration with a certain emphasis on similarity metrics. At the end of the thesis, multiple possible approaches that can add the registration to the segmentation task will be discussed. The main area of the application will be the cardiac sequences of BOLD MRI; which is a imaging technique to measure the the difference in magnetic properties between oxyhemoglobin and deoxyhemoglobin

in capillaries and venous vessels. The developed methods in this thesis will add as a benchmark for the varying contrast imaging.

1.2 Research Questions and Objectives

Based on the motivation of myocardial registration and segmentation of the myocardium, we specify the main research questions addressed by the thesis. In the following we enumerate these questions and describe organization of the material. We illustrate also the scientific and engineering contributions and publications [1] offered by the thesis by answering the research question as follows:

Research Question 1

Does BOLD effect challenge segmentation and registration?

There is a vast amount of literature for the tasks of registration and segmentation of standard cine MR. We test a selection of methods on CP-BOLD MRI, which are specifically powerful to address varying contrast issues and compare the results. This question is addressed in chapters 4, 5 and 6. The results are published in [76], [77], [81] and [82] .

Research Question 2

How to obtain features to successfully align and extract the myocardium region?

We rely on sparse models more specifically dictionary learning to detect features. We add texture information to appearance for learning meaningful features to represent the myocardium at multiple scales. This question is addressed in chapter 4 and the results are published in [76].

Research Question 3

How to segment the myocardium using the data-driven features?

Automatic myocardial segmentation is one of the essential tasks to automatize cardiac ischemia detection. We used the features defined on training subjects with myocardium and background dictionaries to segment the myocardium. This research question is

[1] To disseminate research work, parts of this thesis have been published by the author or are currently under consideration for publication.

investigated in chapter 4 and the corresponding results are reported also in [76].

Research Question 4

How to incorporate motion information for myocardial segmentation?

The motion of myocardium could be used as a powerful tool for automatic segmentation. CP-BOLD MRI has no automatic segmentation method in the current state of the art. We hypothesize that this is due to the lack of exploiting the unsupervised techniques in a sparse representation setting, which can be an effective tool for developing features that are invariant to temporal and inter-subject variabilities, yet unique and descriptive. In addition, we also argue that the temporal consistency assumption for myocardial segmentation of standard CINE MR is a special case of the more generalized spatio-temporal variability observed in CP-BOLD MR. We investigate this issue in chapter 5 and report the results in [77] and [82].

Research Question 5

How to produce a robust measure of similarity for registering cardiac phases?

Nonrigid image registration is an essential step in medical imaging, including automatic segmentation, motion tracking and morphometric analysis [111]. However, since most of the proposed algorithms rely on a (dis)similarity metric build based on the assumptions of consistent intensity and local shape, images with pathologies and locally varying intensity may not be accurately aligned. We hypothesize that it is due to the lack of appropriate features, which are invariant to particular type of appearance and shape deformation observed in CP-BOLD images. This question is addressed in chapters 4 and 7. We publised the results regarding this question in [81].

Research Question 6

How to optimize registration and segmentation cost functions jointly?

Registration and segmentation of anatomical structures are two well studied problems in medical imaging. Optimizing segmentation and registration jointly has been proven to improve results

for both challenges. Joint registration and segmentation aims to solve an optimization function, which enables better performance for both of the tasks. Sequential registration and segmentation uses the result of one process to guide the second task. Simultaneous registration and segmentation aims to solve two problems at the same time using Markov Random Fields (MRF) approach. In chapter 7 we investigate these methods based on dictionary learning residuals and sparse representations defined on pre-trained dictionaries. We have demonstrated the results regarding this publication in [83].

1.3 Scientific Contributions

There are multiple contributions of this work in the area of medical image analysis:

1. The experiments demonstrate that BOLD contrast significantly affects the accuracy of segmentation algorithms.
2. We experimentally validate the fact that BOLD contrast significantly affects the accuracy of registration algorithms.
3. A set of compact features are designed using Multi-Scale Discriminative Dictionary Learning (MSDDL), which can effectively represent the myocardium in CP-BOLD MR.
4. A dictionary learning-based image descriptor (DLID) in a registration framework is proposed.
5. We have employed a joint motion and sparse representation based technique, where the motion not only generates a rough estimate of the myocardium, but also guides the sparse representation stage to a smooth solution.
6. We developed a MRF based joint registration and segmentation scheme on dictionary residuals and sparse coefficients to extract segmentation masks and accurate deformations of the cine typr cardiac data.

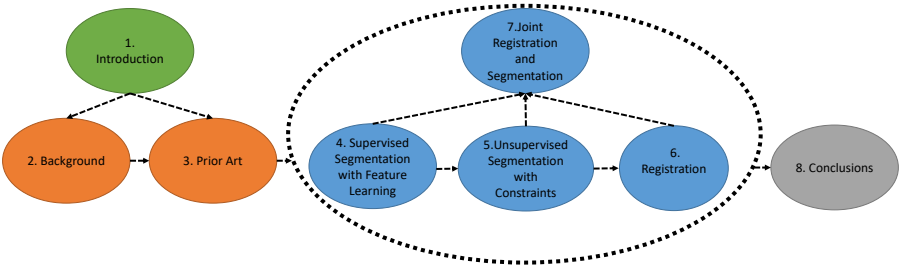

Figure 1.1: Thesis structure and dependencies among chapters.

1.4 Structure of the Thesis

This thesis is divided into eight chapters as illustrated in figure 1.1. The core registration and segmentation of the myocardium, i.e. chapters 4, 5, 6, and 7 as represented in the figure by a dotted circle. The arrows among chapters represent their dependencies. However, these dependencies do not indicate a compulsory reading and are only a guide to illustrate the relations among them. The details of the chapters as follows:

Chapter 2 briefly examines the background of coronary heart disease, CP-BOLD MRI and machine learning methods that are utilized in this thesis. The major focus is on available algorithms for the tasks of registration and segmentation CP-BOLD MRI. We give a detailed explanation of sparse learning techniques in this chapter and investigate K-SVD and OMP algorithms in particular.

In **Chapter 3**, we investigate the current state of the art of cardiac MR segmentation and registration. First, we scan the available methods that have been utilized for the last 20 years for the tasks of registration and segmentation separately. Then, we move to the methods that propose a joint segmentation and registration scheme. Finally, we summarize the current available techniques for CP-BOLD MRI and highlight possible paths the thesis will be built on.

Chapter 4 summarizes the data-driven feature learning tech-

nique (MSDDL) in particular, for the challenge of myocardial segmentation. A multi-scale discriminative dictionary learning approach is proposed for supervised learning and sparse representation of the myocardium, to improve the myocardial feature selection. This method combines appearance with Gabor and HOG features in a dictionary learning framework to sparsely represent features in a low dimensional space.

Chapter 5 investigates two methods for unsupervised segmentation thanks to incorporation of motion patterns as an additional information into the framework. The first technique (UMSS) is a fully unsupervised technique for segmenting myocardium from the background in both standard CINE MR and CP-BOLD MR. The appearance is combined with motion information in a dictionary learning framework to sparsely represent important features in a low dimensional space and separate myocardium from background accordingly. Our fully automated method learns background-only models and a one class classifier provides myocardial segmentation. The second method is a fully automated unsupervised 2D+time myocardial segmentation framework. We build a joint motion and appearance method that relies on dictionary learning to find a suitable representation subspace. The method is based on variational pre-processing and spatial regularization using Markov Random Field (MRF), to further accurately extract segmentations and time-series. Furthermore, a novel segmental analysis method attuned for BOLD time-series is utilized to demonstrate the effectiveness of the proposed method in preserving key BOLD patterns.

Chapter 6 presents the incorporation of data-driven features into a myocardial registration framework. The sparse representations, which are defined by a discriminative dictionary learning approach for source and target images, are used to improve myocardial registration. The differences of these distinctive sparse representations are used to define a similarity term in the registration framework.

Chapter 7 includes the methodologies and experiments of our joint registration and segmentation framework. Initially, we create a

database of dictionaries for myocardium and background, to come up with an initial segmentation, which relies on classification when projecting on discriminatory dictionaries of an input CP-BOLD subject stack. Then, the stack is registered in the cardiac cycle using the sparse coefficients of the projections to establish a new similarity metric. The obtained segmentation is refined till convergence via updated dictionaries. Finally, extensions to the framework of joint registration and segmentation are investigated. The main focus is to generate intensity and rotation invariant features and accelerate the calculation of sparse representations. A joint optimization scheme for both registration and segmentation is proven to aid the solutions of both challenges. A MRF-based joint optimization and segmentation scheme is proposed to accurately extract and register the myocardium region.

Finally, in **Chapter 8** we summarize the findings presented throughout thesis, and offer concluding remarks alongside with limitations and possible directions for future work.

2 Background

In this chapter we investigate the medical and technical background linked with the topics covered in this thesis.

Medical background starts with the fundamentals of cardiac anatomy and medical image acquisition techniques. Then, cardiovascular disease and its significance is pointed out. Afterwards, an overview about currently available MR acquisition techniques, alongside with MR basics, is investigated. Finally, the advantages and recent advancements of cardiac BOLD MRI are listed.

Technical background gives an overview of machine learning techniques utilized in this thesis. First machine learning classification methods that are used i n this thesis are introduced. Second, the available sparse coding techniques and dictionary learning techniques in literature are investigated. Finally, major algorithms of interest for this thesis, Orthogonal Matching Pursuit (OMP) and K-SVD are explained.

2.1 Medical Background

2.1.1 Cardiac Ischemia

Coronary Heart Disease (CHD), also known as coronary artery disease or ischemic heart disease, is the most common form of heart diseases [16]. CHD is a result of progressive build-ups of fat deposits within the walls of coronary arteries. These deposits form plaques, including fatty acids, cholesterol, calcium mineral,and fibrous connective tissues [109]. Such deposition are often present for decades prior to the demonstration of symptoms. During the time, the accumulated plaques causes narrowing of the lumen, leading to the limited or blocked supply of oxygen-rich blood to the myocardium, the heart muscle. This causes blood starvation to the cells of the myocardium and induces myocardium ischemia [126]. Figure 2.1 demonstrates the development of cardiovascular disease and ischemia in myocardium.

The muscle and cavity of the left ventricle can be divided into a variable number of segments. Based on autopsy data the American Health Association (AHA) recommends a division into 17 segments for the regional analysis of left ventricular function or myocardial perfusion [17]. In Figure 2.2, the seventeen AHA segments are illustrated, which are the state of the art guidelines to detect the location of the cardiac disease.

2.1.2 Basic MR Physics

The MR physics is heavily dependent on the spin of hydrogen nuclei, protons ($H+$), which creates magnetic moments due to the electric charge of protons. The water contained in human body allows to use hydrogen atoms for imaging. When these spinning magnetic moments, are placed in a strong external magnetic field B_0 they are not aligned any more. After this excitement the protons aim to come back to alignment to the magnetic field B_0 . Meanwhile they generate a net magnetic field M parallel to B_0. At Larmor

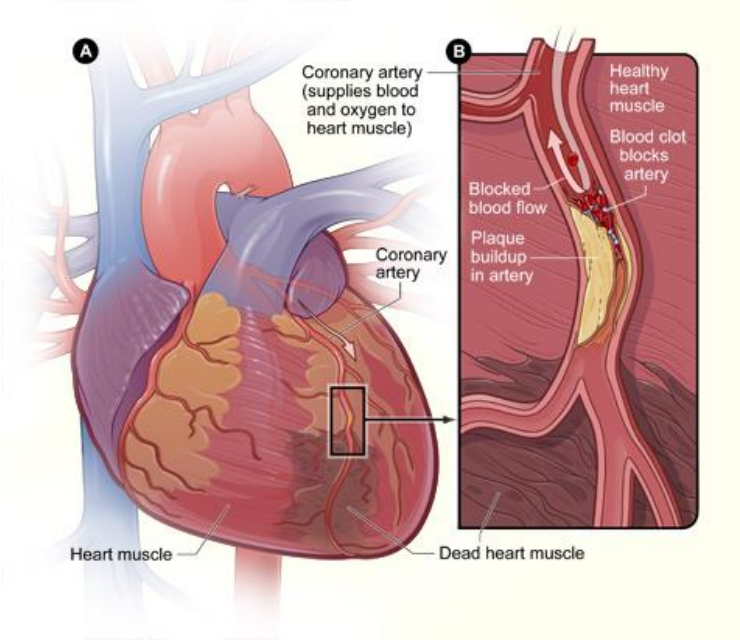

Figure 2.1: Coronary Heart Disease and Ischemia (adopted from [79]).

frequency ω_0, the nuclei precess around the direction of B_0. The precess is proportional to the strength of B_0 i.e. $B\omega_0 = \gamma_0$, where γ is a constant value for hydrogen. M is adjusted when applying a radio frequency (RF) pulse that is perpendicular to the main magnetic field γ. The aim is to generate resonance frequency [148]. To achieve this goal RF should be the same as that of the spins (Larmor frequency ω_0). When this RF is removed the relaxation starts, the spins start realigning to B_0, causing M to return parallel to B_0 again. During the relaxation, the spins release energy, emitting an RF signal. The conductive coil receives this signal and sends to a computer for image reconstruction [74].

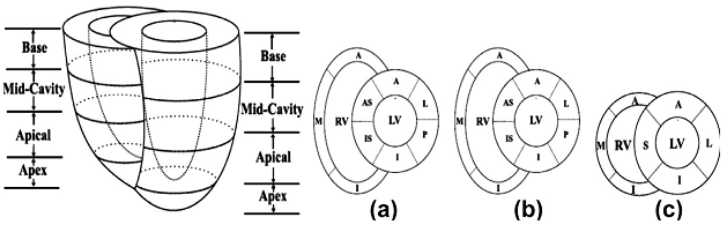

Figure 2.2: Seventeen AHA prescribed segments for the heart (a) basal SAX view, (b) mid-LV SAX view, (c) apical SAX view. (antero-septal (AS), anterior (A), lateral (L), posterior (P), inferior (I), and infero-septal (IS) (adopted from [17]).

K-Space

Recording and reconstruction of the MR signal is normally explained with k-space. The spatial position of protons are defined with the differences of Magnetic field gradients. A gradient (G) is applied to the main magnetic field, B_0, such that $B_0 = B_0 + G$. This gradient will cause the Larmor frequency of spins to vary according to the magnetic field strength B_0. The changed Larmor frequencies defines spatial positions of the spin encoded by gradient G [74]. The K-space is the fundamental space, where MR data is acquired. After establishing the K-space the real MR Image is reconstructed as visualized in Figure 2.3 .

2.1.3 Cardiac Perfusion MRI

The concept of injecting a tracer into blood and detecting its transit and distribution in the heart muscle for the assessment of myocardial perfusion is well established in Computer Tomography (CT) and Positron Emition Tomography (PET). Currently in the clinical practice injected contrast agents have been used to assess perfusion with cardiovascular magnetic resonance (CMR) [74]. The use of a gadolinium-based contrast agent for the assessment of myocardial perfusion with CMR has been extensively validated and success-

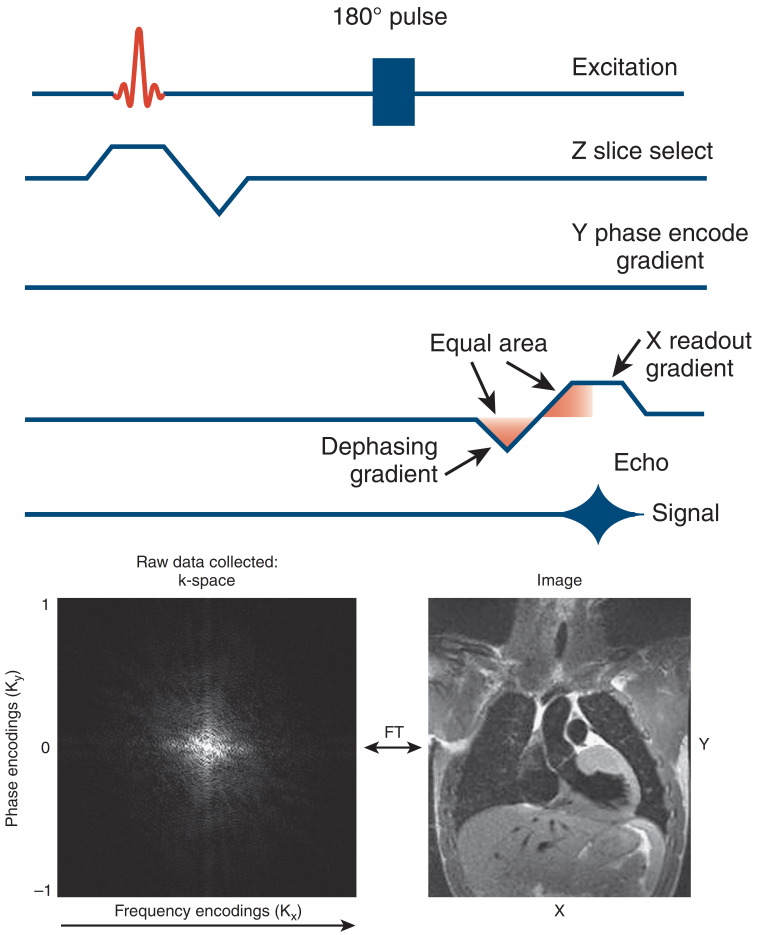

Figure 2.3: MRI acquisition and K-space construction adopted from [74].

fully applied in patient studies. The need for quantitative analysis of perfusion studies is also receiving increasing acceptance.

Under normal conditions, the blood flow resistance of the coronary circulation is determined vessels that are smaller than 300 mm in diameter. The adequate supply of oxygen and metabolites to the myocardium is tightly coupled to myocardial blood flow. Adequate and approximately constant blood flow is maintained through autoregulation and can compensate under resting conditions for up to 80% diameter coronary artery stenosis [74].

With more severe narrowing in a vessel the distal perfusion way is fully vasodilated, even under resting conditions, and no further augmentation of blood flow is feasible. In healthy subjects, myocardial blood flow can increase three to four times with maximal vasodilation. This means that differences in myocardial blood flow between a diseased region coronary artery and the territory of a normal coronary artery are amplified with maximal vasodilation. The techniques have matured to a point where they are applicable in clinical studies, despite the additional time required for quantitative analysis. There is already compelling evidence that CMR is superior to nuclear imaging for the assessment of myocardial perfusion [74].

First Pass Perfusion Imaging (FPP)

Myocardial perfusion imaging is based on measuring the delivery of a contrast agent to the myocardium during the first pass following a bolus injection. The signal intensity is enhanced by the contrast agent, which shortens the T1 relaxation time and results in a brighter signal using a T1-weighted imaging sequence. Regions with lower regional blood flow will appear hypointense and may be detected given adequate image quality. Quantitative measurement of blood flow may be made through analysis of the dynamics of the myocardial signal intensity measurement as a function of time. Myocardial flow reserve may be estimated by comparing the flow measurements acquired at rest and at stress. Stress perfusion is most commonly studied using vasodilation such as adenosine

or dipyridamole. Vasodilators increase the blood flow in normal vessels while stenotic vessels have a reduced vasodilator response. Regions with prior myocardial infarction may app ear hypointense despite normal blood flow, following revascularization, due to the low flow into scar tissue. Therefore, the interpretation of perfusion images usually also incorporates viability assessment by delayed enhancement imaging [29].

2.1.4 Cardiac BOLD MRI

Recently, there is a growing interest in endogenous contrast mechanisms for cardiovascular magnetic resonance. Blood is a magnetically inhomogeneous medium in which the magnetic susceptibility of red blood cells is strongly dependent on the blood oxygen saturation, defined as the percentage of hemoglobin that is oxygenated. Since the susceptibility of blood plasma is generally invariant, the cooperative binding of oxygen to the heme molecules in the red blood cells results in a detectable susceptibility difference between plasma and the red blood cells. This susceptibility variation gives rise to local magnetic field inhomogeneities, resulting in local frequency variations that lead to changes in $T2^*$ and apparent T2 of whole blood. The BOLD effect originates from the difference in magnetic properties between oxyhemoglobin and deoxyhemoglobin in capillaries and venous vessels. Despite its major application in brain imaging, the effect was also noted in the heart already more than fifteen years ago [50].

It has since been considered of great potential in detecting ischemic territories in the heart [27]. In contrast to first-pass bolus perfusion, BOLD imaging can be performed in the steady-state and hence the strict spatiotemporal limitations in resolution are significantly relaxed. To this end, BOLD imaging potentially allows for higher spatial resolution to better resolve transmural ischemic patterns during extended measurement times [145]. The disadvantage relative to contrast perfusion imaging is the limited dynamic range of the BOLD effect which is on the order of a few percent signal

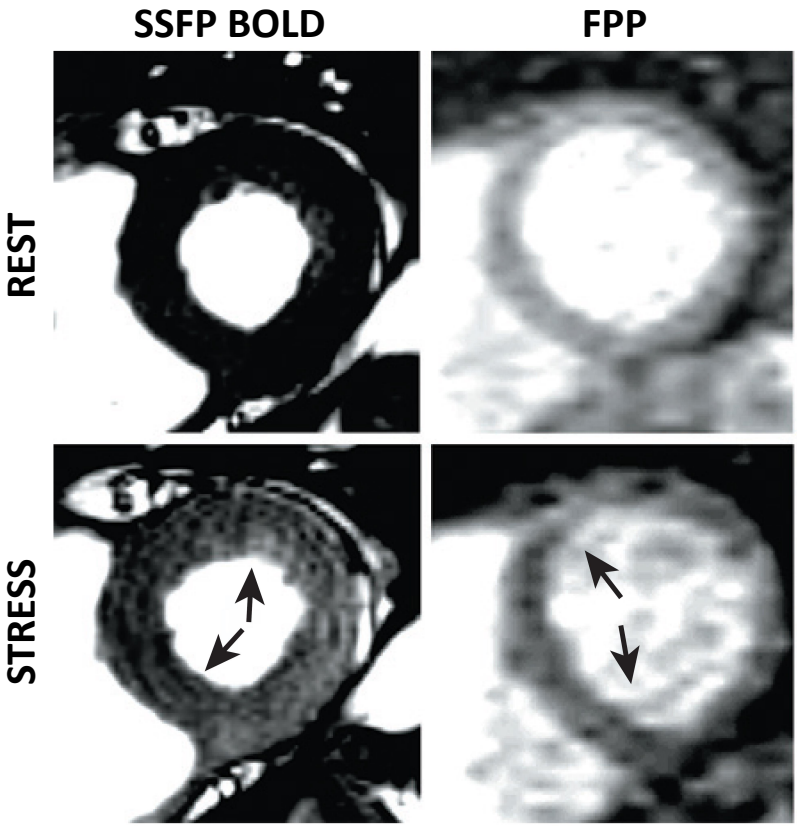

Figure 2.4: The Comparison of BOLD and FPP in a patient study [27].

change between oxygenated and deoxygenated areas. To quantitatively assess the BOLD effect in the heart and the gain in BOLD contrast-to-noise ratio at 3T relative to 1.5T careful considerations are necessary. There is still the need for advanced image processing of BOLD images for them to be used in clinical practice. A comparison of BOLD images of a patient alongside with the related FPP images under rest and stres conditions is illustrated in Figure 2.4. The arrows are showing ischemic region of the myocardium with serious Left Anterior Descending narrowing (LAD) [133].

2.2 Technical Background

In this section, an overview of the technical background and fundamentals of the utilized machine learning techniques are introduced. First part gives a general overview of machine learning techniques of significance for the remainder of the thesis. Second part investigates sparse coding problem in detail, which is an essential part of the thesis. Finally, we introduce dictionary learning techniques and discriminative dictionary learning. We will primarily focus on the utility of machine learning techniques on cardiac image segmentation and registration tasks.

2.2.1 Machine Learning

Machine learning approaches rely heavily on the data according to a defined model. Medical image analysis challenges can be modelled with successful models to automate the and processing. The algorithms in machine learning can be designated for multiple tasks such as clustering, classifications, regression, dimensionality reduction and density estimation. In this thesis, we will mainly focus on using machine learning for classification and clustering in the context of medical image analysis.

Machine learning algorithms can be classified into supervised, semi-supervised and unsupervised learning algorithms according

to the type of the input data. Throughout the thesis there are examples of supervised and unsupervised learning algorithms. Supervised learning refers to using labelled data for training of an algorithm. In Chapter 4 we use an algorithm that relies on manual annotations of the myocardium from training subjects to classify the myocardial region of the testing subject. Some examples of supervised learning algorithms are support vector machines (SVM), naive bayes classifier, random forests etc. Unsupervised learning are trained on unlabelled data and the goal is to discover the hidden structure of the unlabelled data with the proposed model. Some examples for unsupervised learning are k-means, principal component analysis (PCA), manifold learning, dictionary learning etc. In Chapter 5 we utilize an unsupervised learning of myocardium pixels with the aid of dictionary learning to classify the myocardium. Semi-supervised learning on the other hand works on the labelled and unlabelled data jointly in the cases of a scarcity of the labelled data.

More precisely in this work we utilize variations of dictionary learning and Support Vector Machines (SVMs). Dictionary learning is a strong way to learn sparse representations of the data in an unsupervised setting. These well established representations are well-suited input for a supervised classifier, which enables strong classifier definitions. We utilize this data-driven approaches both for image registration and image segmentation throughout the thesis. In the remainder of this chapter we will introduce Support Vector machines (SVMs), sparse coding techniques, and dictionary learning with a certain emphasis on the utilized techniques in the later chapters of the thesis.

SVM

In machine learning, support vector machines (SVMs, also support vector networks) are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis [24]. Given a set of training examples, each

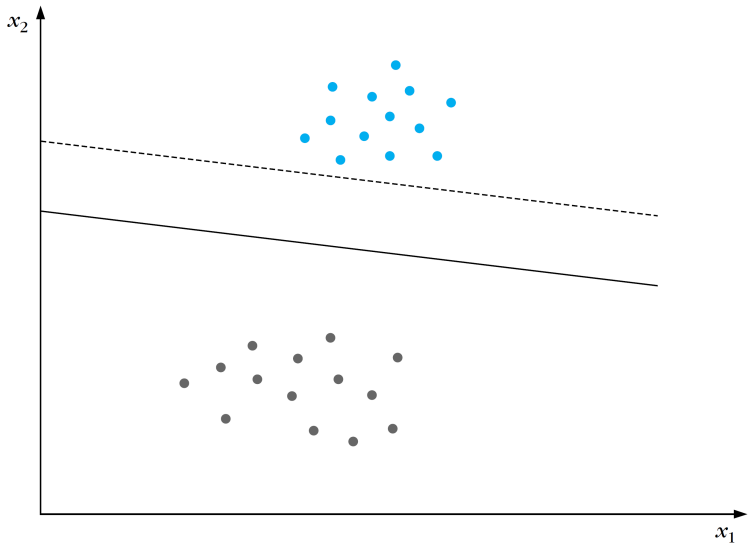

Figure 2.5: Support Vector Machines

marked for belonging to one of two categories, an SVM training algorithm builds a model that assigns new examples into one category or the other, making it a non-probabilistic binary linear classifier. An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that margin is as wide as possible as shown in Figure 2.5. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on.

Let us first take a look at the traditional two-class support vector machine. $\Omega = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$ defines a dataset with points $x_i \in R^d$ in a space, where x_i is the i-th input data point and $y_i \in \{-1, 1\}^d$ is the i-th output pattern, indicating the class membership [125].

SVMs can create a non-linear decision boundary by projecting the data through a non-linear function ϕ to a space with a higher

dimension. This means that data points which can not be separated by a straight line in their original space I are moved to a feature space F a hyperplane separates the data points of one class from an other. When that hyperplane would be projected back to the input space I, it would have the form of a non-linear curve [125].

The hyperplane is represented with the equation $w^T x + b = 0$, with $w \in F$ and $b \in \mathbb{R}$. The hyperplane that is constructed determines the margin between the classes; all the data points for the class 1 are on one side, and all the data points for class 1 on the other. The distance from the closest point from each class to the hyperplane is equal; thus the constructed hyperplane searches for the maximal margin ("separating power") between the classes. To prevent the SVM classifier from over-fitting with non linearly separable data, slack variables ξ_i are introduced to allow some data points to lie within the margin, and the constant $C > 0$ determines the trade-off between maximizing the margin and the number of training data points within that margin (and thus training errors). The objective function of the SVM classifier is the following minimization formulation:

$$\begin{gathered} \underset{w,b,\xi_i}{min} \frac{\|w\|^2}{2} + C \sum_{i=1}^{n} \xi_i \\ \text{subject to} \\ y_i(w^T \phi(x_i) + b) \geq 1 - \xi_i \quad \text{for all} \quad i = 1, \ldots, n \\ \xi_i > 0 \quad \text{for all} \quad i = 1, \ldots, n \end{gathered} \tag{2.1}$$

When this minimization problem (with quadratic programming) is solved using Lagrange multipliers, it gets really interesting. The decision function (classification) rule for a data point x then becomes:

$$f(x) = sgn(\sum_{i=1}^{n} \alpha_i y_i K(x, x_i) + b) \tag{2.2}$$

where α_i are the Lagrange multipliers; every $\alpha_i > 0$ is weighted

in the decision function and thus "supports" the machine; hence the name Support Vector Machine. Since SVMs are considered to be sparse, there will be relatively few Lagrange multipliers with a non-zero value.

One-class SVM

Traditionally, many classification problems try to solve the two or multi-class tasks. The goal of the machine learning application is to distinguish test data between a number of classes, using training data. One interesting application is a problem with only one defined class of data. An example of such problem is anomaly detection cases, where you have information about normal cases and try to classify the anomalies. A method for this task, which gained much popularity the last fifteen years, is the One-Class Support Vector Machine [107]. To cope with this problem, one-class classification problems (and solutions) are introduced. By just providing the normal training data, an algorithm creates a (representational) model of this data. If newly encountered data is too different, according to some measurement, from this model, it is labelled as out-of-class.

The Support Vector Method for novelty detection by Schölkopf et al. [107] basically separates all the data points from the origin and maximizes the distance from this hyperplane to the origin. This results in a binary function which captures regions in the input space where the probability density of the data lives. Thus the function returns $+1$ in a small region (capturing the training data points) and -1 elsewhere.

The quadratic programming minimization function is slightly different from the original stated above, but the similarity is still clear:

$$\min_{w,\xi_i} \frac{\|w\|^2}{2} + \frac{1}{\upsilon n}\sum_{i=1}^{n}\xi_i - \rho$$
$$\text{subject to} \tag{2.3}$$
$$y_i(w \cdot \phi(x_i) + b) \geq \rho - \xi_i \quad \text{for all} \quad i = 1, \ldots, n$$
$$\xi_i > 0 \quad \text{for all} \quad i = 1, \ldots, n$$

In the previous formulation the parameter C decided the smoothness. In this formula it is the parameter υ that characterizes the solution. This way it sets an upper bound on the fraction of outliers (training examples regarded out-of-class) and, it is a lower bound on the number of training examples used as Support Vector. Due to the importance of this parameter, this approach is often referred to as υ-SVM.

Again by using Lagrange techniques and using a kernel function for the dot-product calculations, the decision function becomes:

$$f(x) = sgn((w \cdot \phi(x_i)) - \rho) = sgn(\sum_{i=1}^{n} \alpha_i K(x, x_i) - \rho) \tag{2.4}$$

This method thus creates a hyperplane characterized by w and ρ which has maximal distance from the origin in feature space F and separates all the data points from the origin.

2.2.2 Sparse Coding

In its most general definition, the principle of sparsity, or parsimony, consists of representing some phenomenon with as few variables as possible [54]. In recent years, a large amount of multi-disciplinary research has been conducted on sparse models and their applications. In statistics and machine learning, the sparsity principle is used to perform model selection, which is automatically selecting a simple model among a large collection of them.

In signal processing, sparse coding interest in representing data with limited linear combinations of dictionary elements [72]. Subse-

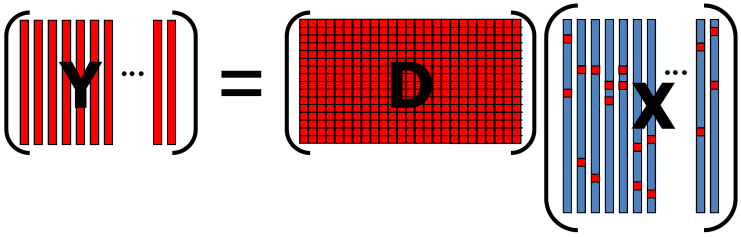

Figure 2.6: Sparse Coding Process. The signals in Y are represented by a linear combination of a given dictionary D with coding coefficient matrix X (adopted from [72]).

quently, the corresponding tools have been widely adopted by several scientific communities such as computer vision, neuroscience and bioinformatics. More specifically, we focus on applications where the dictionary is learned and adapted to data, yielding a compact representation that has been successful in various contexts. Sparse coding is a technique that is originally developed to explain the early visual processing in brain [85]. In a pioneer exploratory experiment, Olshausen et al. [85] demonstrated that dictionary learning could easily discover underlying structures in natural image patches.

Figure 2.6 illustrates an example of the sparse coding process. Given a set of basis vectors (atoms) $D \in \Re^{m \times K}$, a signal (a column of Y defined as y_i) can be represented as linear combination of atoms in D. D can be predefined or learnt from the data itself. Sparse representation of the signal can be represented as:

$$Y = DX \tag{2.5}$$

Here each column of $x_i \in \Re^{K \times 1}$ is the coding coefficient vector. There is no unique solution for X in sparse setting, since $K > m$ creates an under-determined system. This means the number of unknown parameters is larger than the number of equations. There are multiple optimization schemes to approach this under-

determined system imposing prior information over coefficients.

l_2 norm

One example regularization scheme adds the l_2 norm constraint on the coefficients:

$$\hat{x} = \underset{x}{min} \|x\|_2 \quad \text{subject to} \quad y = Dx \tag{2.6}$$

The equation 2.6 can be solved via the pseudoinverse of D, but this leads to a dense solution.

l_0 norm

In the applications of sparse coding a small number of significant coefficients are desired and a signal representation of linear combination of minimum number of atoms is sought. For this purpose, l_0 norm constraint is more suitable. The new formulation with l_0 norm is defined as:

$$\hat{x} = \underset{x}{min} \|x\|_0 \quad \text{subject to} \quad y = Dx \tag{2.7}$$

where the number of non-zero entries are bounded by l_0 norm.

OMP

There are multiple ways to solve the NP-hard problem of equation 2.7. One greedy approach that we refer to in this thesis is the Orthogonal Matching Pursuit (OMP) algorithm [120]. OMP, detailed in algorithm 1 constructs a sparse solution to a given signal by iteratively building up an approximation; the vector y is approximated as a linear combination of a few atoms of D, where the active set of atoms to be used is built atom by atom in a greedy fashion. At each iteration, a new atom that best correlates with the current residual is added to the active set. Each iteration stops until a pre-defined number of atoms in the active set is reached by the sparsity level.

Algorithm 1 OMP Algorithm

Require: Decomposition of Signals y_i
Input: Signals y_i $i = 1, \ldots, N$, Dictionary $D \in R^{m \times K}$, $\hat{y} = \emptyset$
Output: Decomposed signal $\hat{Y}$after pth iteration, Residual R^p

1: **Initialization** $R^0 = y$
2: **while** $i \leq p$ **do**
3: $l = \underset{l=1,\ldots,l}{argmax} |\langle g_l, R^{(i)} \rangle|$ Finding the atom in dictionary D with maximum correlation residual
4: $R^{(i+1)} = R(i) - x_l d_l^i$
5: $\hat{y} = \hat{y} + \langle R^{(i)}, d_l^{(i)} \rangle d_l^{(i)}$
6: $i = i + 1$
7: **end while**

Due to the simplicity of the implementation and enabling working with a number of pre-defined atoms makes OMP superior to other algorithms.

l_1 norm

One other approach is to replace the l_0 norm with the l_1 norm, when the equation is sparse enough.

$$\hat{x} = \min_x \|x\|_1 \quad \text{subject to} \quad y = Dx \tag{2.8}$$

or

$$\hat{x} = \min_x \|x\|_1 \quad \text{subject to} \quad \|y - Dx\|_2 < \varepsilon \tag{2.9}$$

equation 2.9 is a relaxed version of equation 2.8 with small difference between y and Dx. The constrained problems above need some relaxation method to be solved. One way to convert them into an unconstrained problem is using Lagrangian multipliers:

$$\hat{x} = \min_x \|y - Dx\|_2^2 + \lambda \|x\|_1 \tag{2.10}$$

where λ is a Lagrangian multiplier and controls the sparsity level. The first term is a reconstruction error term, which favors the low representation error and the second term adds a sparsity penalty to enforce sparseness. Least absolute shrinkage and selection operator (LASSO) is a common approach to solve this optimization problem [115].

2.2.3 Dictionary Learning

In the setting of sparse coding we assumed to know the (over-complete dictionary) D before hand. However, the dictionary can also be learnt directly from the data [101]. One classical way to do this is to choose a set of training signals or a basis of an over-complete wavelets, curvelets, Fourier transforms and etc. A more recent approach is to learn dictionaries based on training signals instead of predefined dictionaries. Given a set of training signals $Y = [y_1, y_2, \ldots, y_i, \ldots, y_n] \in \mathbb{R}^{m \times N}$, a dictionary D can be defined to represent each signal in Y sparsely:

$$\langle \hat{D}, \hat{X} \rangle = \underset{D,X}{argmin} \|Y - DX\|_2^2 \quad \text{subject to} \quad \|X\|_0 \leq L \qquad (2.11)$$

where L is the sparsity parameter and columns x_i of $X \in \mathbb{R}^{K \times N}$ represents sparse coding coefficients. Since the optimization of equation 2.11 is defined over D and X; this problem can be solved by fixing one parameter and applying the optimization on the other parameter. This optimization strategy starts with initializing the dictionary by using randomly selected training signals. Then, the sparse solutions X of the training signals Y are computed by keeping the dictionary D fixed. After that, the objective function in equation (2.9) can be optimized over D by keeping the sparse solutions X fixed. This alternating optimization process is repeated until some convergence criterion is reached, such as a number of iterations or a desired approximation error. It should be mentioned that finding the global optimal solution cannot be guaranteed by using this

iterative optimization strategy. The method of optimal directions (MOD) [30] and the K-SVD [2] are two efficient algorithms to learn dictionaries which utilize variants of this iterative optimization strategy. In practice, it has been observed that K-SVD converges with fewer iterations than MOD. In the next section, we will give a detailed introduction of the K-SVD algorithm.

K-SVD

The K-SVD algorithm is inspired from the k-means clustering algorithm, which is also an NP-hard problem [2]. The aim of k-means clustering is to partition all the signals into K clusters, in which each training signal belongs to the cluster with the nearest mean. It employs an iterative approach to find the solution of K clusters and there are two steps at each iteration: In the first step, each training signal is assigned to its nearest cluster; in the second step, the K clusters are updated as the centroids of their assigned training signals. The K-SVD follows a similar iterative two-step process to learn the dictionary and find the sparse solutions. After initializing the dictionary D, the solution of the sparse coefficients is found by keeping D fixed, followed by a second stage searching for a better dictionary. The K atoms in the dictionary are updated separately in the dictionary update stage. This is a direct generalization of the k-means algorithm, in which K clusters are also updated separately. The iterative process of the K-SVD is illustrated in Algorithm 2. The iterative process is repeated to update the K atoms of the dictionary D using the singular value decomposition (SVD) decomposition, thus the name K-SVD. In Algorithm 2, OMP is used for sparse coding. It should be mentioned that the K-SVD algorithm is flexible and can work with other sparse coding methods [2].

The major difference between the K-SVD algorithm and other dictionary learning (DL) methods is that the sparse coding coefficients X are not fixed in the dictionary update step. In the K-SVD algorithm, an atom in D and its corresponding row in X are updated simultaneously as shown in Algorithm 2. This accelerates

Algorithm 2 K-SVD Algorithm

Input:Training Signals $y_i \; i = 1, \ldots, N$

Output: An overcomplete dictionary $D \in \Re^{m \times K}$and sparse coding coefficients $X \in R^{K \times N}$

1: Initialize dictionary D with K randomly selected training signals
2: **while** converged **do**
3: **Sparse Coding:**
4: **for** each training signal y_i, use OMP to compute the corresponding coding coefficients x_i **do**
5: $\min_{x_i} \|x_i\|_0 \quad subjectto \quad y_i = Dx_i, i = 1, \ldots, N$
6: **end for**
7: **Dictionary Update:**
8: **for** $j = 1, \ldots, K$, update the jth atom d_j and the jth row x_T^j of the coding coefficients X **do**
9: Find the groups that use d_j:$w_j = \{i \in \{1, \ldots, N\} : x_T^j(i) \neq 0\}$ and x_R^j is obtained by eliminating the zero entries in x_T^j.
10: Compute representation error matrix: $E_j = Y - \sum_{i \neq j} d_i x_T^j$.
11: Obtain E_j^R by selecting the columns of E_j corresponding to w_j.
12: Apply SVD decomposition $E_j^R = U \Sigma V^T$, update the atom d_j with the first column of U, and update x_R^j with the first column of V multiplied by $\Sigma(1,1)$.
13: **end for**
14: **end while**

the convergence of the learning process, making the K-SVD algorithm more appealing. Despite the fact that the K-SVD algorithm converges fast, it is still computationally expensive at each iteration as a SVD decomposition must be calculated K times and all the N training signals are used for sparse coding at each iteration.

2.2.4 Discriminative Dictionary Learning

Dictionary learning aims to learn a (over-complete) dictionary in which only a few atoms can be linearly combined to well approximate a given signal. DL methods are originally designed to learn a dictionary which can faithfully represent signals, therefore they may not work well for classification tasks. The original goal of these works was not inference or classification per se, but rather representation and compression of signals, potentially using lower sampling rates than the Shannon-Nyquist bound [49]. Algorithm performance was therefore measured in terms of sparsity of the representation and fidelity to the original signals. Nevertheless, the sparsest representation is naturally discriminative: among all subsets of base vectors, it selects the subset which most compactly expresses the input signal and rejects all other possible but less compact representations. Researchers recently developed several approaches to learn a classification-oriented dictionary [134]. The central idea in sparse representation classifier (SRC) is to represent a test sample as a linear combination of samples from the available training set. Sparsity manifests because most of non-zeros correspond to bases whose memberships are the same as the test sample. Therefore, in the ideal case, each object is expected to lie in its own class subspace and all class sub-spaces are non-overlapping (as illustrated in Figure 2.7).

By exploring the label information, most DL-based classification methods learn such an adaptive dictionary mainly in two ways: either directly forcing the dictionary discriminative , or making the sparse coefficients discriminative (usually through simultaneously learning a classifier) to promote the discrimination power of the

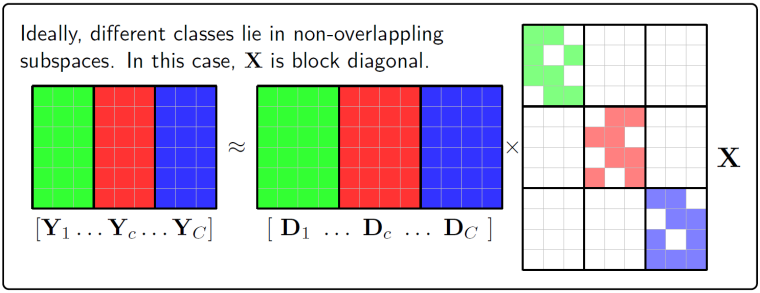

Figure 2.7: Ideal structure of coefficient matrix for sparse representation classifiers. Y_C represents signals from different classes, D_C class specific dictionaries and X_C stands for class-specific sparse representation (adopted from [128]).

dictionary. For the first case, [99] advocate learning class-specific sub-dictionaries for each class with a novel penalty term to make the sub-dictionaries incoherent. For the latter Zhuolin et al. [150] added a label consistence term on K-SVD to make the sparse coefficients more discriminative, thus the discrimination power of the dictionary is further improved. Yang et al. [141] propose Fisher discrimination DL method to simultaneously learn class-specific sub-dictionaries and to make the coefficients more discriminative through Fisher criterion. [128] proposed an extension of FDDL with capability to capturing shared features. Another method underlining the significance of shared features is COPAR algorithm [49] , which learns class-specific features and common pattern tools explicitly. Most of these algorithms are tested on face and scene recognition problems.

Recently, in the context of medical image analysis the utility of such techniques has been recognized from the community. Tong et al. [117] applied the principles of SRC on the problem of hippocapmus segmentation using image patches. A similar approach [118] with inclusion of graph cuts is implemented for the purpose of pancreas, kidney, spleen and liver. segmentation challenges. Benkarim

et al. [10] applied the idea of label-consistent K-SVD for the challenge Caudate and Accumbens segmentation on Brain MRI. Another area of interest is the histo-pathological image classification between healthy and diseased cases. In their work Vu et al. [127], [129] proposed a class-specific DL, which not only specifies the dictionary corresponding to the class identity of the sample, but also learns poorly representing samples from other classes.

3

Prior Work

Two of the most fundamental problems in analysis of cardiac MRI sequences are segmentation and registration. The registration and segmentation of cardiac MRI images are studied in the past 20 years. There is an immense amount of literature on this topic. The majority of the methods in the literature, performs registration and segmentation consecutively for image analysis. For establishing voxel correspondences of a varying contrast image cardiac sequence, registration algorithms are not sufficient; due to the inconsistent movement of different structures in the image. This deficit of registration algorithms can be neglected using the segmentation as a guidance of registration. In recent years there has been many algorithms proposed for solving the issues as a joint problem. This way, the propagation of errors from one to the other is avoided. In this chapter, a literature review of the proposed solutions to these

This chapter is partly based on:

- Oksuz, I., Mukhopadhyay, A., Bevilacqua, M., Yang H.J., Dharmakumar, R., Tsaftaris S.A.," Effect of BOLD Contrast on Myocardial Registration", International Society of Magnetic Resonance in Medicine Meeting (ISMRM), 2015.

challenges are presented separately and jointly. Besides, the possible implementations on cardiac BOLD MR images are investigated, with certain emphasise on specific attributes of this imaging modality. We have also added some additional related works paragraph for each chapter to set the table for the proposed methods during the thesis.

This chapter is structured as follows: in section 3.1 theoretical background of segmentation algorithms is illustrated; in section 3.2 registration algorithms are reviewed; in section 3.3 the joint registration and segmentation methods are discussed. In section 3.4 the current available techniques in BOLD MRI are investigated with a certain focus on the influence of BOLD contrast on the algorithms. Finally, in section 3.5 we summarize all the background and introduction to set the table for the main part of the thesis.

3.1 Segmentation of the Myocardium in Cardiac MRI

Segmentation attempts to reduce the variation in image appearance to a set of discrete labels. In cardiac MRI, these labels can correspond to multiple anatomical structures like myocardium, left ventricle (LV), right ventricle (RV), aorta etc. In this section, we will investigate the algorithms, which deals with this problem as an explicit issue. There are various reviews in literature for this task in [94, 95, 114] for cardiac MR segmentation. To the best of our knowledge, there is only one review [37], which investigates image processing techniques in cardiac MR perfusion.

Cardiac image segmentation presents a vast of challenges in the short axis view. Segmentation of the myocardium on these images consists in delineating the outer wall, also called epicardium and the inner wall, called endocardium (Figure 3.1). Each contour to be delineated presents specific segmentation difficulties.

The epicardial wall is at the frontier between the myocardium

and surrounding tissues, which have different intensity profiles and show poor contrast with the myocardium. Still segmentation difficulties exist, that mostly originate from gray level inhomogeneities in the blood flow, and particularly because of thepresence of papillary muscles and wall irregularities inside the heart chambers,which have the same intensity profile as the myocardium (Figure 3.2). They can thus prevent from clearly delineating this wall. According to clinical standards, they should not be taken into account for endocardial wall segmentation.Because the endocardial wall is less difficult to segment that the epicardial one, and since it is the only contour required to compute the ventricular volume, some works only focus on the endocardium segmentation [95].

In accordance with the motivation of the thesis, we are diving the segmentation algorithms for the myocardium proposed in the literature into three:

- Deformable models and level sets
- Atlas based models
- Neural Networks

However, the boundaries of the algorithms are not strict and there are many hybrid methods for the segmentation task. The hybrid methods are going to be investigated in each chapter in the thesis according to the relevance to the proposed algorithm.

The segmentation task is handled alongside other tasks in literature [95]. One of the main issues is the localization of the heart in MRI to ease the segmentation procedure. Lin [58] et al. proposed an algorithm to localize the heart and propagate an atlas based approach on top of it. Also Kurkure et al. [51] focused on the localization of the heart, with the help fuzzy affinity. However, their algorithm needs seed points for the initialization step. Furthermore, their approach is a 2D solution and has shortcomings in 3D approaches, due to the lack of assumptions of shape and appearance. Moreover, the blood signal intensity varies from base to apex with

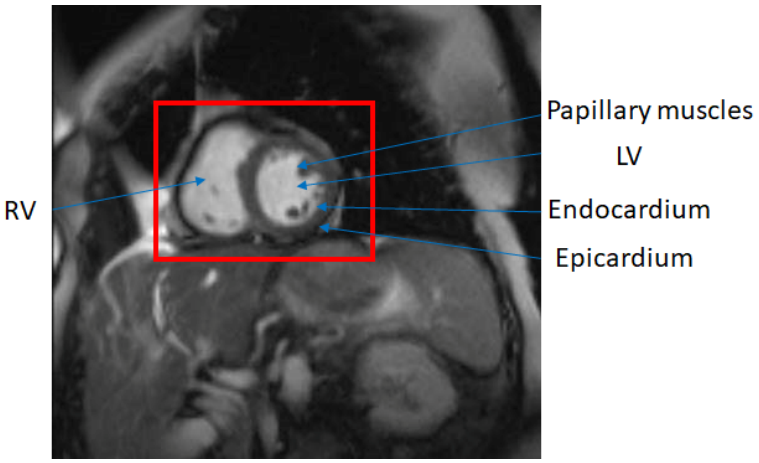

Figure 3.1: A full size short-axis cardiac MR image and a ROI identifying the heart (adopted from [95]).

no predictable pattern due to the variable coil intensity rendering, which creates additional problems for the 3-D segmentation with this method.

3.1.1 Deformable Methods

One popular algorithm in cardiac MRI segmentation has always been the active contour approaches. We investigate the active contours in two separate groups, namely: parametric active contours and geometric active contours. Deformable models, or active contours (also known as snakes) technique is a popular model-driven technique based on parametric curves, surfaces or volumes, that deform under internal and external forces. The external energy forces the contour to move toward the image data. The internal energy controls the contour based on a regularizing smoothness constraint [95]. Additional energy terms can constrain the deformable model to achieve better results. The general deformable model energy function can be written as:

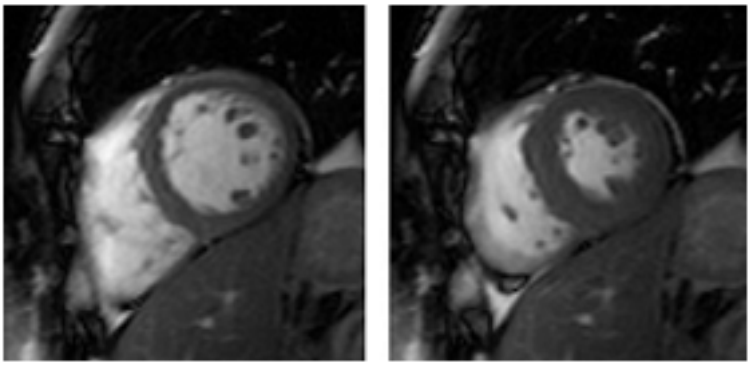

Figure 3.2: Cropped short-axis cardiac MR image of End-Diastole (ED) on left and End-Systole (ES) on right (adopted from [95]).

$$E = E_{ext} + Eint \tag{3.1}$$

External energy term is usually defined as:

$$E_{ext} = -\|\nabla|G_\sigma(x,y) * I(x,y)|\|^2 \tag{3.2}$$

where $G_\sigma(x,y)$ is a two dimensional gaussian function with standard deviation σ and $I(x,y)$ is the intensity value. Internal energy is defined as:

$$E_{int} = \frac{1}{2}(\alpha(s)\|Iv_s(s)\|l^2 + \beta(s)\|v_s s(s)\|^2) \tag{3.3}$$

where α and β are weights, the first order term defines the stretching and the second order term defined the curvature.

Since the edge map of the image can be misleading due to the noise and missing data, Gradient Vector Flow (GVF) was proposes a new external energy function to handle the edge function smoothly [140]. GVF is derived based on the minimization of the energy function proposed in equation 3.4 where v is the Gradient Vector Flow. μ is the smoothing weight, ∇I is the image gradient, and $\|\nabla I^2\|$ penalizes the edge information.

$$E_{\text{GVF}} = \int\int \mu(\|v\|^2) + \|\nabla I^2\| \cdot \|v - \nabla I\|^2 \tag{3.4}$$

There has been many publications relying on this equation for segmentation of LV. [56] proposed a GVF method with merging the global shape information with line shape information. Their approach is an early example of the technique and suffers from high user interaction. [136] proposes an approach based on selective smoothing direction gradient vector flow (SSDGVF) snake model. Their method incorporates shape prior to SSDGVF algorithm. Automatic localization of the cardiac endocardium contour, and elliptic shape constraint are presented to the equation. One of the recent works on GVF [25] combines GVF with morphological operations to segment LV.

Level sets represent an implicit function which deforms based on regional intensity or edge-based feature and is able to develop topological changes. The initial contour at time zero (C_0) corresponds to the zero level set of the function ϕ.

$$C_0 = (x, y)|\phi(x, y, 0) = 0 \tag{3.5}$$

Level set equation can be parametrized as:

$$\phi_t + F|\nabla\varphi| \tag{3.6}$$

where F represents speed function and depends on geometry and image gradient. Given an initial contour (or contours), an implicit function is defined and deformed at each pixel where the zero-level set determines the actual position of the curve(s) as a function of time. Level set approaches are stable, but are computationally costly. [1] uses a level set approach, where they minimize an energy function that contains area, length, and intensity variations inside and outside a contour such that:

$$E = \alpha L(\text{contour}) + \beta V(\text{inside of contour})$$
$$+\gamma \int_{\Omega} |I - c_0|^2 H(\phi) d\Omega + \rho \int_{\Omega} |I - c_1|^2 (1 - H(\phi)) d\Omega$$

where α,β,γ,ρ are the weighting parameters, I is the intensity, c_0 and c_1 represent the average intensity inside and outside the curve. $H(\phi)$ is the so called Heaviside function defined as follows:

$$H(\phi) = \begin{cases} 0, & u \geq 0 \\ 1, & u < 0 \end{cases}$$

Lecellier et al. [52] presents an inclusive overview of region-based active contours for segmentation with a focus on cardiac MRI. Deformable models provides certain advantages such as parameterization independence, ease of implementation, ease of initialization from a simple 2D curve and automatic handling of topological changes. However, it suffers from existence of noise, low contrast and objects complexity in medical images. The solution to these weaknesses is to incorporate a prior knowledge for more plausible results. Chen et al. [21] in their work displacement coded MRI, optimized an algorithm with level set. They use elliptic constraints to limit the evolution of the curve to myocardium region. In [47] shape based LV segmentation in first pass perfusion images used level sets with probabilistic shape constraints to estimate myocardium region. This method is a computational costly approach for multiple slices.

Deformable models have the ability to match particular shapes by changing the parameters of energy functions. These methods are able generate smoothed curves without the need for training data. Additionally, it is possible to integrate shape prior constraints in developed techniques. However, these techniques suffer in occluded objects and noisy images. Moreover, they are heavily dependent on parameters and initialization.

Incompressibility Constraint

The myocardium is believed to be nearly incompressible during systole and diastole due to the high water content [143]. Myocardial volume changes have been quantified during systole and diastole. It is believed that the myocardial volume is relatively constant during a cardiac cycle, varying about $3.5 - 5\%$. The conservation of volume of the myocardium has been utilized as an additional term in [146]. The small compressibility of the heart is usually attributed to the compressible blood vessel lumens and the difference in the total volume of blood in the myocardium depending on the phase of the cardiac cycle. This constraint on 3D images incorporates the speckle statistics and myocardial volume incompressibility using probability models. The compressibility ρ is statistically modelled as a Gaussian $(G(V_0, r))$ such that:

$$\rho(\phi_{in}, \phi_{out}) = \frac{1}{\sqrt{2\phi\sigma}} exp(\frac{(V - V_0^2}{2\sigma^2}) \tag{3.8}$$

here the volume of space between the endocardium(ϕ_{in}) and epicardium (ϕ_{out}) is the myocardial volume and $\sigma_0 = \frac{1}{120} V_0$. This simulation allowed about 5% variation in the myocardial volume. In their work [146], they combine the speckle statistical model and the incompressibility model in a level set framework in order to segment the cardiac endocardium and epicardium from ultrasound images.

3.1.2 Atlas Based Segmentation

An atlas describes the different structures present in a given type of image. It can be generated by manually segmenting an image or by integrating information from multiple segmented images from different individuals. Given an atlas, an image can be segmented by mapping its coordinate space to that of the atlas, using a registration process. Widely applied to brain segmentation, this technique has also been used for cardiac segmentation. As shown in Figure 3.3,

the principle is to register the labelled atlas onto the image to be segmented, and then apply the obtained transformation to the atlas to obtain the final segmentation. Segmentation can thus easily be propagated throughout the cardiac cycle using the same principle. In the literature, the construction of an anatomical heart atlas is based either on a single segmented image [64], an average segmentation result obtained over a population of healthy volunteers [65] or a cadaver atlas [147]. The atlas can be matched on a new individual using non-rigid registration (NRR), a transformation that accounts for elastic deformations, which will be further discussed in the following chapters. NRR consists in maximizing a similarity measure between a source image S (the atlas) and a target or reference image X (the unsegmented image). Since the atlas and the MR image can have non corresponding gray levels, the similarity criterion must only account for statistical dependencies between them. The most widely used criterion for NRR is the normalized mutual information (NMI) measure. Based on individual and joint gray level distributions, NMI is defined as:

$$E - NMI(S, X) = \frac{H(S) + H(R)}{H(S, R)} \tag{3.9}$$

Atlas-based segmentation techniques require a wide range of training data and a good registration method tailored for the images. Recently; patch-based multi-atlas approaches paired with label fusion [8] have become state of the art for cardiac segmentation. Atlas-based models offer a powerful approach; a collection of atlases is used to represent the population. These approaches have been used successfully for segmentation, e.g. multi-atlas segmentation of the brain [40] and heart [8], but are limited by the small number of atlases used. This means that these approaches only perform well if the images are similar to those in the atlas database. While approaches for atlas selection have been proposed, these require very large databases of labelled images, which is infeasible in practice and non-generalisable.

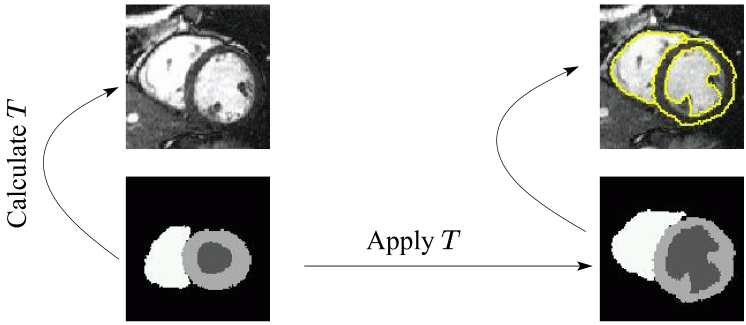

Figure 3.3: Anatomical atlas based segmentation principle: (i) computation of the transformation T between the atlas and the image and (ii) deformation of the atlas by T (adopted from [63])

3.1.3 Neural Networks

Given a large labelled data for training, multi-layer neural networks have been shown are effective automatic feature extractors for high-dimensional datasets [53]. Convolutional neural networks (CNN) are a neural network variant modelled to take advantage of data with regular structure, such as the spatial grid matrix of images [43].

Neural Networks have been utilized extensively in recent years for the task of medical image analysis as detailed in a review [61]. Ronneberger et al. [100] has pioneered an increasing interest for utilizing neural networks in the context of medical image analysis with their approach of U-nets. In the context of cardiac MR segmentation, Avendi et al. [7] combined CNNs and a deformable model to perform LV segmentation, though only for the endocardial wall and only in the ED and ES phases. Three separate networks were trained; consisting of a two-layer CNN for the initial localization of heart region of interest, and two separate three-layer fully-connected networks for LV segmentation at basal and mid-ventricular slices, and for apical slices, respectively. Tran [119] used a 15-layer CNN to perform complete left and right ventricular segmentation. Most

recently, Tan et al. [113] proposed to use radial distances of the LV walls from the center of the blood pool rather than per-pixel binary image masks, and that the radial distances to be computed via neural network regression as opposed to per-pixel classification using two sets on neural networks. Most of this work relied on adjacency and boundary conditions and neglected the underlying semantic information. A more recent work [84] addresses this issue by incorporating global shape information into neural networks. In general we can identify, that supervised methods require lots of data for training and a robust feature generation and matching framework. While deep learning can help incorporate the latter, it does need significant training data.

3.1.4 Evaluation Metrics for Segmentation

To evaluate performance of segmentation we used two metrics in this thesis, which are classically used when evaluating segmentation [98]. We used the Dice overlap measure, which is defined between two regions A and B as:

$$D(A,B) = \frac{2\|A \cap B\|}{\|A\| \cup \|B\|}.$$

To evaluate the match of the ground truth annotation to an algorithm's result in terms of distance, we relied on the Hausdorff distance between two contours C^A and C^B :

$$HD(C^A, C^B) = \max\{\underset{a \in C^A,}{max}\ \underset{b \in C^B,}{min}\ d(a,b), \underset{b \in C^B}{min}\ \underset{a \in C^A}{max}\ d(a,b)\}$$

where d presents the distance of points $a \in C^A$ and $b \in C^B$.

3.2 Registration of Cardiac MRI

Image registration is the problem of finding a coordinate transformation that spatially aligns two or more images. It is a common

necessity in applications of medical imaging. One of the images, called the moving image, is deformed to the other image, the fixed image. The quality of alignment is defined by a cost function C, which measures the similarity of the fixed image and the deformed moving image. A high similarity leads to a low cost function and vice versa [111]. An example of a cost function is the mean squared difference of voxel intensities. The coordinate transformation that relates the fixed and moving image is estimated by iteratively minimizing the cost function. Registration definition in cardiac MRI can lead to different application fields:

- *Dynamic sequence image registration:* Images of the same subject, which involves dynamic stacking static images. These images are acquired at different time steps from images sequence, which are used to quantify and capture the anatomy of the heart.
- *Multi-temporal image analysis:* Images of the same subject are acquired at different times and under different physical conditions. Registration of these images will enable us to monitor the changes in cardiac anatomy.
- *Multi-subject image analysis:* Images of different subjects are registered for deformation-based morphology (e.g. inter-subject registration).
- *Construction and use of atlases:* Images acquired from different sites and different times are aligned for the purpose of segmentation and statistical analysis of anatomical shapes.

There are a vast review papers investigating the multiple aspects of registration algorithms. Sotiras et al. [111] is a great resource for the definitions and concepts in medical image registration. More specifically in cardiac context, [73] gives a generic survey about image registration in cardiac MRI. On the other hand, [131] is focusing on the field of cardiac motion recovery for the survey. [60] compares

the validation techniques for the image registration algorithms, which is still an open topic in research.

Strongly bounded with purpose of the specific approach, algorithms depending on different information in images can be proposed for registration. Current registration methods can be divided into two types, namely pixel intensity-based and feature-based. In cardiac MRI literature both methods have been utilized in different settings. The general intensity based registration algorithms can be summarized as an energy minimization shown in the following equation:

$$\hat{\phi} = \underset{\phi}{argmin}[\underbrace{E_D(I_S \circ \phi, I_T)}_{1} + \underbrace{R_\phi}_{2}] \tag{3.10}$$

where $\hat{\phi}$ is the sought transformation in between two images. ϕ is a transformation, which can be assumed element of Hilbert space or manifold. This transformation has to be parametrized for discrete representation (e.g. B-Spline Free Form Deformations). Term 1 in the equation is a difference measure between the target image I_T and the warped source image $I_S \circ \phi$ (e.g. Sum of squared differences(SSD), Mutual Information(MI)). The term 2 is the regularization term, which models the behaviour of the model, incorporates prior knowledge and constrains the problem (e.g. bending energy).

3.2.1 Non-rigid Registration Using Basis Functions

This class of techniques attempts to extract the non-rigid motion of anatomical objects using a set of basis function such as splines that have inherent smoothness properties. The smooth nature of the B-spline basis functions lead to more congruent results. Rueckert et al. [103] used a two-step registration technique: 1. Global registration using affine registration, 2. Local registration using a spline based similarity matching. The similarity function can be different measures such as NMI and SSD. The advantage of B-spline based Free Form Deformation is that it offers the capability to be used as a

multi-modal algorithm to provide dense and pixel-wise results.The basic idea of B-spline FFD is to transform an object by manipulating an underlying spline-based mesh of control points Φ. The resulting transformation defines the shape of the 3D object. The energy term that leads the motion of the control points usually consists of the similarity function and a spatial velocity smoothness constraint but can be extended to any functional.

$$T(x,y,z) = \sum_{l=0}^{3}\sum_{m=0}^{3}\sum_{n=0}^{3} B_l(u)B_m(v)B_n(v)\phi_{i+l,j+m,k+n} \qquad (3.11)$$

where T is a transformation and n_x; $n_y$and n_z define the number of control points in x, y and z directions. B_l is the lth basis function for the uniform non-rational case is defined below for order up to cubic.

The advantages of the basis function based techniques (such as spline) are their inherent smoothness and their ability to reduction of the computation to control points. Additionally no training is needed and the same framework can be extended for other applications. The disadvantages of such a framework are the dependence of the results on the nature of the basis functions, number of control points, as well as their position. The optimization technique may not lead to the best results if the control points do not properly cover the complex portions of the shape, e.g., with intrusions and protrusions.

One interesting method proposed to estimate motion and strain from 3D echocardiography is [26]. The method enforces time consistency by representing the 4D velocity field as the sum of continuous spatiotemporal B-Spline kernels. They used SSD as image similarity metric and the regularization term is based on the incompressibility of myocardial tissue. Alessandrini [5] et al. proposed to use monogenic signal theory in the context of of pyramidal B-spline registration. Their algorithm relies on the conversation of monogenic phase over time and proven to adapt well to intensity distortions.

3.2.2 Feature-based Registration

Appearance information may lead to a local minima in the objective function and ambiguous matching when pixels of same anatomical structure take different values of intensity [111]. One way to address this issue is to increase the dimensionality of the feature space. A way to augment the feature space is by introducing local information through the use of attributes that represent the geometric structure of the underlying anatomy. These approaches are referred to as feature-based registration algorithms. Recently, Heinrich et al. [39] borrowed an idea from image denoising literature to propose a new descriptor for image registration. The features are defined as similarities between neighboring patches. The similarity metric is constructed as a vector-difference between the set of patches in two images. Local information may also be incorporated by exploiting the local frequency representations obtained as response to Gabor filters [87]. Gabor features have proven successful for cardiac image registration as they are able to capture information across different scales and orientations [88]. Ou et al. [88] used the idea of mutual saliency to detect landmarks from the consistent points for registration. This technique relies on a set of optimized Gabor features. In chapters 6 and 7 we apply feature-based for the task of CP-BOLD MR registration and define a new feature-based similarity criteria in chapter 6.

3.2.3 Registration Evaluation

Image registration is important for many applications, including longitudinal evaluations in individuals, comparison between individuals, creation of population atlases, use of atlas-linked information in individual cases, delivery of precision therapies, and many others. Non-rigid image registration is a more general approach than the widely used affine and rigid methods, but requires more complex methodology and computational effort to implement. Evaluating the performance of non-rigid image registration algorithms is a

difficult task since point-wise correspondence from one image to another is not unique. There is rarely a ground truth correspondence map to judge the performance of a registration algorithm.

Intensity variance metric is a common method used to measure image registration performance is to register a population of images with a target image and average the intensities of the registered images. The idea is that the better the registration algorithm is, the closer each registered image looks to the target image and the sharper the intensity average image. One way to measure the sharpness of the intensity average image is to compute the variance of the registered intensity images. The voxel-wise intensity variance (IV) of a population of M images registered to image j is computed as:

The inverse consistency metric measures the inverse consistency error between a forward and reverse transformation between two images. Ideally the forward transformation equals the inverse of the reverse transformation implying a consistent definition of correspondence between two images, i.e., correspondence defined by the forward transformation should be the same as that defined by the reverse transformations. Thus, composing the forward and reverse transformations together produces the identity map when there is no inverse consistency error. The inverse consistency error is defined as the squared difference between the composition of the forward and reverse transformations and the identity mapping.

The transitivity metric evaluates how well all the pairwise registrations of the image population satisfy the transitivity property. The transitivity property is important to minimize correspondence errors when two transformations are composed together. Ideally, transformations that define correspondence between three images should project a point from image A to B to C to A back to the original position. The transitivity error for a set of transformations is defined as the squared error difference between the composition of the transformations between three images and the identity map.

Many NIR methods have been developed, but are especially dif-

ficult to evaluate, since pint-wise image correspondence is usually unknown. There is no "Gold Standard" to evaluate the performance. In an effort to ganerate a benchmark myocardial tracking and deformation algorithms Tobon-Gomez et al. [116] utilized landmark points for cardiac MR and ultrasound images. This paper was a result of a MICCAI conference challenge on a publicly available data and showcased the performance of different deformation models. Even though there are many metrics to evaluate the registration performance, the choice of the metric usually challenge specific.

3.3 Joint Registration and Segmentation of Cardiac MRI

The challenge in merging the registration and segmentation is to guarantee convergence for preventing the inaccurate estimates of either registration or segmentation without affecting each other. The total available information in image data sets could be exploited in an efficient fashion, since estimates of labels are manipulated with registration criteria and partially registered algorithms will be governed by combined class model.

Markov Random Fields (MRF) is one of the most popular approaches to merge image registration and segmentation. Due to its weak dependency on manual annotations, some recent studies [33] have focused on registration-based segmentation, where an atlas image is registered to a target image. The resulting transformation is then applied to the labeled atlas, which yields a segmentation of the target image. Whereas various registration algorithms are applicable for this task, state-of-the-art performance has been achieved using MRFs on a control grid hierarchy. While being widely used in medical imaging, image registration alone cannot solve the segmentation task by itself, as it is considered to be an ill-posed problem. This is because anatomical correspondences, which are not guaranteed to exist, are computed using surrogate criteria such as intensity similarity [33]. In alternating registration and segmen-

tation (ARS), the estimated solution for one sub-problem is used as prior knowledge in the other sub-problem, whereas simultaneous registration and segmentation (SRS) aims at optimizing both goals at the same time. Most of the existing literature in the field have followed the ARS approach. The method proposed in [137] alternates between estimating a rigid transformation using Powells' method and updating the segmentation using iterative conditional modes (ICM) in an MRF. [138] proposed an ARS approach based on MRFs. Their method alternates between solving one MRF that optimizes registration parameters and updates segmentation probabilities, and a second MRF solving the segmentation. There have been also studies on alternating registration and segmentation in variational formulations, which also rely on iteratively updating the registration and segmentation solutions, for example, by using gradient descent. For example, a method which alternates between updating the deformation field using a quasi-Newton optimization method and evolving a contour in the target image while constrain the distance between deformed atlas and said target contour was proposed in [28]. It is also possible to combine registration with model-based segmentation as shown in [66], where point-based registration and active-shape segmentation trained from manually annotated images are employed for mutual benefit.

There have been successful applications of joint registration and segmentation on brain MRI Literature especially on tumor segmentation and growth modelling. . Gooya et al. [35] proposes registration of an atlas, deformed using a complex tumor growth model, to the patient's space. The parameters are learned using the Expectation Maximization algorithm in a sucesfull way to pose a joint optimization problem. Parisot [91] et al. proposed to combine the modeling of both problems where the unknown variables correspond to a two layer graphical model, one that represents the 3D deformation field and another that refers to the 3D binary segmentation map. This graphical model is superimposed to the volume domain. The two layers are interconnected with a combined cost

that relaxes the registration in the presence of tumors, while at the same time performs a segmentation-by-registration using the segmentation costs as criterion. Linear programming and duality are used to determine the optimal solution of the combined problem.

3.4 Cardiac BOLD MRI

Cardiac BOLD MRI is a current state-of-the-art approach, capable of simultaneously capturing BOLD changes and wall motion in a single sequence. This technique relies on endogenous contrast to relate the 1D time series to myocardial oxygenation. The most significant advantage of this modality is the ability to identify ischemic territories at rest, without the need of a stress agent [123]. In order to analyse the CP-BOLD data deeper the myocardium region in the sequence should be accurately segmented and registered in between the time series.

CP-BOLD identifies the ischemic myocardium by examining changes in myocardial signal intensity patterns as a function of cardiac phase [123]. However, visualizing and quantifying such changes requires significant post-processing, including myocardial segmentation to isolate the myocardium from the rest of the anatomy. In particular, although CP-BOLD is a cine type acquisition, automated myocardial segmentation and registration algorithms developed for standard CINE under-perform, due to the spatio-temporal intensity variations of the myocardial BOLD effect [106], an example of which is shown in Figure 3.4. Thus, in addition to violating shape invariance (as with standard CINE MR), the principal assumption of appearance invariance (consistent intensity [95]) is violated in CP-BOLD MR as well.

3.4.1 Segmentation

Fully supervised myocardial segmentation (i.e., separating myocardium from the rest of the anatomy) developed for standard CINE

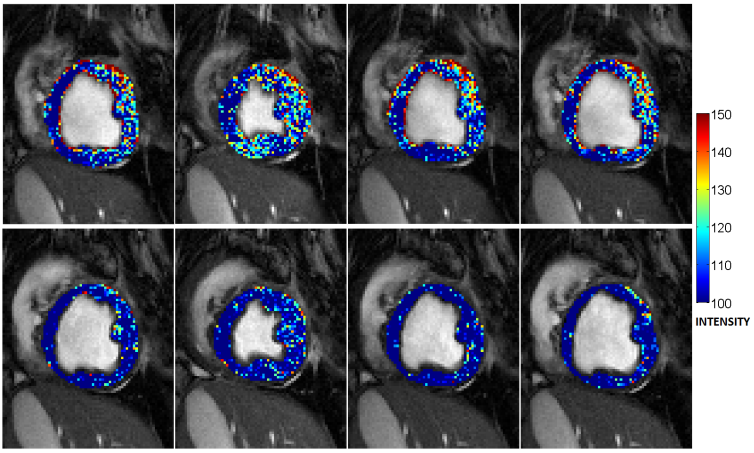

Figure 3.4: CP-BOLD MRI sequence has more intensity variation compared to standard Cine MRI. Exemplary cardiac phases of CP-BOLD MR (top row) and standard CINE MR (bottom row) obtained from the same subject under baseline conditions (absence of ischemia). The myocardium is color coded and overlaid on the grey-level image. We highlight the challenge of appearance variation in CP-BOLD MR which is minimal in the case of standard CINE MR.

MR, however, under-perform in the case of CP-BOLD MR due to the spatio-temporal intensity variations of the myocardial BOLD effect [106,123]. In [121], a dynamical programming based myocardial contour tracking algorithm is proposed. This algorithm relies on fitting and propagating a mask conforming to the six segment model of the LV utilizing the elastic matching information. However, this method is only limited to epicardial segmentation and does not offer the necessary voxel correspondences for the cardiac sequence. In chapter 4, we hypothesize that this is due to the lack of exploiting the unsupervised techniques in a sparse representation setting, which can be an effective tool for developing features that are invariant to temporal and inter-subject variabilities, yet unique and descriptive. In addition, we also argue that the temporal consistency assumption for myocardial segmentation of standard CINE MR is a special case of the more generalized spatio-temporal variability observed in CP-BOLD MR. Consequently, developing generalized features for CP-BOLD MR should also address the problems of myocardial segmentation of standard CINE MR; which is illustrated in chapters 4 and 5.

CP-BOLD provides both BOLD contrast and information of myocardial function [123]in a single acquisition that can be seen together as a movie (i.e., similar to Standard CINE MRI acquisition). Either at stress [122] or at rest (i.e., without any contraindicated provocative stress) [12, 123], it is the spatio-temporal signal intensity patterns due to the BOLD effect which are modulated by the disease that enable the diagnosis. Currently, CP-BOLD myocardial segmentation requires tedious manual annotation. Despite advancements in myocardial segmentation in Standard CINE MRI (which is similar to CP-BOLD but with little or no BOLD contrast), discussed at length in chapter 5, most methods when used on CP-BOLD MR images for the same task, produce unsatisfactory results. Figure 3.5A illustrates this by overlaying ground truth and algorithmic results for several state-of-the-art methods showing significant segmentation errors. These errors have deleterious effects on BOLD

signals, as Figure 3.5B shows. Instead of the expected behaviour across the cardiac cycle [123] which is seen when ground truth manual segmentations are used, significant deviations due to over- and under-segmentation are observed.

3.4.2 Registration

As Figure 3.6 illustrates, time series of intensity vary as a function of cardiac phase when BOLD effect is present –it appears maximal in systole and minimal in diastole. In disease this effect is not present. Such spatio-temporal intensity variations of the myocardial BOLD effect cause the methods developed for standard CINE MR registration to under-perform.

To showcase the influence of BOLD contrast on the registration we considered an experimental set-up with a state-of-the-art registration algorithm [5]. Our set-up tests the registration algorithm in a minimal motion to higlight the performance. We have considered the registration of the first frame (fixed) and the last frame (moving) in the acquisition, which typically reflect minimal motion (diastole), and can thus evaluate more accurately the effects of BOLD contrast and the presence of disease. To provide a visual example of how BOLD contrast and ischemia affect the registration of a single algorithm [5], in registering frames in this set-up, the displacement vectors of those registrations are shown color-coded in Figure 3.7. Under baseline conditions, where cardiac motion should be minimal between diastole, the algorithm finds small and consistent displacements throughout the myocardium; however, when BOLD contrast is present (as in CP-BOLD) greater variability is observed. Under ischemia greater regional variability is observed both in standard Cine and CP-BOLD, consistent with the expected changes in cardiac contractility in the presence of ischemia (from LAD stenosis).

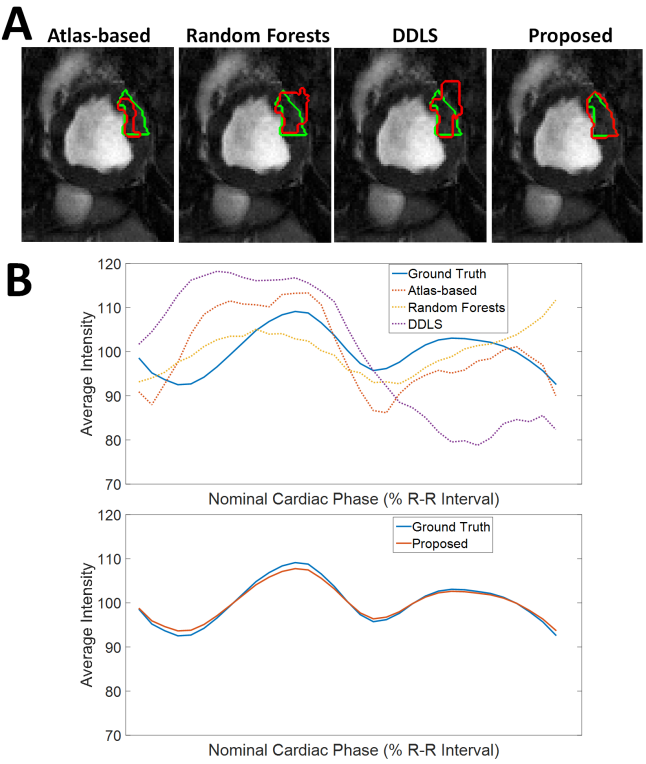

Figure 3.5: BOLD contrast challenges myocardial segmentation algorithms. **A:** Results of various algorithms (shown in red) for myocardial segmentation of the anterior region overlaid to ground truth (green) manual delineations. Algorithms used: atlas-based [124], Random Forests on Appearance and Texture features (a baseline) and a state-of-the-art DL method (DDLS) [117]. **B:** Corresponding time series of the Anterior region from different methods compared to the one obtained based on ground truth segmentation. Overall errors in segmentation lead to deviations in the estimated time series, which will ultimately lead to low accuracy in ischemia detection. Our proposed method in chapter 5 achieves high segmentation accuracy (last image in A); which leads to a better estimate of the time series (bottom part of B). [In typical CP-BOLD acquisition settings, with ECG-triggering, first and last points in the R-R interval correspond to diastole and systole tends to appear around 30%.]

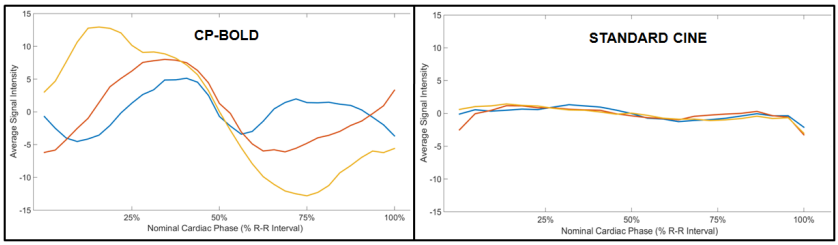

Figure 3.6: Exemplary plots of time series extracted from the same corresponding regions in the same subject under baseline (absence of disease) conditions using CP-BOLD MR and standard Cine. Observe how in CP-BOLD, intensity varies with cardiac phase, but in standard CINE MR this variation is minimal.

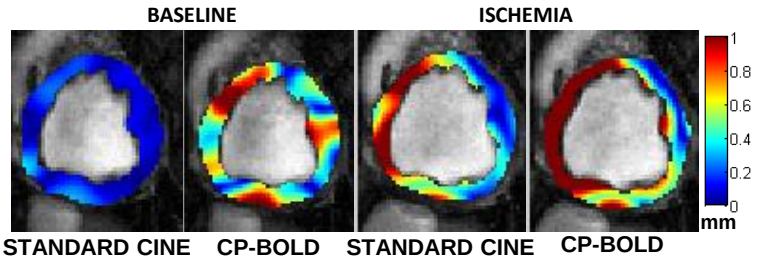

Figure 3.7: Color-coded pixel-wise magnitudes of displacement vectors [5] in mm overlaid on the original images within the myocardium.

3.4.3 Ischemia Detection

There is still a great need of image analysis and feature extraction algorithms for computer aided diagnosis of ischmeic pixels in CP-BOLD. Most importantly, it has been shown that CP-BOLD can identify ischemic territories at rest, without the need for provocative stress [123], contraindicated for some of the population. Thus, one scan captures physiological and functional details of the heart in a non-invasive and repeatable manner with great image quality. In [106], a dictionary learning method for generating synthetic data is proposed. The method utilizes the circulant nature of cardiac BOLD time-series. This work adds additional significance for the accurate registration and segmentation cardiac BOLD MRI. Moreover, a more recent work by Bevilacqua et al. [12] has showed that varying intensity patterns can guide the process of ischemia detection, thanks to an anomaly detection algorithm. The proposed algorithm suffers from low quality of image segmentation and registration; which causes a drop of accuracy in estimated timeseries and consequently ischemia detection.

3.5 Summary

In this chapter, we have investigated the vast literature of cardiac image registration and segmentation algorithms. We also have underlined the need for accurate segmentation and registration algorithms in the context of myocardial registration and segmentation. An accurate segmentation and registration scheme will output accurate timeseries, which will enable pixel-level determination ischemia likelihood in CP-BOLD MRI. This is a crucial task for successful clinical translation of CP-BOLD MRI. In the next part of the thesis, we will investigate the performance of the state of the art cardiac image analysis tools on our challenge and present our novel solutions for the tasks of segmentation and registration. In Chapter 4, we will highlight the drawbacks of the literature for

the myocardial segmentation and investigate a novel sparse feature learning techniques for myocardial segmentation. In Chapter 5 the possibility of using motion as distinctive feature for myocardium is considered. In Chapter 6, the utility of data-driven features are explored in a novel registration setting. In Chapter 7 the ways to merge two optimization problems are studied. In each chapter we added a related work section to highlight the corresponding literature to each method we developed.

Part II

Data-Driven Feature Learning for Segmentation and Registration

4

Supervised Segmentation with Feature Learning

There is a need for definition of appropriate features for the task of CP-BOLD MR myocardial segmentation. There is no automated CP-BOLD MR segmentation algorithm currently, and semi-automated methods based on tracking are currently employed [121]. We hypothesize that it is due to the lack of appropriate features, which are invariant yet unique and descriptive under the particular type of appearance and shape deformation observed in CP-BOLD images. Rather than relying on low-level features used often for myocardial segmentation of standard CINE MR which are inconsistent for CP-BOLD MR, a more generalized feature learning method should be developed to accommodate the myocardial BOLD effect while still being reliable in the CINE MR case.

This chapter is based on:

- I. Oksuz, A. Mukhopadhyay, M. Bevilacqua, R. Dharmakumar S. A. Tsaftaris, "Data-driven feature learning for myocardial segmentation of CP-BOLD MRI", *Functional Imaging and Modeling of the Heart (FIMH)*, pp. 189–197, 2015.

A patch-based discriminative dictionary learning technique (which has been used also in echocardiography [42]) is adapted to learn features from previously segmented data in a fully supervised manner. The motivation behind the choice of a sparse dictionary is to employ a compact and high-fidelity low-dimensional subspace representation which is able to extract semantic information of the myocardium as well [135]. The key observation behind this strategy is that, though the patch intensity level varies significantly across the cardiac cycle, sparse representations based on learnt dictionaries are invariant across the cardiac cycle, as well as unique and robust. The discriminative dictionary learning strategy is designed to facilitate this key observation regarding CP-BOLD. Briefly described, during training two separate dictionaries are learnt at multiple scales for the myocardium and background. In this regard, also a discriminative initialization step (discarding patches with high values in intra-class Gram matrix) is introduced to promote diversity in initialization, and a discriminative pruning step (discarding training patches with high values in inter-class Gram matrix) to further boost the discriminative abilities of the dictionaries. During testing, multiscale sparse features are used.

The main contributions of this chapter are twofold. First, the experiments demonstrate that BOLD contrast significantly affects the accuracy of segmentation algorithms (including segmentation via registration of an atlas, level sets, supervised classifier-based and other dictionary-based methods) which instead perform well in standard CINE MR. Second, to address the hypothesis a set of compact features are designed using Multi-Scale Discriminative Dictionary Learning, which can effectively represent the myocardium in CP-BOLD MR. The method has been evaluated on canine subjects, which makes the problem even more challenging (lower accuracy is expected) due to the smaller size of myocardium.

4.1 Background

Recently, Atlas-based segmentation techniques have received significant attention. The myocardial segmentation masks available from other subject(s) are generally propagated to unseen data in Atlas-based techniques [8] using non-rigid registration algorithms, e.g., diffeomorphic demons (dDemons) [124], FFD-MI [34] or probabilistic label fusion [8]. Level set class of techniques uses a non-parametric way for segmenting myocardium with weak prior knowledge [18, 55].

Segmentation-only class of techniques mainly focuses on feature-based representation of the myocardium. Texture information is generally considered as an effective feature representation of the myocardium for standard CINE MR images [144]. The patch-based static discriminative dictionary learning technique (DDLS) [117] and Multi-scale Appearance Dictionary Learning technique [42] have achieved high accuracy and are considered as state-of-the-art mechanisms for supervised learning of discernible myocardial features from previously segmented data.

In this chapter, the segmentation-only approach with the major feature of considering multi-scale appearance and texture information is followed as the input of a discriminative dictionary learning procedure. Our experimental results suggest a significant improvement over the existing segmentation only and segmentation with registration techniques in the case of CP-BOLD MR images, whereas state-of-the art performance in the case of standard CINE MR.

4.2 Multi-Scale Discriminative Dictionary Learning (MSDDL)

General image segmentation strategies are developed on the assumption that both appearance and shape do not vary considerably across the images of a given sequence. Cardiac motion affects the shape invariance assumption, and varying CP-BOLD signal inten-

sities violate the appearance invariance assumption as well. To overcome this issue, dictionary learning techniques can be leveraged to learn better representative features. To this end, we propose a Multi-Scale Discriminative Dictionary Learning (MSDDL) method (detailed in Algorithm 3). The features learnt via dictionary learning are tested in a rudimentary classification scheme solely for the purpose of comparing to other methods.

4.2.1 Feature Generation with MSDDL

Given some sequences of training images and corresponding ground truth labels (i.e. masks), we can obtain two sets of matrices, $\{Y_k^B\}_{k=1}^K$ and $\{Y_k^M\}_{k=1}^K$, where the matrix Y_k^B contains the background information at a particular scale k (each scale is characterized by a different patch size), and Y_k^M is the corresponding matrix referring to the myocardium. Information is collected from image patches: squared patches are sampled around each pixel of the training images. More precisely, the i-th column of the matrix Y_k^B (and similarly for the matrix Y_k^M) is obtained by concatenating the normalized patch vector of pixel intensities at scale k, taken around the i-th pixel in the background, along with the Gabor and HOG features of the same patch. Our MSDDL method takes as input these two sets of training matrices, to learn, at each scale k, two dictionaries, D_k^B and D_k^M, and two sparse feature matrices, X_k^B and X_k^M. E.g. , the i-th column of the matrix X_k^B, $x_{k,i}^B$, is considered as the discriminative feature vector for the particular pixel corresponding to the i-th column in Y_j^B. Dictionaries and sparse features are trained via the well known K-SVD algorithm [2]. One main modification to K-SVD is the use of the "intra-class Gram matrix" to promote diversity in the initialization step. The Gram matrix is used the literature [150] to ensure atoms of the dictionaries to be incoherent. Our idea here is to have a subset of patches as much diverse as possible to train dictionaries and sparse features. For a given class considered (let us say background) and a given scale k, we can define the intra-class Gram matrix as $G_k^B = (Y_k^B)^T Y_k^B$. To ensure a proper discriminative

initialization, patches that correspond to high values in the Gram matrix are discarded from the training before performing K-SVD. Notably, we sort the training patches w.r.t. the sum of their related coefficients in the Gram Matrix, and we prune them by choosing a certain percentage.

A second proposed modification relates to a pruning step, which is performed after K-SVD to remove undesired elements from each dictionary trained. In this case, at each scale k, an "inter-class Gram matrix" is computed ($G_k^{BM} = (D_k^B)^T D_k^M$): the atoms of each dictionary are sorted according to their cumulative coefficients in G^{BM}, and a chosen percentage of them is discarded to ensure mutual exclusiveness between the dictionaries of the two different classes. The philosophy behind this operation is similar to the one of the discriminative dictionary learning algorithm proposed in [99], where the norm of the inter-class Gram matrix appears in the optimization formulation as a constraint to be minimized. By pruning the undesired dictionary atoms all at one time, we actually adopt a greedier and low-complexity approach to the same problem. Moreover, we believe that, instead of globally minimizing the Gram matrix norm, directly removing the most "problematic" patches, which create ambiguity between background and myocardium, is more effective in our case.

4.2.2 Building a Rudimentary Classifier for Segmentation

When considering the same patch-based approach in a segmentation problem, we have a set of test matrices $\{\hat{Y}_k\}_{k=1}^K$, obtained by sampling patches at multiple scales from the test image, and concatenating intensity values of these patches, along with Gabor and HOG features. The goal is to assign to each pixel of the test image a label, i.e. establish if the pixel is included in the background or the myocardial region. To perform this classification, we use the multi-scale dictionaries, $\{D_k^B\}_{k=1}^K$ and $\{D_k^M\}_{k=1}^K$, previously learnt with MSDDL. The Orthogonal Matching Pursuit (OMP) algorithm [120] is used to compute, at each scale k, the two sparse feature matri-

Algorithm 3 Multi-Scale Discriminative Dictionary Learning (MS-DDL)

Require: Multi-scale training patches for background and the myocardium:
$\{Y_k^B\}_{k=1}^K$ and $\{Y_k^M\}_{k=1}^K$
Ensure: Multi-scale dictionaries for background and the myocardium:
$\{D_k^B\}_{k=1}^K$ and $\{D_k^M\}_{k=1}^K$

for $k = 1...K$ **do**
  for C={B,M} **do**
    Evaluate Y_k^C
    Compute the intra-class Gram matrix G_k^C
    Discard atoms with high values in G_k^C
    Learn dictionary and sparse feature matrix with the K-SVD algorithm

$$\underset{D_k^C, X_k^C}{\text{minimize}} \|Y_k^C - D_k^C X_k^C\|_2^2 \quad \text{s. t.} \quad \|x_{k,i}^C\|_0 \leq L$$

  end for
  Compute the inter-class Gram matrix G_k^{BM}
  Discard from D_k^B and D_k^M atoms with high values in G_k^{BM}
end for

ces $\hat{X}_k^B$ and $\hat{X}_k^M$. A certain patch at scale k, $\hat{y}_{k,i}$ will be assigned to the class that gives the smallest dictionary approximation error. More precisely, if $\|\hat{y}_{k,i} - D_k^B \hat{y}_{k,i}^B\|_2$ is larger than $\|\hat{y}_{k,i} - D_k^M \hat{y}_{k,i}^M\|_2$, at scale k the patch is assigned to the background; otherwise, it is considered belonging to the myocardial region. In this study, we employed a simple majority voting across all scales to obtain the final classification for each pixel of the test image.

4.3 Results

This section offers a qualitative and quantitative assessment of our proposed method w.r.t. state-of-the-art methods, to demonstrate its effectiveness for myocardial segmentation. It is particularly important to note that our method significantly outperforms all methods from current literature in both baseline and ischemia cases of CP-BOLD MR, whereas yields state-of-the-art results for both baseline and ischemia cases of standard CINE MR.

4.3.1 Data Preparation and Parameter Settings

2D short-axis images of the whole cardiac cycle were acquired at baseline and severe ischemia (inflicted as stenosis of the left-anterior descending coronary artery (LAD)) on a 1.5T Espree (Siemens Healthcare) in the same 10 canines along mid ventricle using both standard CINE and a flow and motion compensated CP-BOLD acquisition within few minutes of each other. All quantitative experiments are performed in a strict leave-one-subject-out cross-validation setting.

As for the parameters of MSDDL, in this framework we have empirically chosen a dictionary of 1000 atoms for foreground and background respectively, a sparsity of 4, a number of scales $K = 3$, and $9 \times 9, 11 \times 11$ and 13×13 as patch sizes. We tested the parameter sensitivity within a reasonable range.

4.3.2 Visual Comparison of the Discriminativeness of the Learnt Dictionaries and Features

The feature patches learnt by MSDDL are discriminative enough for representing the myocardium separately from the background. In particular a set of feature patches of size 11×11 (without HOG and Gabor) learnt for the myocardium and background are shown in Figure 4.1 to illustrate the discriminativeness of the learnt feature patches.

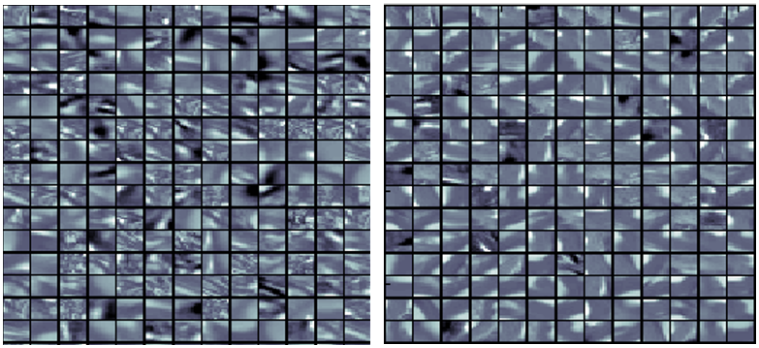

Figure 4.1: Exemplary set of dictionary atoms (without HOG and Gabor) for Background (left) and Myocardium (right) learnt from patches of size 11×11 on CP-BOLD MR.

The motivations behind choosing each step of the proposed MSDDL strategy and the effectiveness of the features learnt by this technique are highlighted in Figure 4.1, where the Cosine Similarity metric [20] is used to determine the most similar patches to a given patch in the MSDDL feature space. When selecting a patch inside the myocardium, without texture and Gram filtering, similar patches are found outside the myocardium too. Adding texture improves somewhat localization, but when considering also Gram filtering, the discriminative strengths of the approach are more evident, since few similar patches are found only within the myocardium. We can observe the gradually improved similarity obtained by incorporating texture features (HOG and Gabor) and Gram filtering over the original patch-based multi-scale dictionary learning technique for a myocardial patch of CP-BOLD MR. Similar behavior is noticed in case of standard CINE MR as well.

4.3.3 Quantitative Comparison

As segmentation quality metric, the Dice coefficient, which measures the overlap between ground truth segmentation masks and

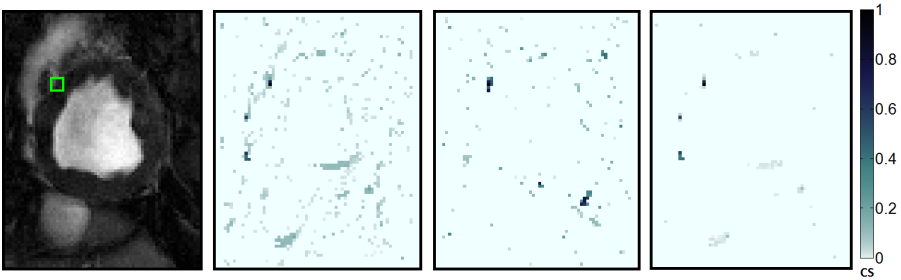

Figure 4.2: Cosine Similarity (CS) between the learnt features showing the advantage of adding texture and Gram filtering. Test patch denoted by a green square in the raw image (first column), MSDDL only on appearance (second column), with texture (third column), and with proposed Gram filtering (final column).

those obtained by the algorithm(s), is employed as described in Chapter 3. For our implementation of Atlas-based segmentation methods, the registration algorithms dDemons [124] and FFD-MI [34] were used to propagate the segmentation mask of the end-diastole image from all other subjects to the end-diastole image of the test subject, followed by a majority voting to obtain the final myocardial segmentation. For level-set class of methods, a hybrid approach of [18] for endocardium and [55] for epicardium is used.

For supervised classifier-based methods, namely Appearance Classification using Random Forest (ACRF) and Texture-Appearance Classification using Random Forest (TACRF) we used random forests as classifiers to get segmentation labels from different features. To provide more context we compare our approach with dictionary-based methods, DDLS and RDDL. DDLS is an implementation of the method in [117], whereas for RDDL we used the discriminative dictionary learning of [99] within the same classification framework that we described in Section 7.2.2. Finally to showcase the strengths of our design choices we considered two additional variants of MSDDL, one without Gram filtering (MSDDL No GF) and one without

texture information as well (MSDDL No GF No Texture). Note that the former is similar to [42] without level-set refinement.

As Table 5.2 shows, overall, when standard CINE acquisition is used, most algorithms perform adequately and the presence of ischemia slightly reduces performance. However, when BOLD contrast is present, other approaches fail to accommodate changes in appearance due to contrast, but MSDDL obtains consistent performance. Specifically, Atlas-based methods are shown to perform well in standard CINE cases but poorly in CP-BOLD. ACRF and TACRF, instead, show very low performance in both standard CINE MR and CP-BOLD MR.

Among dictionary-based techniques, DDLS performs well in standard CINE MR, but underperforms in CP-BOLD MR. Our MSDDL method outperforms all approaches. When comparing it with its variants, it shows that both texture and appearance are important and that the pruning steps based on the Gram matrix are extremely beneficial. Even when we replaced our dictionary learning algorithm with RDDL, an algorithm that forces discrimination by explicitly penalizing the inter-class Gram matrix norm, the results are unimpressive.

These findings are also statistically significant using a paired t-test between the results of MSDDL and the second-best performing one, i.e. DDLS [117]. For both baseline and ischemia cases of CP-BOLD MR, MSDDL shows improved performance compared to DDLS ($\star, p < 0.001$). In the case of standard CINE MR although differences appear small they are still statistically significant, i.e. ($\dagger, p < 0.05$) and ($\ddagger, p < 0.01$) for baseline and ischemia respectively.

4.4 Discussion

Rethinking the assumptions underlying the design of analysis algorithms for standard CINE MR is critical for successfully developing the appropriate analytical tools necessary to meet the new challenges posed by myocardial CP-BOLD MR. In particular, this

Table 4.1: Dice coefficient (mean(std)) for segmentation accuracy in %.

	Baseline		Ischemia	
Methods	CINE	BOLD	CINE	BOLD
Atlas-based methods				
dDemons [124]	60(8)	55(8)	56(6)	49(7)
FFD-MI [34]	60(3)	54(8)	54(8)	45(6)
Level set-based methods				
CVL [18] [55]	50(8)	43(11)	45(9)	37(10)
Supervised classifier-based methods				
ACRF	57(3)	25(2)	52(3)	21(2)
TACRF	65(2)	29(3)	59(1)	24(2)
Dictionary-based methods				
DDLS [117]	71(2)	32(3)	66(3)	23(4)
RDDL [99]	42(15)	50(20)	48(13)	61(12)
MSDDL No GF No Texture	52(8)	51(7)	45(4)	51(6)
MSDDL No GF	62(5)	52(4)	53(5)	57(7)
MSDDL	75(3)†	75(2)⋆	72(2)‡	71(2)⋆

chapter pin-pointed the challenges the BOLD effect poses on these assumptions made when segmenting the myocardium and quantitatively analysed the adverse effect on algorithmic performance. In addition, in this chapter we showed that by learning appropriate features to best represent texture and appearance in CP-BOLD, it is possible to improve the performance of automated algorithms for myocardial segmentation. This chapter also showed overall low performance of state-of-the-algorithms even for standard CINE MR in canine subjects, which can be attributed to the small size of the myocardium.

The MSDDL algorithm does not exploit the temporal information across cardiac phases and doing so should increase performance in future extensions. Figure 4.3 shows segmentation results

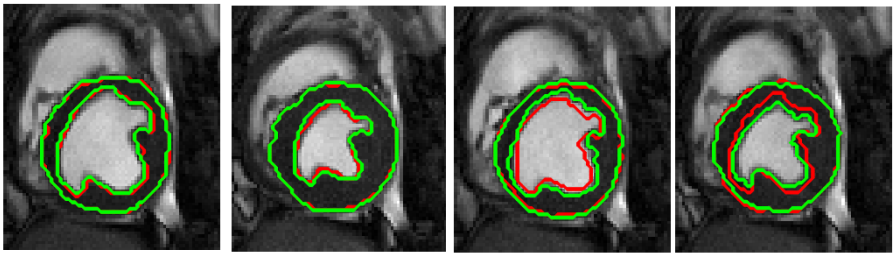

Figure 4.3: The segmentation results showcase the performance of MSDDL for different phases. The contours (red) show good alignment with the ground truth contours (red). However, there are some inconsistencies in segmentations, which should be addressed with spatio-temporal information.

of a BOLD acquisition under baseline conditions. Most of the segmentations show good alignment with the ground truth. However, some phases the segmentations are inaccurate due to lack of smoothness terms in our cost function. In the light of the learnt features in this chapter, we will introduce an unsupervised motion incorporated segmentation algorithm in Chapter 5. Moreover, we will investigate the effectiveness of the feature definitions from this chapter for the task of registration in Chapter 6. The data-driven learning mechanisms proposed in this chapter will also be instrumental in part III of the thesis for joint segmentation and registration.

5

Unsupervised Segmentation with Constraints

In this chapter we investigate two methods we developed for solving the problem of myocardial segmentation in an unsupervised way opposed to the previous chapter, which used a fully supervised mechanism. The first method models background using motion patterns to segment the myocardium region. The later one extracts a region of interest, smooths the data and concatenates motion information with appearance in a dictionary learning setting. Both techniques are unsupervised and rely heavily on different motion patterns of the myocardium region compared its surroundings.

This chapter is based on:

- A. Mukhopadhyay, I. Oksuz, M. Bevilacqua, R. Dharmakumar S. A. Tsaftaris, "Unsupervised Myocardial Segmentation for Cardiac MRI", *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*, pp. 12–20, 2015.
- I. Oksuz,A. Mukhopadhyay R. Dharmakumar S. A. Tsaftaris, "Unsupervised Myocardial Segmentation for Cardiac BOLD", *IEEE TMI*.

5.1 Unsupervised Motion and Sparsity based Segmentation (UMSS) method

In this section, rather than relying on low-level features often used for representing the myocardium when developing segmentation methods for standard CINE MR, which are inconsistent for CP-BOLD MR, a fully unsupervised motion and sparse representation-based feature selection technique has been developed to accommodate the myocardial BOLD effect. The only assumption is that the myocardium moves differently than its surrounding background anatomy. This strategy is also motivated by the findings of Chapter 4 where sparse representation using dictionaries are shown to be invariant under intensity changes. In addition, the sparse representation is capable of retaining semantic information of the myocardium [135]. This essentially enables myocardial segmentation in cardiac MR image sequences (i.e. CINE stack) without any form of manual intervention e.g., landmark selection, ROI selection, spatio-temporal alignment to name a few.

The unsupervised motion and sparse-representation strategy is designed to facilitate this key observation. Each frame is coarsely segmented (over-segmented) based on the optical flow vectors, inspired by [89]. The appearance and motion of the coarsely-segmented background is sparsely represented in a patch-based discriminative dictionary learning technique. A one-class Support Vector Machine (SVM) [107] is employed on the learnt sparse representation of all the pixels in the image sequence to classify myocardium from the background.

The main contributions of this section are twofold. First of all, we revisit fully unsupervised myocardial segmentation technique employing no manual intervention and minimal myocardial motion-pattern assumption for solving general myocardium segmentation problem in standard and emerging cardiac MR imaging modalities. Secondly, we have employed a joint motion and sparse representation based technique, where the motion not only generates

a rough estimate of the myocardium, but also guides the sparse representation stage to a smooth solution based on the motion of the myocardium. Among the very few unsupervised techniques which are fully automated, most similar ones to MSDDL method are those that consider motion as a way to propagate an initial segmentation result to the whole cardiac cycle [45, 57, 93]. Grande et al. [23] integrates smoothness, image intensity and gradient related features in an optimal way under a MRF framework by Maximum Likelihood parameter estimation. Their deformable model estimates the walls based on the MRF along the short axis radial direction. A recent work [15] uses synchronized spectral networks for group-wise segmentation of cardiac images from multiple modalities.

Another idea is to exploit motion and temporal information within the acquired data. In [62] a graph cut algorithm is utilized by simultaneously exploiting motion and region cues. The method uses terminal nodes as moving objects and static background with the intention to extract a moving object surrounded by a static background. Spottiswoode *et al.* [110] used the encoded motion to project a manually-defined region of interest in the context of DENSE MRI. Both of these methods are semi-automated and need interaction to achieve high accuracy. Earlier, we proposed a supervised multiscale discriminative dictionary learning (MSDDL) procedure [76] in Chapter 4. However, unlike the proposed method, only appearance and texture features are considered for sparse representation in MSDDL. In general we can identify, that supervised methods require lots of data for training and a robust feature generation and matching framework. While deep learning can help incorporate the latter (e.g. [7], [80], [119]), it does need significant training data, which can be augmented with in different scenarios [22]. In this section, we instead propose a fully unsupervised method that incorporates motion information in a discriminative dictionary learning framework.

Our proposed Unsupervised Motion and Sparsity based Segmentation (UMSS) method (as shown in Figure 5.1) for segment-

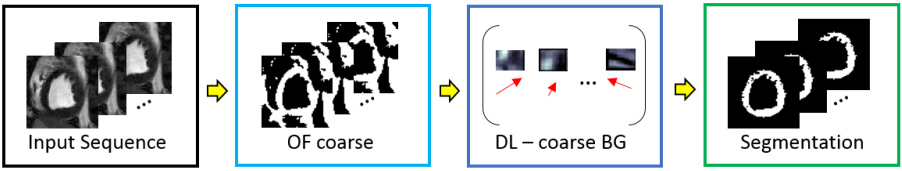

Figure 5.1: Description of the proposed method.)

ing 2D Cardiac MR (both standard CINE and CP-BOLD) image sequences is described here in details.

5.1.1 Optical Flow Based Coarse Segmentation

Our first step is to compute optical flows between two subsequent frames (I_t, I_{t+d}) of the given image sequence using [59]. The motion boundary of the optical flow can be measured simply by calculating the gradient. We have computed the coarse segmentation by applying a threshold T_c on the gradient vectors as shown in Algorithm 4.

5.1.2 Dictionary Learning of Background

Given a sequence of images $\{I_t\}_{t=1}^{T}$ and corresponding coarse segmentation labels obtained from Optical Flow motion boundary as described earlier, we can obtain a matrix, $\{Y = [Y^{cB} Y^{cF}]\}$, where the matrix Y^{cB} and Y^{cF} contains the coarse background and Foreground information respectively. Information is collected from image and motion patches: squared patches centered around each pixel of the image and its corresponding motion matrix. More precisely, the p-th column of the matrix Y^{cB} is obtained by concatenating the normalized patch vector of pixel intensities and motion vectors taken around the p-th pixel in the coarse background. The Dictionary Learning part of our method takes as input this matrix Y^{cB}, to learn a dictionary D^{cB} and a sparse feature matrix X^{cB}. Dictionaries and sparse features are trained via the K-SVD algorithm [2]. We use the "Gram matrix" ($G^{cB} = (Y^{cB})^T Y^{cB}$) to promote diversity

Algorithm 4 Unsupervised Motion and Sparsity based Segmentation (UMSS)

Require: Image sequence from single subject
Ensure: Predicted Myocardium masks across the sequence

1: Calculate Optical Flow f_p at each pixel p between pairs of frames (I_t, I_{t+d})
2: Measure motion boundary from gradient of Optical Flow $B_p = 1 - exp(-\lambda \nabla f_p)$
where λ is the parameter controlling steepness, $B_p \in [0, 1]$.
3: Compute Coarse segmentation $C_p \in \begin{smallmatrix} cB, if B_p < T_c \\ cF, if B_p \geq T_c \end{smallmatrix}$
4: Collect all $C_p \in cB$ and Calculate $Y_p^{cB} = [I_{p\pm\Delta}; f_p \pm \Delta]$
5: Discard atoms with high values in intra-class Gram matrix G^{cB}
6: Learn dictionary and sparse feature matrix with the K-SVD algorithm

$$\underset{D^{cB}, X^{cB}}{\text{minimize}} \|Y^{cB} - D^{cB} X^{cB}\|_2^2 \quad \text{s. t.} \quad \|x_{p \in cB}\|_0 \leq L$$

7: Train one-class SVM on X^{cB} using Equation 5.1
8: Test on all sparse features $X \in x_{p \in (cF \cup cB)}$ for final classification

in the initialization step. The idea is to have a subset of patches as much diverse as possible to train dictionaries and sparse features. To ensure a proper discriminative initialization, patches that correspond to high values in the Gram matrix are discarded from the training before performing K-SVD. We sort the training patches w.r.t. the sum of their related coefficients in the Gram Matrix, and we prune them by choosing a certain percentage.

5.1.3 One-class SVM for Segmentation

The goal of the segmentation problem is to assign to each pixel of the image sequence a label, i.e. establish if the pixel is included in

the background or the myocardium. To perform this classification, we use the coarse Background dictionary $\{D^{cB}\}$ previously learnt with Discriminative Dictionary Learning technique for sparse representation of the appearance and motion features. We compute the sparse feature matrix $X = [X^{cB} X^{cF}]$ for all the pixels of the image sequence with OMP [120]. The classification is performed by constructing the classifier from only the sparse-features of coarse Background class X^{cB} (the p-th column of the matrix X^{cB}, x_p^{cB}, is considered as the discriminative feature vector for the particular pixel p) using a one-class SVM framework [107]. Supposing for each pixel p of coarse Background class, there is a high dimensional feature space F, then each sample is represented in F by $\Phi(x_{p \in cB})$ and the objective function is formulated as follows:

$$\underset{W \in F, \eta \in \mathbb{R}^l, b \in \mathbb{R}}{\text{minimize}} \frac{1}{2} W^T W + \frac{1}{\nu l} \sum\nolimits_{p \in cB} \eta - b \quad \text{s. t.} \quad W \cdot \Phi(x_{p \in cB}) \geq b - \eta, \eta \geq 0 \tag{5.1}$$

Here, W is the normal vector that represents the support, b is the threshold of function f, $\eta_{p \in cB}$ is the slack variable and ν is the parameter that represents the fraction of sample that should be accepted as the other class. During testing, sparse features for all the pixels of the image sequence, stored in matrix $\hat{X}$ are fed to the classifier learnt on the coarse Background features, to classify the myocardial region as the other class. In addition, a Hough transformation-based post processing step is employed by fitting parametric circles to enforce the shape constraint of the myocardium. The Hough transform is applied to the final output of the segmentation algorithm by fitting circles to the segmentation and fill the circular gaps of the generated by the algorithm. The inconsistency of the segmentation due to is adressed in Chapter 7 with joint registration to enforce local smoothness.Note that the Hough transform step is a rudimentary way of ensuring shape-based constraints. Treating it in a more sophisticated way, using probabilistic models (e.g. Graph-cut with coarse Background as a sink as in 7),

improves the performance.

5.2 Results of UMSS

This section offers a qualitative analysis and quantitative comparison of our proposed UMSS method w.r.t. state-of-the-art methods, to demonstrate its effectiveness for myocardial segmentation. Note that our method outperforms all supervised methods from current literature in both baseline and ischemia cases of CP-BOLD MR, whereas yields state-of-the-art results for both baseline and ischemia cases of standard CINE MR.

5.2.1 Parameter Settings

Parameters of UMSS in this chapter we have empirically chosen a d of 5 and a threshold T_c of 0.4 for coarse segmentation based on Optical Flow, 9×9 as the patch size,and a pruning of 10% for Gram Filtering. Each sparse feature has been represented by 5 non-zero elements whereas a dictionary of 10 atoms is chosen for coarse Background representation. We computed the myocardium segmentation across the whole stack of image sequences for each subject and tested the parameter sensitivity within a reasonable range. In Figure 5.2 we showcase the sensitivity of our algorithm to the parameters ν and λ. We selected the optimum λ of 0.5 and ν of 0.2

5.2.2 Quantitative Comparison

As segmentation quality metric, the Dice coefficient, which measures the overlap between ground truth segmentation masks and those obtained by the algorithm(s), is employed. For our implementation of *Atlas-based segmentation methods*, the registration algorithms dDemons [124] and FFD-MI [34] are used to propagate the segmentation mask of end-diastole image from all other subjects to the

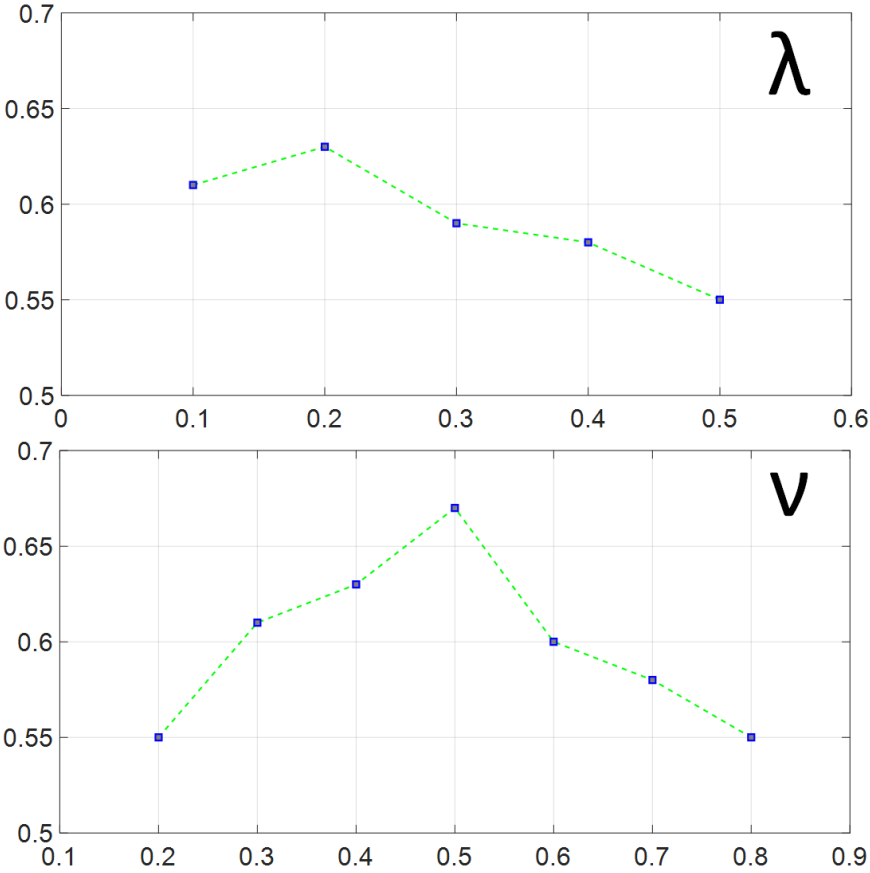

Figure 5.2: The sensitivity of the algorithm performance (Dice Score) to λ and ν parameters.

end-diastole image of the test subject, followed by a majority voting to obtain the final myocardial segmentation. For *supervised classifier-based methods*, namely Appearance Classification using Random Forest (ACRF) and Texture-Appearance Classification using Random Forest (TACRF) random forests are used as classifiers to get segmentation labels from different features. To provide more context, we compare our approach with *dictionary-based methods*, DDLS and RDDL. DDLS is an implementation of the method in [117], whereas the discriminative dictionary learning of [99] is used for RDDL. Finally to showcase the strengths of our design choice of sparse representation using discriminative dictionary learning, we have considered two additional variants of UMSS, without Dictionary Learning (UMSS No DL) and without concatenating optical flow features with intensity for Dictionary Learning (UMSS No Motion). All quantitative analysis for supervised methods are performed using strict leave-one-subject-out cross validation.

As Table 5.2 shows, overall, for standard CINE, most algorithms perform adequately and the presence of ischemia slightly reduces performance. However, when BOLD contrast is present, other approaches fail to accommodate changes in appearance due to contrast, but UMSS obtains consistent performance. Specifically, Atlas-based methods are shown to perform well in standard CINE but poorly in CP-BOLD. ACRF and TACRF, instead, show very low performance in both cases. Among dictionary-based methods, DDLS performs well in standard CINE MR, but under-performs in CP-BOLD MR. When comparing with its variants, UMSS shows that both Dictionary Learning and motion information are extremely beneficial.

5.2.3 Qualitative Analysis of UMSS

The quality of myocardial segmentation by UMSS for both baseline and ischemia cases across standard CINE and CP-BOLD MR is shown in Figure 5.3. Note that UMSS results in very smooth endo- and epicardium contours which closely follow ground truth contours generated by the experts and can be attributed to the suc-

Table 5.1: Dice coefficient (mean (std)) for segmentation accuracy in %.

	Baseline		Ischemia	
Methods	CINE	CP-BOLD	CINE	CP-BOLD
Atlas-based methods				
dDemons [124]	60(8)	55(8)	56(6)	49(7)
FFD-MI [34]	60(3)	54(8)	54(8)	45(6)
Supervised classifier-based methods				
ACRF	57(3)	25(2)	52(3)	21(2)
TACRF	65(2)	29(3)	59(1)	24(2)
Dictionary-based methods				
DDLS [117]	71(2)	32(3)	66(3)	23(4)
RDDL [99]	42(15)	50(20)	48(13)	61(12)
Proposed Unsupervised method				
UMSS No DL	25(9)	26(12)	19(5)	18(7)
UMSS No Motion	49(15)	42(19)	51(14)	53(12)
UMSS	62(20)	71(10)	65(14)	66(11)

cessful representation of myocardial motion.

5.3 Motion incorporation for Myocardial Segmentation with Local Smoothness

This section presents a fully automated and unsupervised method for CP-BOLD MRI with the goal of faithfully preserving the key patterns necessary for diagnosis. The bottom of Figure 3.5C illustrates the results of our method, which does not require any form of manual intervention e.g., landmark selection, ROI selection, spatio-temporal alignment. It builds upon a dictionary approach introduced in Chapter 4 using a joint appearance and motion model introduced. To increase robustness to the BOLD effect, we introduce

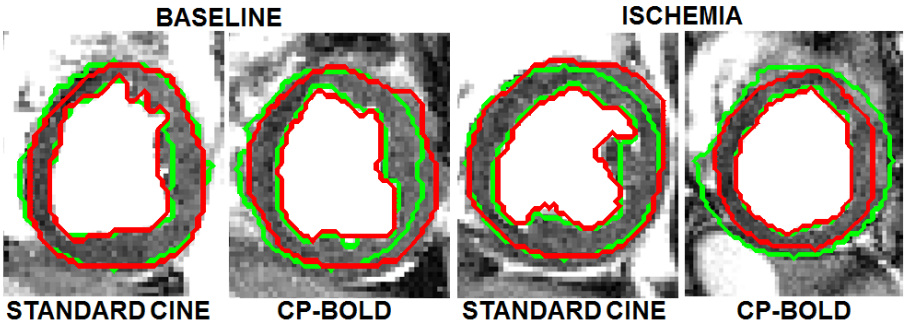

Figure 5.3: Segmentation result (green) of UMSS for both CP-BOLD MR and standard CINE MR at baseline and ischemic condition superimposed with corresponding Manual Segmentation (red) contours delineated by experts.

a pre-processing step that aims to ”smooth out” temporal intensity variations. Subsequently, subject-specific dictionaries of patches of appearance and motion are built from a rudimentary definition of foreground (myocardium) and background (everything else). Projections on these discriminative dictionaries and spatial regularization with a Markov Random Field (MRF) obtains the final result. Extensive experiments show that, not only we obtain higher segmentation accuracy globally and locally around the myocardium, but also that this accuracy translates to better local preservation of BOLD patterns.

There are three main contributions of this section. First, we revisit fully unsupervised myocardial segmentation using discriminative dictionary learning that jointly represents appearance and motion. Second, we utilize a variational inhomogeneity refinement scheme for spatio-temporal smoothing of BOLD signal in a cardiac image sequence. Finally, an extensive segmentation performance analysis both in local and global fashion is proposed.

In the following we detail the proposed method for segmenting 2D(+time) Cardiac MRI data. The method does not rely on manual

intervention and its only assumption is that motion patterns of the myocardium differ from those of surrounding tissue and organs. Our proposed method consists of three main blocks which are illustrated in Figure 5.4 and described briefly below and in detail in the next sections.

The pre-processing block aims to overcome BOLD effects by temporal smoothing using a Total Variation based method and to localize the myocardium to initialize the next step. The extraction of the LV blood pool is solely based on apperance and completely differs with the approach in [77], which relies on optical flow based thresholding.

The second block uses Dictionary Learning to obtain a segmentation. Subject-specific foreground and background dictionaries are trained from the two extracted regions from the entire cardiac sequence. These dictionaries are used to calculate the residuals of the cardiac image to be segmented.

The final block introduces spatial regularization using Markov Random Field (MRF) approach, that is applied on the residuals of the two dictionaries to achieve the final segmentation of the myocardium. This block ensures the local smoothness of the extracted region.

5.3.1 Pre-processing

The overriding goal is to reduce the BOLD effect and obtain regions that patches can be drawn from for learning the dictionaries. This happens in few steps that we detail below and visually in Figure 5.4A. First a Total Variation based filtering technique is used to smooth images to reduce the BOLD effect. Then, a process based on multi-level histogram thresholding is used to find the center of the Left Ventricle (LV) (on the mid-ventricular images we use here). We then segment the LV blood pool with region growing. Finally, aided by the distance transform we identify candidate foreground and background regions to sample from.

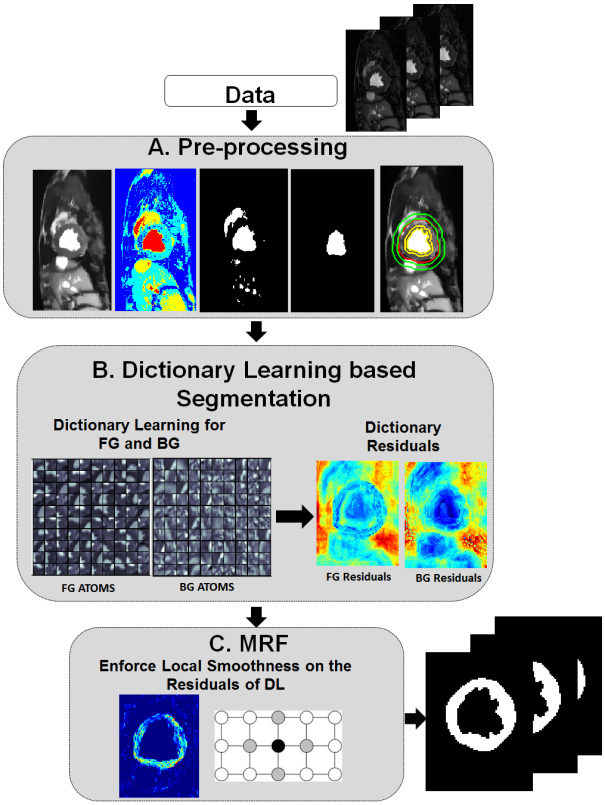

Figure 5.4: Description of the unsupervised segmentation method. Block A aims to find a rough segmentation of the myocardium. Otsu-thresholding to generate four classes color-coded (second column), top two classes are extracted to capture LV blood pool (third column), LV is extracted using a region-growing (fourth column) Rough background and foreground regions are detected (fifth column). In Block B two subject-specific dictionaries are trained on foreground and background on appearance and motion. In Block C an MRF-based post-processing algorithm on the residuals of the two dictionaries is utilized to have smooth boundaries.

Total Variation based smoothing: The BOLD effect poses a significant problem to all state-of-the-art segmentation algorithms as demonstrated in [76] and discussed in the introduction. One way to create robustness is to learn intensity invariant features. However, [76] also demonstrated superior performance when using standard CINE MR. Inspired by this observation, we aim to identify a process that essentially converts the difficult CP BOLD MRI's appearance into a more manageable standard CINE MR like appearance. Variational methods are used extensively in image denoising problems, most famous being the pioneering Rudin-Osher-Fatemi model [102]. Most of the video denoising methods derived from [102] actually work on a frame-by-frame basis. This approach is not suitable in our case since the BOLD effect is spatio-temporal across the cardiac cycle. In this work, we adopted the augmented Lagrangian method [78] developed in [19] to solve the BOLD inhomogeneity refinement problem in a space-time volume. We have employed the ℓ_1-norm Total Variation (ℓ_1-TV) using the augmented Lagrangian method introduced in [19] for solving both the problems together. The energy functional we have used for this particular minimization problem is:

$$\underset{f}{\text{minimize}}\ \frac{\mu}{2}\|u - v\|_1 + \| \bigtriangledown u\|_2,$$

where v is the input $2D+t$ image series and u is the processed image series. The main reason behind choosing ℓ_1-norm over ℓ_2-norm is the fact that appearances of different anatomies are piece-wise constant functions [38]. They also demonstrate quantized levels (i.e., a function can only take a given energy level without any other level existing between two anatomies), within a certain anatomy and sharp edges across anatomical boundaries. These boundaries and anatomies can be better preserved when using the ℓ_1-norm as shown in Figure 5.6.

LV center point detection and blood pool extraction: To extract the blood pool, first multiple thresholds are found using Otsu's histogram thresholding [86] for each image in the cycle to obtain a

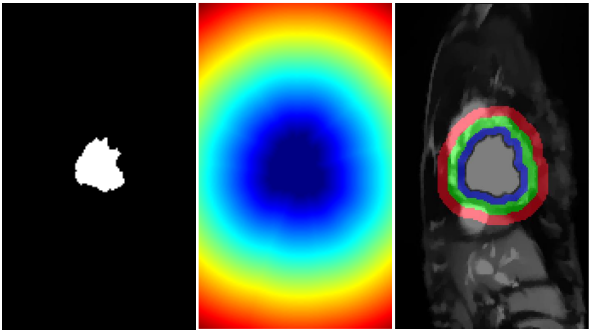

Figure 5.5: Extracting candidate background and myocardium regions. LV blood pool (left); Distance transform from the LV blood pool boundary (middle); Rudimentary background and foreground classes (right). Only pixels within the blue and red rings (right panel) are used to sample patches for dictionary learning. The green ring acts as boundary in between these two regions to reduce the chance of false positives.

four-class segmentation: loosely capturing blood pool (brightest in both standard CINE and BOLD weighted imaging), partial volume between myocardium and blood pool (second brightest), myocardium (third brightest) and other (most dark) –adapting broadly ideas from [97]. The brightest two classes are used to extract the blood pool region. Then, the region that fits most closely a circle (of a roughly known diameter) is found, which eventually is used to determine the middle point of LV blood pool. Finally, a region-growing approach is employed to delineate the LV blood pool.

Finding foreground and background regions to sample from: The distance transform from the LV blood pool is used to define two ring-like areas identifying foreground and background regions to sample from as visualized in Figure 5.5. In this paper we use a ring thickness of $R = 6$mm at end systole for all rings involved. We

vary the thickness value of the myocardium in experimental results to test the robustness. The thicknesses are normalized according to the cardiac phase to ensure that these regions do not include false positives with the following function: $R\frac{f}{f+|f_t - f_{ES}|}$; where f represents the total number of cardiac phases, f_t represents the frame number of the current phase and f_{ES} is the end systolic frame. End systolic frame is defined around 30% of the cardiac cycle in accordance with ECG triggering. The regions for foreground M^F (blue ring in figure. 5.5) and background M^B (red ring in Figure 5.5) will be utilized to draw patch samples to learn the dictionaries.

The goal of the last two steps is to obtain a soft definition of where to sample patches from for myocardium and background. Any similar methodology will suffice. Experiments in Figure 5.13 show the precision of the last two pre-processing steps does not have a major influence on the performance of the overall algorithm.

5.3.2 Dictionary Learning

Joint learning of discriminative dictionaries for segmentation problems is a recent idea also developed in our earlier study [76]. The different motion patterns of the myocardium and background is utilized here to guide myocardial segmentation. Sparse representation of motion guides the sparse representation stage to a smooth solution based on the motion of the myocardium. The discriminative dictionary learning idea has been proposed earlier in atlas-based segmentation of brain MRI [10, 117] and abdominal CT [118] but without the context of motion. In this chapter we have developed a method to jointly model the appearance and motion within a discriminative dictionary learning framework by incorporating both appearance and motion within the observations.

Our method builds observations from the concatenation (after raster-scanning) of square patches of appearance (pixel intensities) and corresponding motion (found via optical flow). Specifically, given (1) a series of pre-processed images $I_t, \{t = 1; \ldots, T\}$, (2) the

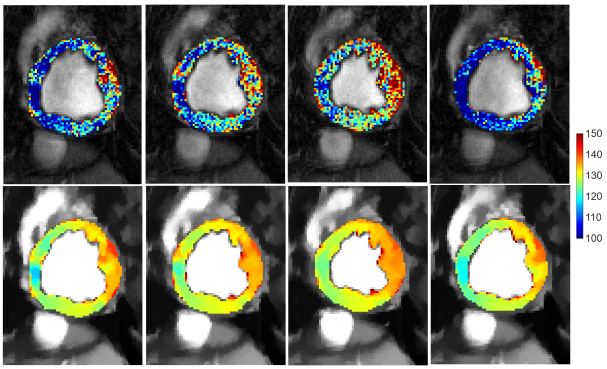

Figure 5.6: Influence of Total Variation based smoothing on different cardiac phases of a healthy subject. Four temporal phases of the same acquisition of a subject before (top) and after pre-processing (bottom), where myocardial intensities have been color-coded to aid visualization. Observe, how myocardial intensities appear smoother and within the same (and shorter) range across the cardiac cycle after TV-based smoothing (bottom row). The smoothing happens on the entire image and visualized only on the myocardium region.

estimated optical flow between subsequent images I_t and I_{t+d} and (3) the corresponding coarse segmentation labels M_t^F and M_t^B obtained as previously described, two matrices were obtained, Y^B and Y^F, where these matrices contain the data from the coarse background and foreground information from the entire cine stack respectively. The j-th column of the matrix Y^F is obtained by concatenating the normalized patch vector of pixel intensities and motion vectors calculated by the method in [14] taken around the j-th pixel in the coarse foreground as shown in figure. 5.7. Both horizontal and vertical components are used for each pixel. The Dictionary Learning part of our method takes as input these two matrices Y^B and Y^F, to learn dictionaries D^B, D^M and a sparse feature matrix X^B, X^F.

In order to achieve discriminative initialization, highly corre-

lated data are disregarded prior to learning. We calculate for a given class C (foreground or background), the intra-class Gram matrix as:

$$G^C = (Y^C)^T Y^C. \tag{5.2}$$

We sort the training patches w.r.t. the sum of their related coefficients in the Gram Matrix, and we prune the top $\%10$ of the atoms to promote discriminative power.

Subsequently, dictionaries consisting of K atoms and sparse features with L non-zero elements are trained via the K-SVD algorithm [2]:

$$\arg\min_{D^F, X^F} \|Y^F - D^F X^F\| \quad \text{s. t.} \quad \forall i \in M^F, \|x_i^F\|_0 \leq L,$$

$$\arg\min_{D^B, X^B} \|Y^B - D^B X^B\| \quad \text{s. t.} \quad \forall i \in M^B, \|x_i^B\|_0 \leq L$$

Considering this patch-based approach in the context of a segmentation problem, given a new image I_t we have an unseen data matrix Y, from the extracted regions obtained from all pixels inside the third ring from the test image. Y contains vectors $y_i; \{i = 1, \ldots, Q\}$ obtained from image and motion patches for all Q pixels. The goal is to assign to each pixel of the image a label, i.e. establish if the pixel is included in the background or the foreground region.

To perform this classification, we use the dictionaries, D^B and D^F, previously learnt. The Orthogonal Matching Pursuit (OMP) algorithm [120] is used to compute, the two sparse feature matrices $\hat{X}^B$ and $\hat{X}^F$ for a given sparsity level.

5.3.3 MRF Based Smoothing

In this study, we employ a frame-by-frame MRF strategy [11] across all image pixels to enforce spatial regularization on the final segmentation for each image I_t. The process ensures local smoothness of

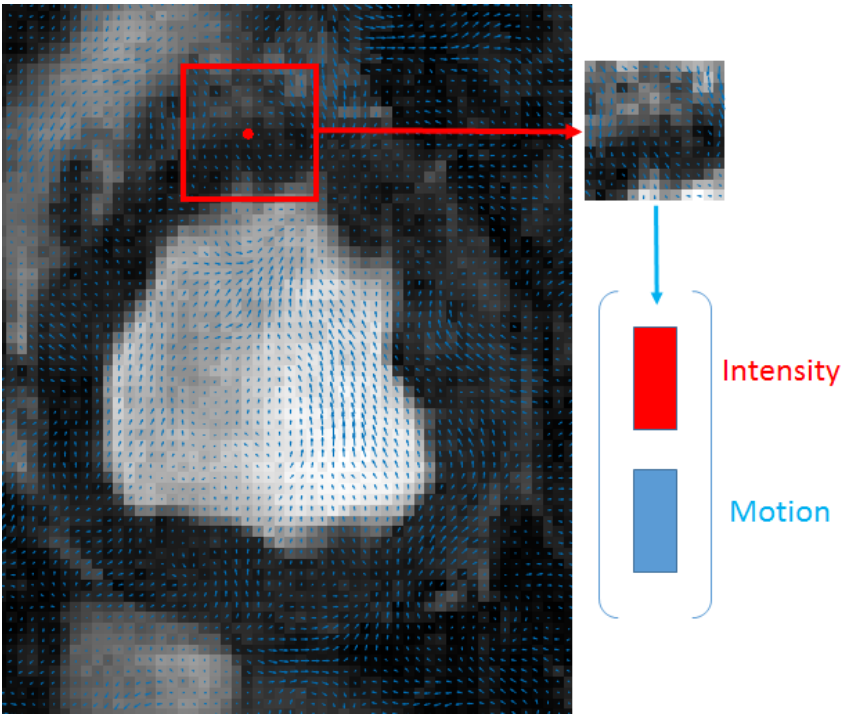

Figure 5.7: The feature vector generation as concatenation of intensities of square patches and corresponding motion vectors inside that patch.

Algorithm 5 Proposed Method

Require: Image sequence from single subject
Ensure: Predicted Myocardium masks across the sequence

1: Calculate Optical Flow $\boldsymbol{f}_p$ at each pixel p between pairs of frames (I_t, I_{t+d})
2: Generate Y^B and Y^F concatenating image intensities and motion information for each patch
3: **for** C={B,F} **do**
4: Discard atoms with high values in intra-class Gram matrix G^F and G^B
5: Learn dictionary and sparse feature matrix with the K-SVD algorithm

$$\underset{D^C, X^C}{\text{minimize}} \|Y^C - D^C X^C\|_2^2 \quad \text{s. t.} \quad \|x_i^C\|_0 \leq L$$

6: Compute the inter-class Gram matrix G^{IC}
7: Discard from D^B and D^F atoms with high values in G^{IC}
8: **end for**
9: Learn residuals R_B and R_F given Y, D^B and D^F with OMP algorithm
10: Test on all residuals R_B and R_F for first classification
11: Use MRF-based segmentation on the residuals R_B and R_F using Equation 5.3

the classification, which is refined according to the labels. Given the residuals for background R_B and foreground R_F the final segmentation is obtained by minimizing the MRF-based energy functional:

$$E_{\text{MRF}}(I_t) = \sum_{p \in I_t} (V_p(I_p) + \lambda V_{pq}(I_p, I_q)) \tag{5.3}$$

where $V_p(\cdot)$ corresponds to the unary potentials representing the data term for node p and $V_{pq}(\cdot)$ corresponds to the pairwise potentials representing the smoothness term for pixels at nodes p and q in a neighborhood N in the image I_t. The data term measures the disagreement between the prior and the observed data, which is based on the residuals of dictionaries. For a pixel p with initial label C: Label$(p) = C$, data term is: $V_p(I_p) = R_C$. The smoothness term is defined as; $V_{pq}(I_p, I_q) = \sum_{q \in N} R_{C'}$ on the nodes that have different class Label$(q) = C'$ in the neighborhood N. The parameter λ controls the trade off between smoothness and data term that govern the final segmentation. The smoothness term penalizes discontinuities in a neighborhood N. In our implementation, the total energy is calculated using the residuals for the possible labels of foreground R_F and background R_B. More precisely, if $R^F = \|y_i^{BF} - D^F \hat{X}_i^F\|_2$ is larger than $R^B = \|y_i^{BF} - D^B \hat{X}_i^B\|_2$, the patch is assigned to the background; otherwise, it is considered belonging to the foreground region for the initial segmentation. The update occurs if the total energy calculated adding the unary and pairwise terms for both labels possibilities and picking the smaller one for the next iteration as detailed in Algorithm 5. The method converges either when there is no change of labels or the maximum number of iterations are reached.

5.4 Experimental Results

This section offers a qualitative and quantitative analysis of the proposed method, as well as quantitative comparison of our pro-

Table 5.2: Dice coefficient (mean(std)) for myocardial segmentation accuracy in % for standard CINE (CINE) and cardiac BOLD images (BOLD)

	Baseline		Ischemia	
Methods	CINE	BOLD	CINE	BOLD
Atlas-based methods				
dDemons [124]	60 (8)	55 (8)	56 (6)	49 (7)
FFD-MI [34]	60 (3)	54 (8)	54 (8)	45 (6)
Supervised classifier-based methods				
ACRF	57 (3)	25 (2)	52 (3)	21 (2)
TACRF	65 (2)	29 (3)	59 (1)	24 (2)
Dictionary-based methods				
DDLS [117]	71 (2)	32 (3)	66 (3)	23 (4)
RDDL [99]	42 (15)	50 (20)	48 (13)	61 (12)
MSDDL [76]	75 (3)	75 (2)	75 (2)	71 (2)
UMSS [77]	62 (20)	71 (10)	65 (14)	66 (11)
Proposed Unsupervised method				
Proposed No TV	65 (6)	59 (7)	63 (8)	57 (9)
Proposed No Gram Filtering	62 (5)	52 (4)	53 (5)	57 (7)
Proposed No Motion	71 (6)	69 (8)	67 (9)	68 (8)
Proposed No MRF	74 (5)	75 (6)	73 (7)	72 (6)
Proposed	77 (10)	77 (9)	74 (7)	74 (6)

posed method w.r.t. state-of-the-art methods, to demonstrate its effectiveness for myocardial segmentation.

Our quantitative analysis consists of comparing our method with others but also looking into regional effects and performance. Unless otherwise noted we use 13×13 patch size, a dictionary of $K = 400$ atoms, a sparsity level of $L = 4$, as parameters. Their influence (and computational performance of the our method) and are discussed in subsection 5.4.6.

Since part of our analysis is to evaluate how errors in segmentation affect the BOLD response (and its patterns) we use cosine

similarity to evaluate the match between two intensity signals (e.g. timeseries), which is defined for two signals S^A, and S^B as:

$$CS = \frac{S^A \cdot S^B}{|S^A||S^B|}$$

where $|\cdot|$ corresponds to ℓ_2 norm of the vector. (We multiply with 100 to report in % in some cases.)

Methods of Comparison and Variants

Our data set consists of the same 10 canines imaged under four different settings as described in Chapter 4. All quantitative analysis for supervised methods were performed using a strict leave-one-subject-out cross validation. For our implementation of *atlas-based segmentation methods*, the registration algorithms dDemons [124] and FFD-MI [34] are used to propagate the segmentation mask of all other subjects to the image of the test subject, followed by a majority voting to obtain the final myocardial segmentation. For *supervised classifier-based methods*, namely Appearance Classification using Random Forest (ACRF) and Texture-Appearance Classification using Random Forest (TACRF) random forests are used as classifiers to get segmentation labels from different features. To provide more context, we compared our approach with *dictionary-based methods*, DDLS, RDDL, MSDDL and UMSS. DDLS is an implementation of the method in [117], whereas the discriminative dictionary learning of [99] was used for RDDL. MSDDL [76] uses a multi-scale supervised dictionary learning approach with majority voting classification. UMSS [77] is a unsupervised method relying only on a motion-based coarse segmentation of background. This method learns background class only with a dictionary and performs classification with one-class SVM. Finally, to showcase the strengths of our design choices, we considered three additional variants of our method, without Total variation pre-processing (Proposed No TV), without Gram filtering (Proposed No Gram Filtering), without concatenating optical flow features with intensity for Dictionary

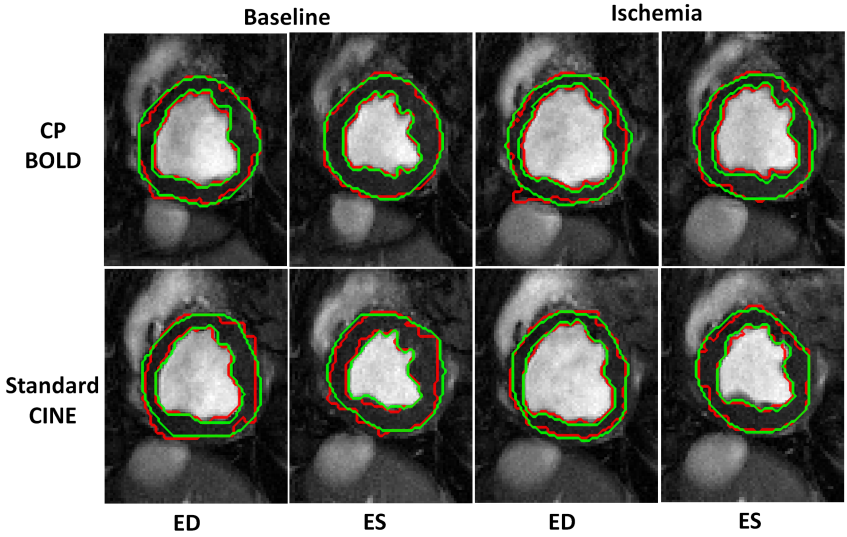

Figure 5.8: Segmentation result (red) of Proposed method for both CP-BOLD MR and standard CINE MR at baseline and ischemic condition for End-diastole (ED) and End-systole (ES) superimposed with corresponding Manual Segmentation (green) contours delineated by experts.

Learning (Proposed No Motion) and without spatial regularization using MRF (Proposed No MRF).

5.4.1 Comparison with State-of-the-art Methods

The visual quality of myocardial segmentation by the proposed method for both baseline and ischemia cases across standard CINE and CP-BOLD MR is shown in Figure 5.8. The End-diastole (ED) and End-systole (ES) phases are picked as exemplary images from the entire cardiac cycle. Note that our method results in very smooth endo- and epi-cardium contours, which closely follow ground truth contours generated by the experts and can be attributed to the successful representation of myocardial motion.

These observations also hold quantitatively when relying on the Dice metric for evaluation. As Table 5.2 shows, overall, for standard CINE, most algorithms perform adequately well and the presence of ischemia slightly reduces performance. However, when BOLD contrast is present, other approaches fail to accommodate changes in appearance due to contrast, but proposed method obtains consistent performance. Specifically, Atlas-based methods, ACRF and TACRF all shown to perform better in standard CINE compared to CP-BOLD. Among dictionary-based methods, DDLS performs well in standard CINE MR, but under-performs in CP-BOLD MR. Observe that the proposed method outperforms MSDDL algorithm, which, unlike the proposed, does not rely on a subject-specific dictionary but on one trained on other subjects. Unsupervised UMSS is also outperformed by the proposed method due to use of two discriminative dictionaries instead of relying on only background dictionary. When comparing with its variants, the proposed shows that all processes including, pre-processing, Gram Filtering, motion information and MRF-based spatial regularization, are extremely beneficial. One issue of interest is the superior performance compared to the supervised methods. This is due to the lack of limited training data and superior performance of achieved with motion incorporation.

5.4.2 Segmental Analysis

In this Section, we perform a detailed regional analysis of the segmentation results by taking into account the spatial distribution of the errors. For each myocardium segmented both manual and automatically, we divide it in 6 radially concentric regions, following the six-segment AHA model for the mid-ventricular slice [17]. Specifically, we take the manually segmented masks and divide them to six radially concentric regions 0°, 60°, 120°, 180°, 240° and 300°. As a reference, a diagram of this process, known as bull's eye view, is shown in Figure 5.9 along with anatomical nomenclature for each six segments.

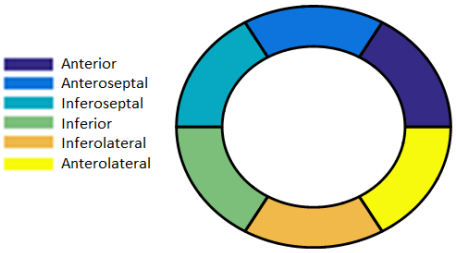

Figure 5.9: Six segments of mid-ventricular myocardial slice

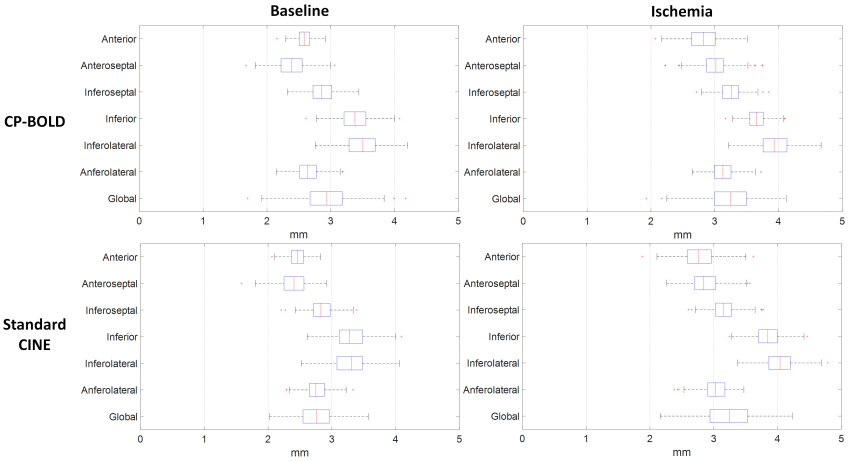

Figure 5.10: Segmental accuracy for CP-BOLD and standard cine MR for epicardium.

5.4.3 Quantitative Analysis

In Figure 5.10 boxplots of the Hausdorff distance metric for the epicardium for CP-BOLD and standard Cine MR are presented. Endocardium results show sub-pixel accuracy on average. The boxes represent the lower quartile, median and upper quartile values; the whiskers represent the whole extension of the error distribution whereas the crosses correspond to outliers. The global error distri-

Table 5.3: Regional segmentation accuracy measured via Dice (mean ± std) in % for Standard CINE and CP-BOLD.

	Baseline		Ischemia	
Regions	Std. CINE	CP-BOLD	Std. CINE	CP-BOLD
Anterior	81±13	83±10	78±10	79±8
Anteroseptal	79±10	82±9	75±10	75±9
Inferoseptal	75±12	72±16	75±12	75±9
Inferior	72±11	70±12	69±11	71±8
Inferolateral	73±8	72±12	71±13	71±11
Anterolateral	82±7	81±9	76±11	74±9

bution shows the presence of two outliers, whereas the remaining segmentations have mean errors lower than approximately 4 mm for images with 1.25 mm spatial resolution. Our reported results of Hausdorff distance are at par with [97]. In the case of Hausdorff distance errors, largest values are located at the inferior region mainly due to the presence of liver.

A comparison is shown in Table 5.3 to indicate the stability of the method when ischemia is present. The Dice overlap measure is calculated for the 6 regions of the myocardium. In general our algorithm is robust to regional complexities of the myocardium. Ischemia appears to slightly influence the performance especially in the regions that are under influence of LAD stenosis (Anteroseptal, Anterior and Anterolateral).

5.4.4 Time Series Analysis for Ischemia Detection

It is important to evaluate quantitatively the influence of segmentation errors on preserving the BOLD effect to reduce errors of ischemia detection methods [12,122,123]. As a benchmark, we used the BOLD signal intensity as obtained via averaging (and normalizing) pixel values in various regions with and without disease obtained from myocardial definitions from ground truth or algorithm

Table 5.4: Cosine Similarity comparison of Timeseries for 6-segmental regions (mean ± std, in %) acquired from the ground truth compared with the proposed method and Atlas-based method [124] for CP-BOLD sequences.

	Proposed		Atlas-based [124]	
Regions	Baseline	Ischemia	Baseline	Ischemia
Anterior	93±2	89±3	89±4	86±5
Anteroseptal	92±5	83±6	89±5	81±8
Inferoseptal	82±5	83±9	80±8	80±11
Inferior	79±4	80±8	75±8	77±11
Inferolateral	81±3	80±9	81±3	80±9
Anterolateral	91±3	83±5	88±5	81±7

results. Figure 3.5C already alludes that our proposed approach outperforms other segmentation methods, and this performance also holds when disease is present (see Figure 5.11). This also holds quantitatively when comparing with an Atlas-based method [124] as an illustrative example, using the cosine similarity metric (see Table 5.4). Evidently, small errors (even 5-10 pixels) in segmentation towards hyperintense (blood pool) or hypointense (lung/liver interface) areas when a myocardial region is as small as 100 pixels in systole have severe effects in preserving the BOLD signal.

5.4.5 Temporal Evaluation of Results:

Since we use motion patterns as a feature, one important evaluation of the algorithm is the temporal performance change. As the heart moves with different velocities between frames at different stages of the cardiac cycle (for example, during diastole there is much less motion than during systole). We illustrated the performance on how well the algorithm performs over different temporal phases of the cardiac cycle in Figure 5.12.

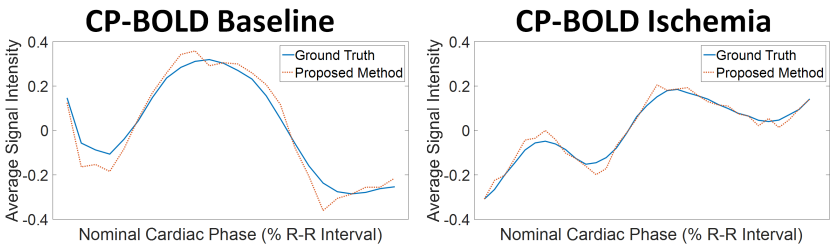

Figure 5.11: Normalized time series obtained by averaging pixel intensities in the anterior region, as defined using ground truth (blue) and automatic segmentation (red dotted line) in a subject at baseline (left) and after LAD stenosis and during ischemia (right). Observe that the time series obtained via the proposed segmentation is consistent with that of ground truth, which eventually result in more accurate ischemia detection.

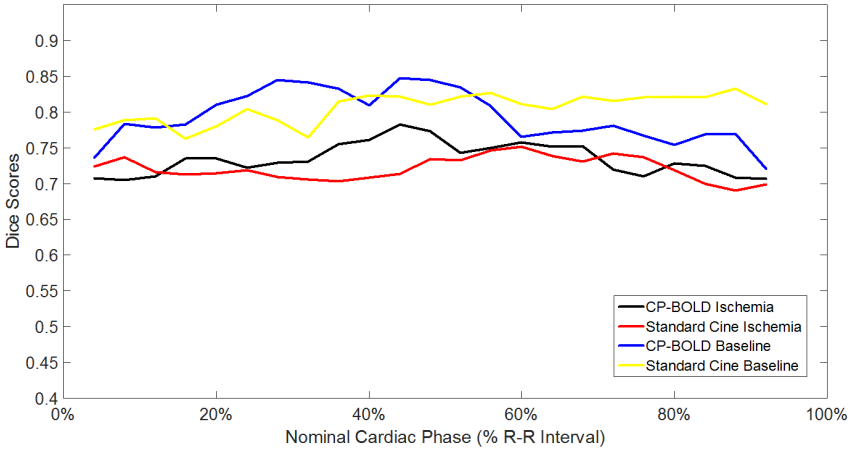

Figure 5.12: Dice accuracy on different temporal frames. The results are the average values for CP-BOLD and standard cine under baseline and ischemia conditions.

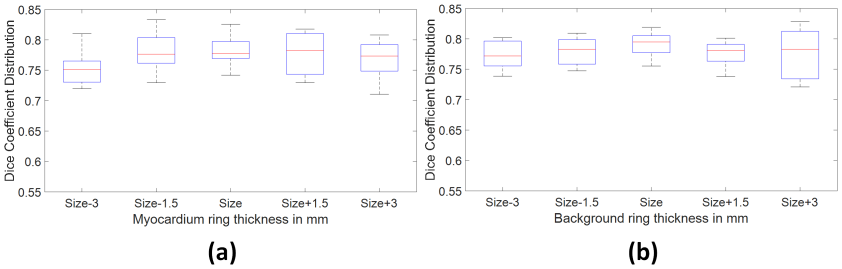

Figure 5.13: Effect of Pre-processing on segmentation accuracy. Rudimentary class thickness is varied from the original size (6mm) for background (a) and myocardium (b). The influence of changing the thickness from 3mm to 9mm of both classes on segmentation accuracy is minimal.

5.4.6 Parameter Analysis and Computational Performance

The purpose of this section is to analyze effects of different parameters of the algorithm as well as discuss computational performance. Overall we evaluate patch size, number of atoms K, and sparsity level L independently. We vary one of the 3 but keep the other two fixed using the following values: patch size of 13×13, $K = 400$ and $L = 4$.

Influence of Pre-processing

Pre-processing consists of identifying both background and myocardium regions to sample from, which depend on the thickness of the rings that define them. Here we vary this ring size (from the initial size of 6mm) keeping all other parameters fixed. Figure 5.13 illustrates that the results remain consistent whether modifying more the background (more false negatives) or the myocardium (more false positives) class. This result demonstrates that we can tolerate imprecision in defining the regions to sample from.

Influence of Patch Size

The patch size is related to the local geometry whilst the neighborhood size reflects the anatomical variability. The Dice coefficient distributions over varying patch are presented in Figure 5.14 a for a dictionary size of $K = 400$ atoms, and sparsity of $L = 4$. As one can observe, the best median Dice coefficient was obtained with a patch size of 13×13 albeit it performed similar to 15×15. This is to be expected as this comes close to the average size of the myocardium given the image size of our dataset.

Influence of Dictionary Size and Sparsity Level

First, experiments were carried out to study the influence of dictionary size K (the number of atoms in each dictionary) on segmentation accuracy with fixed values 13×13 patch size and $L = 4$ sparsity threshold. As illustrated by Figure 5.14 b, 400 atoms provide a good balance of accuracy w.r.t. dictionary size. Note that a larger dictionary does imply higher computational complexity, albeit it also depends on sparsity level.

Thus, experiments were also carried out to study the influence of the sparsity level L (the number of non-zero components in sparse coefficients) on segmentation accuracy. This governs the selection of atoms to be combined for the purpose of representing classes with the dictionaries. Figure 5.14 c shows that sparsity 4 is the most suited level of sparsity for our experiments and indicates the importance of this parameter. It appears that lower sparsity has higher discriminative ability as adding additional atoms it appears to add noisy information.

Computational Complexity

Execution time on a 2.4 GHz processor with an average data set ($192 \times 114 \times 30$) is approximately seven minutes. Most of this time is spent on the dictionary learning stage (approx. 4 minutes).

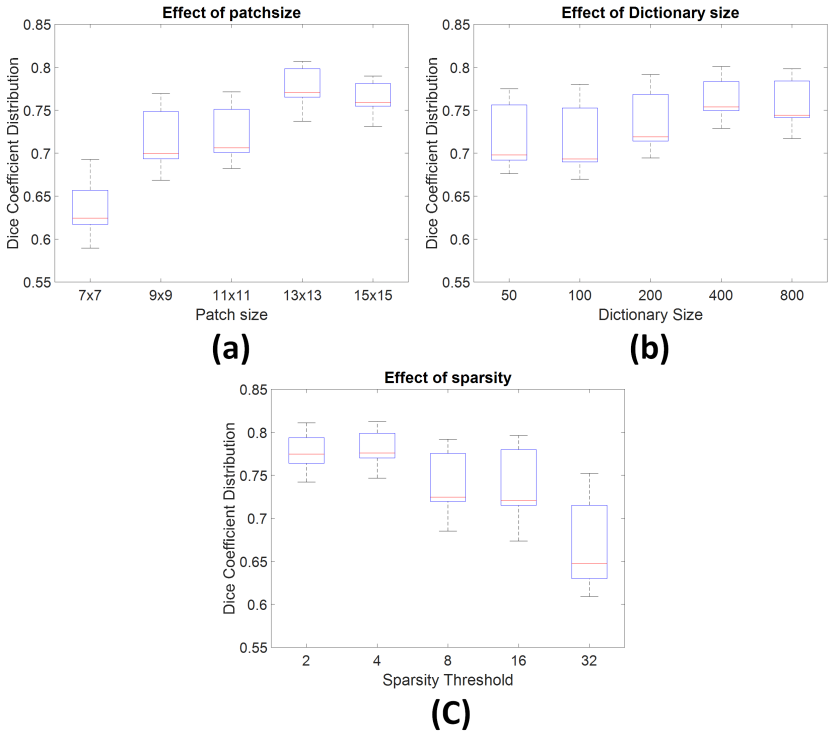

Figure 5.14: Effect of patch size (a), dictionary size (b) and sparsity threshold (c) on segmentation accuracy. The optimal results were obtained using a patch size of 13×13, a dictionary of 400 atoms and a sparsity threshold of 4.

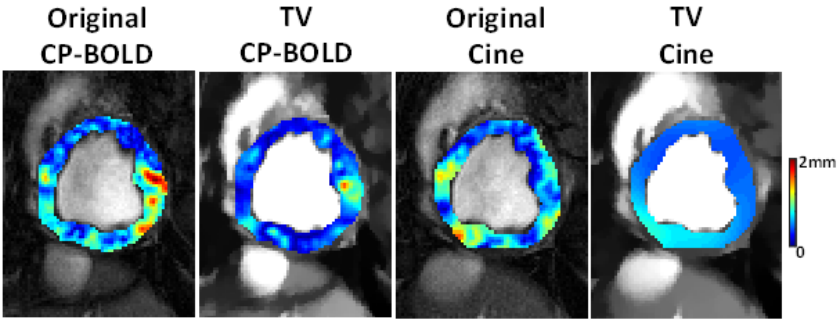

Figure 5.15: Change of motion magnitudes between consecutive phases with TV-smoothing. The expectation of minimal motion is achieved with TV-smoothing.

5.5 Discussion

This chapter motivates us to rethink the standard assumptions regarding the segmentation of the myocardium in MR image sequences, especially to accommodate emerging cardiac MR imaging modalities. In particular, deviating from fully supervised techniques (the performance of which heavily depends on the amount of training data) towards unsupervised ones can benefit in multitude of ways: from operating on no training data, better handling of variability in image contrast to no manual intervention. In addition, this work has shown that unsupervised methods can still deliver state-of-the-art performance even for standard CINE MR. The proposed algorithm does not exploit the spatio-temporal information across cardiac phases and doing so by introducing graph-based formulation should increase performance in future extensions. UMSS can be an effective tool in challenging datasets where inter-acquisition variability prohibits the effectiveness of supervised segmentation strategies.

The results show that the unsupervised automatic segmentation resulting from the proposed method results in an acceptable level

of agreement with the manual segmentations. The main challenge of CP-BOLD data set stems from the inconsistent appearance of patterns. To this end, we have introduced two methodological innovations: discriminative dictionary learning for appearance and motion pattern as well as variational appearance inhomogeneity refinement. The inhomogeneity refinement scheme helps to reduce the intensity inhomogeneity which boosts the performance. The subject-specific dictionary learning approach enables learning of discriminative patterns (represented as atoms). Moreover, MRF-based smoothing improves segmentation outcomes due to influence of spatial smoothing in local neighbourhoods.

An important observation is the significance of subject-specific dictionary learning rather than relying on a training dataset. As experiments showed in Table 5.2, training from other subjects introduces less discriminative atoms inside the dictionaries, which results in poor performance of classification.

As experiments on time series comparisons showed, accurate segmentation translates directly to the fidelity of the signal that we aim to preserve, namely: BOLD contrast. This will have direct effects on fully automating ischemia detection [12].

One interesting interpretation of the results is the the impact of TV-smoothing on the overall method. TV-smoothing contributes the most by reducing the BOLD effect (and standardizing the intensities) as we highlight in experimental results. This mostly improves the accuracy of the motion patterns calculated with optical flow. As for the standard Cine the inherent noise in data is smoothed with this process and the performance is boosted by learning more accurate motion vectors with optical flow. The Figure 5.15 illustrates optical flow for two consecutive phases, where the motion is ideally minimum. The magnitude of the motion calculated on original images is larger than expected motion, which is reduced using TV-smoothing for both standard Cine and CP-BOLD. These aforementioned points regarding the influence of TV-preprocessing on the performance and especially on the motion vectors are added

to discussion.

In conclusion, this chapter underlines the necessity to re-evaluate the standard assumptions and verification metrics regarding the segmentation of the myocardium in cardiac MRI. Development of MR technologies bring new challenges and departing from fully supervised techniques (the performance of which heavily depends on the amount of training data) towards unsupervised ones can provide multiple benefits. Finally, this chapter has shown that global DICE score on its own is not a sufficient performance metric and more analysis can bring about the suitability of segmentation methods for particular MR techniques.

6

Registration

Nonrigid image registration is an essential step in medical imaging for automatic segmentation, motion tracking and morphometric analysis [111]. However, since most of the proposed registration algorithms rely on a (dis)similarity metric build based on the assumptions of consistent intensity and local shape, images with pathologies and locally varying intensity may not be accurately aligned.

There is no CP-BOLD MR myocardial registration algorithm for establishing correspondences in a cardiac sequence. Due to this absence either segmental information [105] or synthetic data sets are used [106], to obtain pixel-wise time series. The main assumption of the this chapter, that it is due to lack of proper similarity criteria. Rather than relying on low-level features used often for myocardial

This chapter is based on:

- I. Oksuz, A. Mukhopadhyay, M. Bevilacqua, R. Dharmakumar S. A. Tsaftaris, "Dictionary Learning Based Image Descriptor for Myocardial Registration of CP-BOLD MR", *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015*, pp. 205–213, 2015.

registration of standard CINE MR, a more distinguishing descriptor should be developed to accommodate the BOLD effect.

In this chapter, a feature-based descriptor as a similarity measure is proposed for the alignment to register the myocardium in the entire cardiac sequence of CP-BOLD. A patch-based discriminative dictionary learning technique [42] is adopted as in chapter 2 to learn features from data. Our motivation is to employ a compact and high-fidelity low-dimensional subspace representation, which is able to extract semantic information of the myocardium pixels [76]. We observe that although the patch intensity level varies significantly across the cardiac cycle, sparse representations based on learnt dictionaries are invariant, as well as unique and robust. The discriminative dictionary learning strategy is designed to facilitate this key observation regarding CP-BOLD.

During training, two dictionaries of patches for myocardium and background are learnt offline. To register two images with unknown myocardium masks, the sparse representations that are obtained on the basis of previously trained dictionaries for background and myocardium, are concatenated and considered as the feature for that particular pixel. The similarity term evaluates the match of the sparse features at every iteration on a pixel level. The sum of squared differences of the sparse representations between the target image and warped source image are utilized as similarity criteria.

There are three major contributions of this chapter. First of all, we propose a sparse representation-based image descriptor in a registration framework, for the first time to the best of our knowledge. Second, we experimentally validate the fact that BOLD contrast significantly affects the accuracy of registration algorithms (including intensity-based and feature-based methods), which instead perform well in standard CINE MR. Finally, we address the fundamental problem in handling BOLD contrast by designing a set of compact features using discriminative dictionary learning, which can effectively represent the myocardium in CP-BOLD MR.

6.1 Background

Automated myocardial registration for standard CINE MR is a well studied problem [114]. Most of these algorithms can be classified into two groups according to similarity criteria used: intensity-based or feature-based. General intensity-based registration algorithms can be summarized as an energy minimization procedure, where the energy functional is [111]:

$$E = \sum_{p \in \Omega} DS(I_s(p), I_t(p+u)) + \lambda E_R, \tag{6.1}$$

where Ω represents the entire image domain, and p denotes a pixel in the domain. Non-rigid registration consists of minimizing a dissimilarity measure *DS* between a source image I_s and a target image I_t, u denotes the displacements and E_R denotes the regularization term. In this chapter, we are particularly interested in the definition of the similarity measure. Sum of squared differences (SSD) and cross correlation (CC), are the earlier metrics utilized in registration. Recently, information theory-based approaches gained attention, e.g., derivatives of Mutual Information (MI), which is based on individual and joint gray level distributions [96].

When registration under inhomogeneity conditions is required, some have proposed modifications on regional intensity distortions (for brain MRI) [112] or spatially intensity variations [149]. Alternatively, feature-based approaches can be used. A recent example is DRAMMS [87], where the similarity is based on optimal Gabor attributes. Another approach, MIND [39], relies on regional information following the footsteps of self-similarity (a method utilized for image denoising) for multi-modality registration.

In this chapter, we concentrate on developing a feature-based metric but also learning features instead of using fixed ones. We use sparse representation coefficients of patches, generated by a dictionary trained offline, to define a similarity measure of alignment. In this chapter, we compare our method with SSD and MI based Free

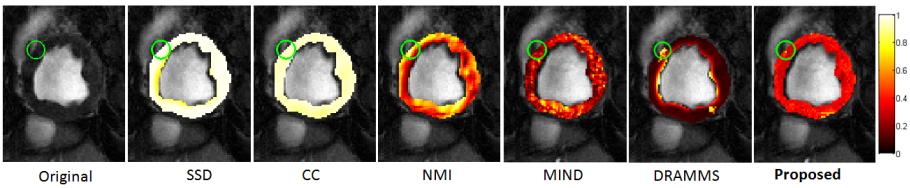

Figure 6.1: Similarity of patches in two consecutive images. First image shows the test patch (green circle) and the remainder shows responses of each similarity metric inside the myocardium. All metrics are normalized and dissimilarity metrics are inverted.

Form Deformations (FFD) [103], optical flow based diffeomorphic demons (ddemons) [124] and symmetric diffeomorphic transformation with CC metric implemented in Advanced Normalization Toolkit (ANTs) [6]. To demonstrate that our proposed approach provides better localized matches, Figure 6.1 shows the values of matches using several criteria when taking a patch from one image and matching it to myocardial locations in another image.

6.2 Dictionary Learning-based Image Descriptor (DLID)

We leverage dictionary learning techniques to learn better representative features. Accordingly, we integrate a Dictionary Learning-based Image Descriptor (DLID) derived from training patches into a similarity term of our proposed registration framework. Features learnt via dictionary learning are used in an image registration framework to evaluate the performance of the proposed descriptor.

6.2.1 Using learnt Features in a Registration Framework

When registering a cardiac sequence $I_1, \ldots, I_t$, we aim to find a deformation that can register each image in the sequence to the first one. Here following the formulation of equation 6.1, we adopt a

regularization in the form of

$$\underset{u}{\operatorname{argmin}} \sum_{p \in \Omega} S(I_1(p), I_t(p+u))^2 + \lambda \operatorname{tr}(\nabla u(p)^T \nabla u(p))^2, \qquad (6.2)$$

where ∇u denotes the gradient of the displacement field. This function is minimized over u with Gauss-Newton optimization as described in [39].

We propose an appropriate similarity term S based on the sparse feature representation of image patches. Assuming two input images, considering I_1 as fixed and I_t as moving, we extract for each pixel location in both, patches, which we represent with appearance and texture features (HoG, Gabor). We create a sparse representation $\hat{X}_p$ for each pixel location for the two images to be registered. The Orthogonal Matching Pursuit (OMP) algorithm [120] is used to compute, two sparse feature matrices $\hat{X}^B$ and $\hat{X}^M$, both with n dimensions, based on previously computed dictionaries D^B and D^M (detailed below). At a certain pixel p of the image, a concatenation of these sparse representation vectors $\hat{X}_p = [\hat{X}^B; \hat{X}^M]$ are used to represent the image instead of the pixel level definitions. The proposed similarity term S at pixel p is defined as the ℓ^1 norm of the difference vector between the sparse representations of the warped source image and the target image as shown in equation 6.3.

$$S(I_1(p), I_t(p+u)) = \| \hat{X}_p^1 - \hat{X}_{p+u}^t \|_1 \qquad (6.3)$$

The proposed metric can take different measures of similarity such as l_2 norm. l_1 norm is utilized due to its efficiency in spacing for sparse but significant inconsistencies between two frames. In the case of CP-BOLD images these inconsistencies are significant and can be better expressed with l_1 norm compared tol_2 norm .

Algorithm 6 Dictionary Learning

Require: Training patches for background and myocardium: Y^B and Y^M

Ensure: Dictionaries for background and myocardium: D^B and D^M

1: **for** C={B,M} **do**
2: Find intra-class Gram matrix G^C and discard atoms with high values
3: Learn dictionary and sparse feature matrix with the K-SVD algorithm

$$\underset{D^C, X^C}{\text{minimize}} \|Y^C - D^C X^C\|_2^2, \text{ subject to } \|x_i^C\|_0 \leq s$$

4: **end for**
5: Compute inter-class Gram matrix G^{BM}
6: Discard from D^B and D^M atoms with high values in G^{BM}

6.2.2 Feature Generation with Discriminative Dictionary Learning

Given some training images (e.g., sequences in the context of cine (BOLD) MRI) and corresponding ground truth labels (i.e., myocardial masks), we obtain two sets of matrices, Y^B and Y^M, where the matrix Y^B contains background information, and Y^M contains information of patches within the myocardium. Information is collected from image patches: $K \times K$ squared patches are sampled around each pixel in the training images. More precisely, the i-th column of the matrix Y^B (and similarly for the matrix Y^M) is obtained by concatenating the normalized patch vector of pixel intensities, taken around the i-th pixel in the background (or myocardium), along with Gabor and HOG features of the same patch. The dictionary learning method takes as input these two sets of training matrices, to learn, two dictionaries, D^B and D^M, with n number of

atoms, and two sparse feature matrices, X^B and X^M, with sparsity s. The i-th column of the matrix X^B, x_i^B, is considered as the discriminative feature vector for the particular pixel corresponding to the i-th column in Y_j^B.

Dictionaries and sparse features are trained via the well-known K-SVD algorithm [2], in an optimization problem shown in Algorithm 1. During initialization we first find the "intra-class Gram matrix" to promote diversity. The idea is to have a subset of patches as much diverse as possible to train dictionaries and sparse features. For a given class considered (let us say background) we can define the intra-class Gram matrix as $G^B = (Y^B)^T Y^B$. To ensure a proper discriminative initialization, patches that correspond to high values in the Gram matrix are discarded from the training before performing K-SVD, and K-SVD is initialized obtaining a random set of patches as initial atoms.

We also use pruning, inspired as a greedy approach of [99], which is performed after K-SVD to remove undesired (similar to other) atoms from each dictionary trained. In this case, an "inter-class Gram matrix" between dictionaries is computed ($G^{BM} = (D^B)^T D^M$), the atoms of each dictionary are sorted according to their cumulative coefficients in G^{BM}, and a chosen percentage of them is discarded to ensure mutual exclusiveness (and better discrimination) between the different dictionaries. These modifications ensure that patches of different origin will have different support and that similar atoms are excluded.

6.3 DLID Results

This section describes qualitatively and compares quantitatively our proposed dictionary learning-based descriptor with state-of-the-art approaches.

6.3.1 Parameter Settings

All quantitative experiments are performed in a strict leave-one-subject-out cross-validation. Parameters and settings were optimized for each method used in comparison. For DLID, in this chapter we have empirically chosen a dictionary of $n = 1000$ atoms for foreground and background respectively, a sparsity of $s = 4$, and as patch size K=9. The regularization weight (λ) is set to 0.8 to ensure smooth deformations.

6.3.2 Visual Evaluation

In an example sequence, we register each image in the sequence throughout the cardiac cycle to the first image using our approach. We take two orthogonal short axis profiles that intersect approximately at the center of the Left Ventricle, and in Figure 6.2 we show the temporal evolution of the profiles with and without registration (left-most and right-most horizontal and vertical profile, respectively). The proposed shows clearly defined structure and the ability to correct for cardiac motion. Notice that BOLD intensity variation is subtle and not perceptible in these images (ie., is not a global change).

6.3.3 Quantitative Comparison

Using again the same process, in a strict-leave-one-out fashion we want to investigate the effect of different similarity metrics in recovering cardiac motion. To evaluate performance, we use again manual delineations of the myocardium provided by experts, and train dictionaries on a set of images and test on one subject. For validation, via segmentation, the myocardial mask from the source image was propagated to the target using the deformation field found with the algorithms, and its overlap with the ground truth mask of the fixed is measured using the Dice overlap metric [88]. Note that these masks are unknown to the algorithms and are used

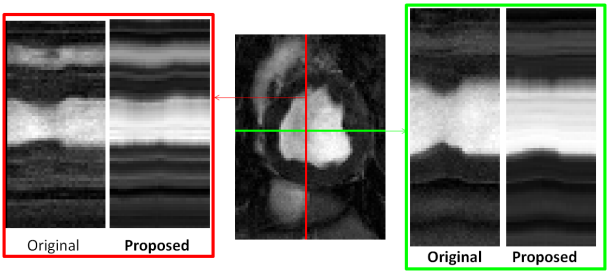

Figure 6.2: Temporal evolution of two orthogonal short axis profiles (red and green line) intersecting approximately at the center of the left ventricle, without registration (original) and with registering every image in the sequence with the first image (proposed).

only for comparison.

Our findings in Table **??**, show that using discriminative features and our similarity term significantly improve the performance for CP-BOLD cardiac sequence registration either under baseline or ischemia conditions w.r.t. other approaches. To highlight the unique challenge of BOLD, we also include results based on standard CINE. Our proposed method, although not its main focus, performs as good as other algorithms even in this case. To emphasize the importance of sparsity and learning we also use directly the ℓ^2 norm between input patches, instead of spare representations. Lower performance in ischemia for all algorithms could be attributed to changes in myocardial contractility.

6.4 Discussion

We propose a new dictionary learning-based image descriptor (DLID) for myocardial registration. The experiments clearly underline the need for a new representation in image registration. Their integration into analytical tools are necessary to meet new challenges posed by myocardial CP-BOLD MR. In particular, this chapter pin-pointed

Table 6.1: Dice overlap comparison (mean $\pm$ std) for registration accuracy in %

	Baseline		Ischemia	
Methods	CINE	BOLD	CINE	BOLD
ANTs (CC metric) [6]	60 ± 11	55 ± 10	55 ± 15	51 ± 12
dDemons [124]	59 ± 11	51 ± 16	58 ± 13	45 ± 13
DRAMMS [87]	67 ± 9	61 ± 7	59 ± 10	54 ± 6
FFD-SSD [103]	49 ± 7	45 ± 16	48 ± 14	39 ± 13
FFD-MI [103]	54 ± 12	48 ± 8	53 ± 6	38 ± 7
MIND [39]	62 ± 7	62 ± 12	61 ± 15	53 ± 9
Proposed w.o. sparsity	55 ± 8	52 ± 11	45 ± 9	42 ± 12
Proposed	63 ± 7	66 ± 9	58 ± 7	60 ± 13

the challenges the BOLD effect poses on common assumptions made when registering the myocardium and quantitatively analysed the performance of the descriptor both under baseline and ischemia conditions. Moreover, in this chapter we showed that by learning appropriate features to best represent texture and appearance in CP-BOLD, it is possible to obtain better correspondences for the entire cardiac sequence. The proposed method can be utilized for other challenges, where spatio-temporal intensity as a biomarker of disease, especially in the presence of motion. One limitation is computational time, since calculating sparse representations is the bottleneck of the problem. The successful application of this post-processing tools are foreseen to be critical in the clinical translation of cardiac CP-BOLD MR.

6.5 Summary

In this part of the thesis, we have proposed novel solutions to the challenges of myocardial segmentation and registration for CP-BOLD MRI. We also have underlined the need for accurate segmentation and registration algorithms in the context of CP-BOLD

MRI with extensive analysis of the data. An accurate segmentation and registration scheme will output accurate timeseries, which will enable pixel-level determination ischemia likelihood in CP-BOLD MRI. In Chapter 4, we highlighted the drawbacks of the literature for the myocardial segmentation and investigated a novel sparse feature learning techniques for myocardial segmentation. In Chapter 5 the motion of the myocardium is used as a distinctive feature in two different contexts. First, the background is modelled with one-class SVM to segment the myocardium. Then motion and appearance is learnt jointly with dictionary learning to extract the myocardium region. In Chapter 6, the utility of similar features in a novel registration setting are explored. In the next part of the thesis, we will investigate techniques to merge two optimization problems.

Part III

Joint Registration and Segmentation

7

Joint Registration and Segmentation

This chapter focuses on using registration and segmentation techniques introduced in Chapters 4 and 5 jointly. Registration and segmentation of anatomical structures are two well studied problems in medical imaging. Optimizing segmentation and registration jointly has been proven to improve results for both challenges. First, we use the external database of training subjects to segment the myocardium and after finishing the segmentation we register the entire sequence to add temporal information to our segmentations. Then we propose a multi-scale scheme to tackle the issues regarding the precision of registration. Finally, we propose to write a cost function with terms for both segmentation and registration to optimize jointly. The experimental results highlight the performance increase

This chapter is partly based on:

- I. Oksuz, R. Dharmakumar S. A. Tsaftaris, "Towards joint segmentation and registration of the myocardium in CP-BOLD MRI at rest", *Society for Cardiovascular Magnetic Resonance Annual Meeting (SCMR)*, 2016.
- I. Oksuz, R. Dharmakumar S. A. Tsaftaris, "Joint Myocardial Registration and Segmentation of Cardiac BOLD MRI", *STACOM*, 2017.

achieved by joint optimization of two tasks.

7.1 Naive Joint Segmentation and Registration

Automated analysis approaches, which can obtain pixel-level determination of ischemia, are desirable since they may lead to improved accuracy in detection of disease. To achieve this, precise segmentation and non-linear registration of the myocardium among the frames (the cardiac phases) in the cine stack would be required. Unfortunately, at present due to BOLD contrast variations, classical approaches to segmentation and registration fail to reach desirable accuracy. In this chapter, we investigate algorithms that jointly finds suitable myocardial segmentation and elastically registers the heart.

7.1.1 Sequential Dictionary Learning based Segmentation and Registration

The main principle of the approach is to use an external initial database of pre-segmented images (and dictionaries for myocardium and background), to first come up with an initial segmentation (relying on classification when projecting on discriminatory dictionaries) of an input CP-BOLD patient stack, then register images in the cardiac cycle (cine acquisition) using the sparse coefficients of the projections to establish a new similarity metric, refining the obtained segmentation, refining the dictionaries (adapting them to the dataset under consideration), till convergence (Figure 7.2). First; 9x9 patches from each image frame is extracted and concatenated for background and myocardium pixels separately. Then Gabor and HOG features are added to generate the vectors Y. Given enriched vectors Y, K-SVD algorithm is utilized to find dictionaries D and a sparse representation X for $Y \approx DX$. Afterwards; orthogonal matching pursuit algorithm is used for dictionaries to calculate the sparse representations. The low representation error among two dictionaries is used for classifying each pixel at the initial seg-

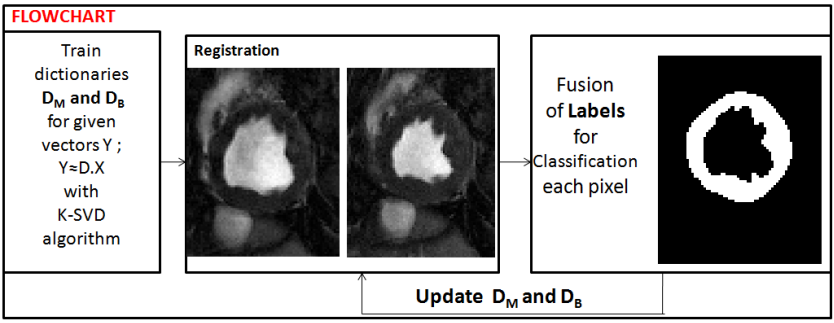

Figure 7.1: Flowchart of sequential registration and segmentation

mentation. The pixels labeled as myocardium and background in this initial segmentation are used for the dictionary learning based image registration algorithm. The registration algorithm relies on supports as X the similarity metric for the registration process. For each image frame It the supports for background X_B and myocardium X_M are calculated from subject specific dictionaries generated from the initial segmentation and supports are concatenated for background and myocardium for each pixel p; $X_p = [X_B; X_M]$. In data term of the objective function of registration E; the similarity metric in between two consecutive frames is defined as L1 norm of the concatenated supports: $S(I_1(p), I_t(p+u)) = \| \hat{X}_p^1 - \hat{X}_{p+u}^t \|_1$ as detailed in Chapter 6. The approach iterates first a refinement step of the segmentation using the registration for the agreement of classification and then finding the D, then X, and the agreement in between different frames is used for classification of the myocardium region.

In an example from a canine the proposed approach identifies the myocardium region accurately (Figure 7.2). As seen in Table 7.1, experiments with different groups of data sets show that the proposed approach obtains superior accuracy when compared to state of the art methods for registration and training the dictionaries on different subjects from the data set.

Table 7.1: Dice Overlap measures of CP-BOLD data for the proposed method compared with training from other subjects and two state of the art methods for image registration

	BASELINE	ISCHEMIA
Proposed Approach	0.68 ± 0.08	0.63 ± 0.06
Proposed without Refinement	0.66 ± 0.09	0.60 ± 0.13
DRAMMS [87]	0.61 ± 0.07	0.54 ± 0.06
MIND [39]	0.62 ± 0.07	0.53 ± 0.09

Figure 7.2: Quantitative Results of sequential registration and segmentation

The experiments clearly underline the need for a new representation of the data in image registration and segmentation. Using the information from the same subject to train the dictionaries improves the accuracy of registration. Although further experiments are necessary to validate this approach, CP-BOLD can open the road to repeatable, truly non-invasive diagnosis of ischemic heart disease.

7.1.2 Multi-Resolution Scheme

The registration schemes are defined on coarse to fine multi-resolution schemes to capture the motion the at different scales. We propose a sequential registration and segmentation scheme to update the

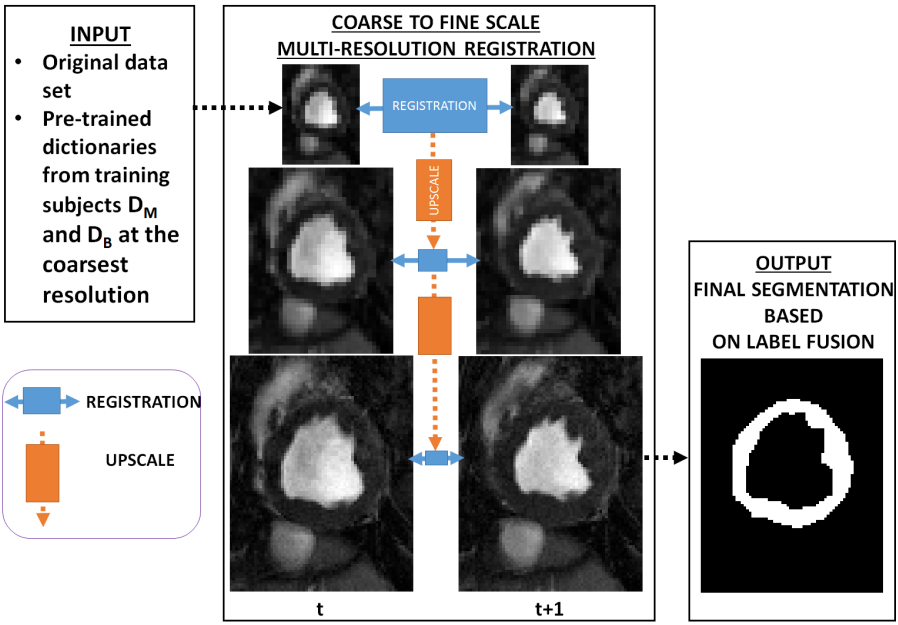

Figure 7.3: Flowchart of the proposed algorithm with a Multi-Resolution Registration Scheme. The images at different cardiac phases are down-sampled and registered at different resolutions (Figure 7.4). The transformations from the coarser scale are used as guidance to update the dictionaries for the finer scale registration process in dictionary learning based registration (Figure 7.5).

dictionaries

The main principle of the approach (Figure 7.3) is to use an external initial database of pre-segmented images (and dictionaries for myocardium and background) to first come up with an initial segmentation (relying on classification when projecting on discriminatory dictionaries) for an unseen CP-BOLD dataset (cine acquisition). Then it registers images in a multi-resolution fashion across the cardiac cycle using a registration algorithm that relies on the sparse coefficients. This step refines the obtained segmentation, which is

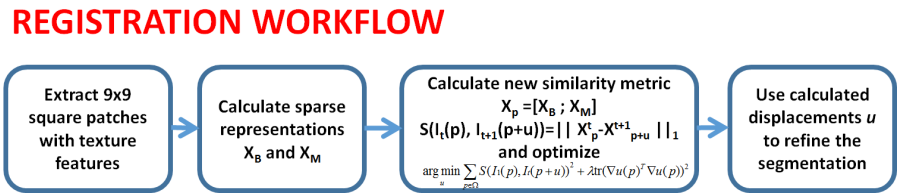

Figure 7.4: Dictionary learning based registration. At every resolution patches are extracted from fixed and moving images to calculate the sparse representations. Dictionary learning based similarity metric is used in an optimization framework to find displacements. These displacements are used to update the dictionaries for the next resolution.

used to update the dictionaries (adapting them to the dataset under consideration), and thus personalize the process. This process is repeated till convergence. To obtain an initial dictionary on prior available segmented data sets, at the coarsest resolution, first 9×9 patches are extracted for each pixel and concatenated with HOG and Gabor features for each patch and dictionaries D_B and D_B are learned similar to the approach in Chapter 4.

Multi-Resolution Registration Refinement is achieved by the proposed method as illustrated in Figure 7.3. For each pixel p when projected on the background or myocardium dictionaries sparse coefficients X_B and myocardium X_M are obtained and concatenated; $X_p = [X_B; X_M]$. They are used to define a similarity metric between two images as $S(I_t(p), I_{t+1}(p+u)) = \| X_t(p) - I_{t-1}(p+u) \|_1$. Starting at the coarsest resolution, after all images in the cardiac cycle have been registered, individual per-image segmentations of the myocardium are obtained and propagated to the first image. Then labels are fused (via majority voting) to refine the segmentation of that image. The segmentation is then upscaled and using the corresponding intensity image, new dictionaries $D^{'}_B$ and $D^{'}_M$ specific to this resolution and image are obtained (Figure 7.5). The process iterates using these new dictionaries and sparse representations. At

Figure 7.5: Upscaling and dictionary update. At every resolution the calculated displacements are used to generate label fusion from multiple cardiac phases on the fixed image. Then, these label maps are upscaled and used to obtain new dictionaries specific to this unseen image in the new resolution.

the finest resolution all images I_{t+j}, $j \neq 0$ are registered to I_t, segmentations obtained via the dictionaries for $I_{t+j}$are propagated to I_t and fused with majority voting to obtain the final segmentation.

To evaluate segmentation accuracy, delineations (ground truth) provided by experts are used. For a given unseen dataset, overlap of the segmentation of It with corresponding ground truth is measured using the Dice overlap metric, and averaged across all t. Multiple paired t-tests were applied to evaluate performance when compared to state-of-the-art registration methods, namely Diffeomorphic Demons [124], Free Form Deformations (FFD) [103], DRAMMS [87] and MIND [39]. Dice accuracies are presented in Table 7.2, across our study population. Utilizing the proposed scheme increases myocardial segmentation accuracy under baseline (, p<0.001) and ischemia ($, p<0.001) significantly. The scheme herein improves myocardial segmentation accuracy in a statistically significant manner by refining and adapting the dictionary to the data using multiple resolutions. When only an external dictionary is used the resulting performance is not different to other methods. However, taking advantage of registration to refine the segmentation results leads to an improvement by $\%8$. Operating at multiple resolutions (capturing more context and spatial information and

Table 7.2: Multi-resolution sequential registration and segmentation results. Dice overlap comparison (mean∓std) of the CP-BOLD myocardial segmentation accuracy.

METHODS	**BASELINE**	*ISCHEMIA*
Refined Segmentation	0.68 ∓ 0.08	0.63 ± 0.06
Initial Segmentation	0.66 ± 0.09	0.60 ± 0.13
Diffeomorphic Demons	0.63 ± 0.08	0.53 ± 0.08
FFD-MI	0.62 ± 0.07	0.51 ± 0.09
DRAMMS	0.61 ± 0.07	0.54 ± 0.06
MIND	0.62 ± 0.07	0.53 ± 0.09
Proposed Approach	$0.72 \pm 0.07^{\#}$	$0.68 \pm 0.08^{\$}$

avoiding local minima) leads to even better improvement ($\%15$).

7.2 Simultaneous Segmentation and Registration

In this section, we propose a joint optimization scheme for registration and segmentation using dictionary learning based descriptors. Our joint registration and segmentation aims to solve an optimization function, which enables better performance for both of these ill-posed processes. We build two dictionaries for background and myocardium for square patches extracted from training images. Based on dictionary learning residuals and sparse representations defined on pre-trained dictionaries, a Markov Random Field (MRF) based joint optimization scheme is built. We propose to update the dictionaries based on the segmentation label changes of the joint optimization. The accuracy of the proposed method is illustrated on challenging Cardiac Phase-resolved Blood Oxygen-Level-Dependent MRI dataset. Precise segmentation and non-linear registration of the myocardium among the frames (the cardiac phases) in the cine stack would be required to achieve accurate timeseries.

Our proposed joint registration and segmentation scheme to generate accurate timeseries information for cardiac sequence. We

adopt a joint optimization scheme [69] to optimize the registration term on sparse representations and segmentation terms for dictionary learning residuals. The motivation behind this choice is the mutual benefit of the both functions, which can be directly translated to accurate registration and segmentation.

There are two contributions of this two sections. First, we define a joint optimization scheme based on dictionary learning residuals and sparse representations for the first time. Moreover, we introduce an iterative dictionary update stage, which takes the spatial smoothness into account to boost discriminative power of the dictionary learning structure. With this, the dictionaries are ensured to be subject-specific and more robust for classifying the myocardium.

7.2.1 Background

Registration and segmentation of organs in medical imaging are two major tasks, which are processed with two independent optimization schemes in most applications. One approach of solving both problems is using a sequential strategy to address both challenges, which results in concatenation of errors of both processes. Instead of a sequential segmentation and registration scheme, which uses the estimated solution one sub-problem as a prior knowledge to the other, joint optimization of two problems can be defined [137], where both problems are solved simultaneously. Early works merged the two processes with partial differential equations [130] and and in particular within level-set formulations [142]. More recent literature relies on joint optimization with single function simultaneously using Markov Random Fields (MRF)s [33]. MRFs are suitable for discrete labeling problems and the labels are defined as segmentation classes and discrete displacement vectors. The concept of utilizing mutual benefits between the registration and segmentation has been studied for the the problem of atlas-based tumor segmentation for brain MRI [92]. [4] proposes to couple segmentation and registration scheme for classifying multiple regions in brain MRI. Mahapatra et al. [68] used a joint optimization scheme

to detect the left ventricle (LV) in standard cine and perfusion MR images.

In joint registration and segmentation, the estimate of one set of parameters of registration should not adversely affect parameters of segmentation. An appropriate optimization scheme aims to balance these influences. Graph cuts is based on maximum-flow approach and is very effective in finding the global minimum or a strong local minimum of discrete MRF energy formulations [13]. However, a number of issues have to be addressed in using segmentation information for MRF-based registration. Registration and segmentation energies have to be combined such that there is no bias for a particular term. The mutual dependence of registration and segmentation has to be factored in the objective function.

In this section, we propose a joint optimization scheme for myocardial registration and segmentation to generate accurate deformations and segmentation masks for the entire cine stack. Our method builds upon the externally trained dictionaries of myocardium and background and uses priors on each problem jointly to extract and register the myocardial region. We introduce a dictionary update scheme to fuse subject-specific local information. Our algorithm generates deformations and segmentations for the entire cardiac sequence of Cardiac Phase-resolved Blood Oxygen-Level-Dependent (CP-BOLD) MRI.

7.2.2 Methods

The details of our method is visualized in Figure 7.6. We extract region of interest around LV blood pool using a similar preprocessing strategy to [97]. We rely on externally trained dictionaries of myocardium and background to define registration and segmentation terms for joint optimization. Then, these terms are optimized using a graphical model. Finally, we update our dictionaries to enrich subject-specific information in the dictionaries.

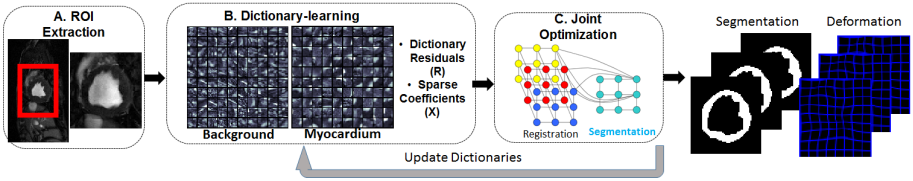

Figure 7.6: Algorithm design for joint segmentation and registration. Region of interest extraction (Panel A). Dictionary learning from training images and calculation of residuals (R) and sparse coefficients(X) (Panel B). Multi-resolution deformation grids and exemplary connections with segmentation grid (Panel C)

Dictionary Learning based Image Segmentation

Dictionary learning based approaches have been used for segmentation of medical images [42]. In our specific algorithmic design, given some sequences of training images and corresponding ground truth labels, we can obtain two sets of matrices, Y^B and Y^M, where the matrix Y^B contains the background information and Y^M is the corresponding matrix referring to the myocardium. Squared patches are sampled around each pixel of the training images from both regions. The i-th column of the matrix Y^B (and similarly for the matrix Y^M) is obtained by concatenating the normalized patch vector of pixel intensities, taken around the i-th pixel in the background, along with the Gabor and HOG features of the same patch. The method detailed in [76] trains two dictionaries, D_k^B and D_k^M, and two sparse feature matrices, X_k^B and X_k^M using the K-SVD algorithm [2] for each class $C = \{B, M\}$:

$$\underset{D^C, X^C}{\text{minimize}} \|Y^C - D^C X^C\|_2^2, \text{ subject to } \|x_i^C\|_0 \leq \text{sparsity}$$

After the training given a new subject, a certain patch will be assigned to the class that gives the smallest dictionary approximation error using Orthogonal Matching Pursuit [120]. If $R_B = \|\hat{y}_i - D^B \hat{x}_i^B\|_2$ is less than $R_M = \|\hat{y}_i - D^M \hat{x}_i^M\|_2$, the patch is assigned

to the background; otherwise, it is considered belonging to the myocardial region.

Graph-based Joint Optimization

In this section, we introduce our dictionary learning based joint optimization scheme for registration and segmentation of the myocardium. The general term for energy of a second-order MRF is defined as:

$$E(L) = \sum_{p \in \Omega} D_p(l_p) + \lambda \sum_{p,q \in N} V_{pq}(l_p, l_q)$$

where p and q denote the pixels, l_p and l_q denotes the registration and segmentation labels of the pixels p and q. λ controls the interaction between data term and smoothness term. The function is optimized over the labels $L = \{C, u\}$, which consists of the segmentation label C and discrete deformation u. We define the general data term $D_p(l_p)$ similar to [68]:

$$D_p(l_p) = D^1_{l_p} + \gamma D^2_{l_p}$$

which consists of two terms, namely segmentation and registration data terms. Segmentation of the myocardium is defined over the dictionary learning residuals R_B and R_M. The penalty of the pixel p to be classified as myocardium is : $\kappa_M(p) = \frac{R_M(p)}{R_M(p)+R_B(p)}$. Similarly, the penalty for the same pixel to be classified as background is $\kappa_B(p) = \frac{R_B(p)}{R_M(p)+R_B(p)}$. Using these penalty definitions D^1_p is defined as:

$$D^1_{l_p} = \begin{cases} \sqrt{\kappa^r_M(p) * \kappa^f_M(p+u)}, & \text{if } C^r(p) = C^f(p+u) = M \\ \sqrt{\kappa^r_B(p) * \kappa^f_B(p+u)}, & \text{if } C^r(p) = C^f(p+u) = B \\ \sqrt{\kappa^r_B(p) * \kappa^f_B(p+u)} + \sqrt{\kappa^t_M(r) * \kappa^f_M(p+u)}, & \text{otherwise} \end{cases}$$

where $\kappa_M^f(p+u)$ corresponds to the penalty associated with myocardium class for the deformed floating image with displacement u. Similarly, $\kappa_B^r(p)$ corresponds to penalty of the reference image for the background class. This term ensures a low penalty for same labels of the displaced image and the reference image. If the floating image and the reference image do favor different segmentation classes the penalty will be high.

The registration penalty term $D_{l_p}^2$ of the data term D_{l_p} is defined as:

$$D_{l_p}^2 = \begin{cases} \parallel X_M^r(p) - X_M^f(p+u) \parallel_1, & \text{if } C^r(p) = C^f(p+u) = M \\ \parallel X_B^r(p) - X_B^f(p+u) \parallel_1, & \text{if } C^r(p) = C^f(p+u) = B \\ \parallel X_M^r(p) - X_M^f(p+u) \parallel_1 + \parallel X_B^r(p) - X_B^f(p+u) \parallel_1, & \text{otherwise} \end{cases}$$

where $X_M^r(p)$ corresponds to the sparse representation defined for D^M for the reference image and $X_M^f(p+u)$ defines sparse representation defined for the floating image at location $p+u$. This penalty is increased for dissimilar representation and also for the points with different segmentation labels.

Regularization term ensures the smoothness of segmentation labels and deformation field. The term favors the same segmentation labels in local neighborhoods and smooth deformations. The regularization term is defined as:

$$V_{pq}(l_p, l_q) = \begin{cases} 1, & \text{if } (C_p = C_q \text{ and } \parallel u_p - u_q \parallel \leq \varepsilon) \\ 1, & \text{if } (C_p \neq C_q \text{ and } \parallel u_p - u_q \parallel \leq \tau) \\ 100, & \text{otherwise} \end{cases}$$

where ε and τ restrict high displacements for local neighborhoods when segmentation labels agree or disagree respectively. To optimize the energy functional $E(L)$, we use graph cuts [13] on discrete labels of registration and segmentation.

Dictionary Update

We propose a dictionary update, which refines the dictionaries to inject subject-specific information. After every run of the MRF-based optimization scheme the estimated segmentation labels C are subject to change. We only extract patches that are corresponding to the points of label changes to update our dictionaries. We add square patches Y_u concatanated with Gabor and HOG features and train our dictionaries with Online Dictionary Learning (ODL) algorithm [71], which uses mini-batches to update the dictionaries. We add the new patches with changed labels for updating dictionaries we trained before. During the update the dictionary learning is initialized with the pre-trained dictionaries and this approach improves the discriminative power of the dictionaries in the next iteration.

7.2.3 Experimental Results

This section offers a qualitative analysis and quantitative comparison of our proposed method w.r.t. state-of-the-art methods, to demonstrate its effectiveness for myocardial segmentation and registration. Note that our method outperforms all supervised methods from current literature in both baseline and ischemia cases.

Implementation Details

2D short-axis images of the whole cardiac cycle were acquired at baseline and severe ischemia (inflicted as stenosis of the left-anterior descending coronary artery (LAD)) on a 1.5T Espree (Siemens Healthcare) in the same 10 canines along mid ventricle using both standard CINE and a flow and motion compensated CP-BOLD acquisition within few minutes of each other. The image resolution is 192×114 and each cardiac cycle has 25 frames approximately. We have utilized a strict leave one out cross validation experiment, where the patch size is defined as 11×11, dictionary size as 100

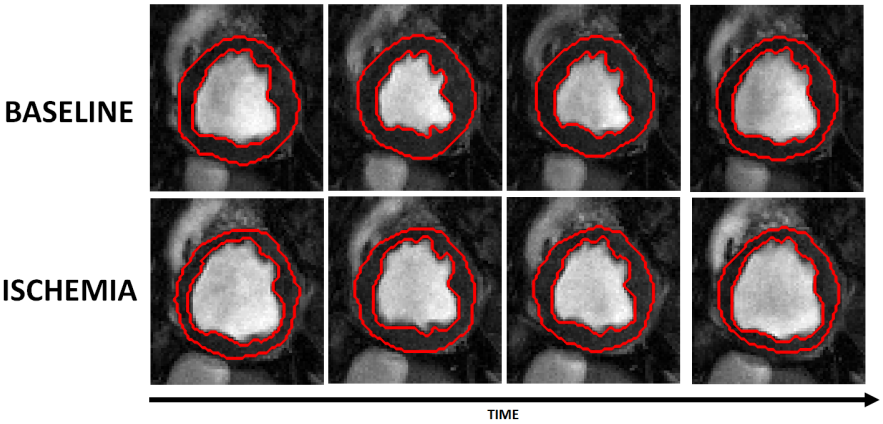

Figure 7.7: Generated Segmentation masks with proposed approach for an exemplary CP-BOLD subject under baseline and ischemia conditions.

and sparsity threshold as 8. The parameters of deformation $\varepsilon = \sqrt{2}$ and $\tau = 3$ are optimized to ensure smooth labels for deformations. $\gamma = 0.7$ gave the optimal contribution of the data terms and $\lambda = 0.9$ ensures the balance of regularization and data terms. We have utilized three scales from coarse to fine for registration. The influence of the control points on each pixel is calculated using cubic B-Splines [34]. The displacement ranges from 2 to 6 pixels.

Visual Evaluation

We demonstrate a set of contours in Figure 7.7 and a deformation grid in Figure 7.8 to highlight the performance of our joint optimization and registration framework. We visualize the deformation grid in between the end systole and end diastole for an exemplary subject under baseline condition. We also illustrate the segmentation and deformation results of CRS [69] compared with our algorithm. Our method generates smooth deformation fields and smooth contours compared to CRS [69].

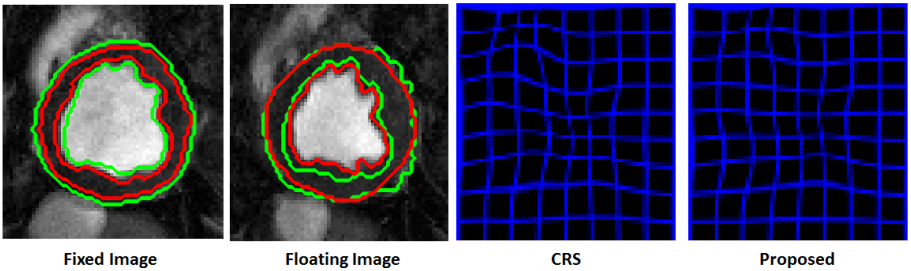

Figure 7.8: Segmentation masks (red contours) and registration grid of proposed approach compared to CRS [69] (green contours) for an exemplary subject under baseline conditions in between end diastole and end systole frames.

Quantitative Comparison

Table 7.3 summarizes our results for Dice overlap measure for myocardium. We compare our algorithm with an atlas-based segmentation technique, which relies on discrete registration performance using mutual information as a similarity metric [34]. Moreover, we include a recent joint registration and segmentation scheme CRS [69], which relies on sum of squared distances and edge-based differences as similarity term for registration. We generate results based on dictionary residuals for each pixel just for segmentation (Segmentation only). In addition, we used a sequential segmentation and registration (Sequential Seg. and Reg.), which first segments myocardium based on residuals and then refines the contours with propagation of the contours via registration based on sparse representations. Finally, we generate a variant of our algorithm, without using the dictionary update (Proposed w.o. update) to highlight the performance increase.

The proposed method outperforms all variants and other techniques in all four datasets. Segmentation information alone shows low performance compared to the variants, which incorporate registration. The sequential segmentation and registration has low per-

Table 7.3: Dice overlap comparison of myocardial segmentation in % for Dice accuracy

	Baseline		Ischemia	
Methods	Cine	BOLD	Cine	BOLD
Atlas-based [34]	60 ± 3	54 ± 8	54 ± 8	45 ± 6
CRS [69]	71 ± 6	70 ± 6	69 ± 5	68 ± 7
Segmentation only	70 ± 5	71 ± 4	69 ± 4	68 ± 4
Sequential Seg. and Reg.	74 ± 6	72 ± 7	71 ± 7	68 ± 8
Proposed w.o update	75 ± 4	76 ± 4	75 ± 5	74 ± 4
Proposed	79 ± 4	79 ± 3	78 ± 5	78 ± 3

formance compared to the proposed method. This low performance is due to the mutual dependence of registration and segmentation that has not been factored in the objective function, which is ensured with the proposed approach. Our method is superior to CRS [69], which relies on edge-based terms for myocardial registration. Ischemia condition generates a slight decrease in the performance for all methods. The proposed dictionary update enables a performance improvement thanks to less coherent dictionaries. The coherence of two dictionaries is calculated before and after the single dictionary update. The average coherence of two dictionaries 0.850 is reduced to 0.780 with the update. We illustrate an example set of dictionaries before and after the update in Figure 7.9.

7.2.4 CAP Dataset

To demonstrate that our method works also non-BOLD, clinical data, we have tested our algorithm on cine cardiac training data set from the MICCAI 2013 SATA Segmentation Challenge. The dataset is part of the Cardiac Atlas Project (CAP) [32] and consists of 83 subjects with a varying in plane resolution from 0.7 mm to 2mm and a varying range of 19 to 30 frames per subject. On mid-ventricular level, we train our algorithm on 30 subjects to learn dictionaries

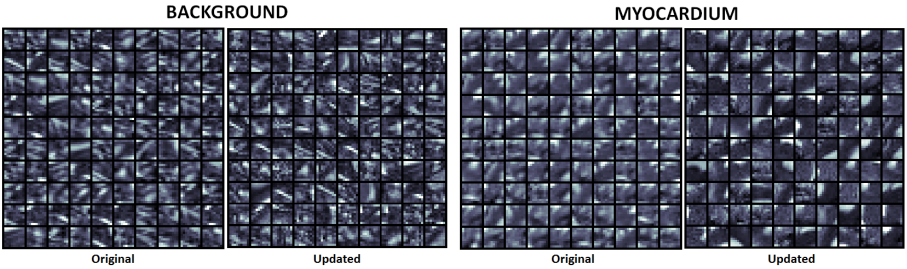

Figure 7.9: Background and foreground dictionaries before and after the dictionary update. Observe the increased number of unique myocardial patterns after the dictionary update.

for background and myocardium. Then, we test on the remaining 53 subjects and we achieve a dice score of 0.81 ± 0.04 compared to 0.80 ± 0.05 of CRS [69] algorithm (where standard deviation refers to variation among subjects and not on leave one cross validation).

7.3 Discussion

In this section, we have investigated the performance joint registration and segmentation scheme for dictionary learning based approach in two different settings. We first concatanate the steps of segmentation and registration one after each other to see the ifluence on the general performance. Then, we proposed a more involved framework to merge two optimization tasks in a single cost function. Our algorithm uses a MRF-based optimization scheme defined on dictionary learning residuals and at each iteration the dictionaries are updated using patches corresponding to the points that changed segmentation labels. This not only boosts the performance by introducing subject specific information, but also adds more discriminative power as showcased with experiments. In future, an interesting direction for further interest is the use of similar joint segmentation registration schemes on different challenges that

involve spatio-temporal information, such as perfusion images and fMRI.

8 Conclusions

8.1 Thesis Summary

The thesis focused on the development of automated image segmentation and registration approaches using patch-based dictionary learning techniques. First a supervised automated segmentation framework that exploits sparse coding techniques is developed. In Chapter 4, we defined features for myocardium and background sparsely using dictionary learning. We use intra- and inter-class Gram filtering to generate mutually exclusive dictionaries using a greedy and low complexity approach. Then, these features are used at multiple scales for myocardial segmentation. In Chapter 5 we use motion as a key feature to guide the segmentation process with two different approaches. The first method (UMSS) models the background pixels using a joint appearance and motion dictionary learning framework. We utilize one-clase SVM to segment the myocardium in a unsupervised fashion. The second method extracts a region of interest, smooths the data and concatenates motion information with appearance in a dictionary learning setting. We also model appearance and motion jointly in a dictionary learning framework. Local smoothness is enforced to the segmen-

tation, which is generated using two separate dictionaries using rudimentary segmentations of background and myocardium. In Chapter 6 we defined a dictionary learning based image descriptor as a similarity metric to establish pixel to pixel correspondences. Our similarity metric is capable to establish better correspondences compared to state of the art similarity metrics. Finally, in Chapter 7 we propose sequential and simultaneous registration and segmentation schemes to address both challenges together. Our framework relies on dictionary updates to add subject-specific patterns into dictionary learning structure, which is proven to improve the performance. In comparison with the state-of-the-art segmentation techniques a significant improvement was observed by using the sparse coding technique, suggesting that the sparsity property of patches can be helpful for segmentation task.

8.2 Research Findings

We have addressed the research questions in the thesis as follows:

Research Question 1

Does BOLD effect challenge segmentation and registration?

Automated myocardial segmentation and registration algorithms developed for standard CINE MR under-perform in CP-BOLD. This is due to the spatio-temporal intensity variations of the myocardial BOLD effect [106], an example of which is shown in Figure 3.4 in cine type acquisition of CP-BOLD. Thus, in CP-BOLD in addition to violations of shape invariance (as with standard CINE MRI) the principal assumption of appearance invariance (consistent intensity) is violated as well. This question is addressed in Chapters 4, 5 and 6.

Research Question 2

How to obtain features to successfully align and extract the myocardium region?

This question is addressed in Chapter 4. In particular, we used a low complexity and a greedy dictionary learning approach to generate two sets of dictionaries from the training images. We concate-

nated texture to appearance features at each patch to characterize the information of the center pixel. This feature representation adds not only contextual information to each pixel it also helps to better distinguish each pixel for segmentation and registration.

Research Question 3

How to segment the myocardium using the data-driven features?

We have used the learnt features for multiple scales to classify the each pixel of the image with MSDDL method in Chapter 4. We learn the dictionaries offline from the training using the appearance and texture features. We capture information at multiple scales to segment the myocardium.

Research Question 4

How to incorporate motion information for myocardial segmentation?

We use motion in two different contexts to aid the segmentation in this thesis. In Chapter 5, we utilized an optical flow based preprocessing to model the background pixels to learn a class-specific dictionary. In this context, we only model one class and classify the myocardium as the remaining pixels using one-class SVM. In our second method, we rely on a region of interest extraction and generate rudimentary classes for background and myocardium. We concatenated the motion to appearance to train two separate dictionaries. Finally, we added local smoothness with a MRF scheme. The incorporation of motion has shown a significant performance increase for segmentation accuracy.

Research Question 5

How to produce a robust measure of similarity for registering cardiac phases?

A new similarity metric, dictionary learning based image descriptor (DLID) is defined in Chapter 6. With this metric, we adrress the low specificity showed by the state-of-the-art similarity metrics for the task cardiac image registration. We illustrated the superior performance of our technique, which based on the sparse representations based on trained dictionaries.

Research Question 6

How to optimize registration and segmentation cost functions jointly?

In Chapter 7, we proposed a sequential and a joint optimization and registration scheme to merge the efforts to minimize two cost functions. The goal of accurately extracting and registering myocardium problem is addressed jointly with sequential and simultaneous setups.

8.3 Limitations

The work presented in Chapters 4, 5, 6 and 7 has been focused on the image registration and segmentation using patches. Although promising results have been achieved, there are some limitations in the current registration and segmentation approaches, which should be further investigated in future work. These limitations are as follows:

- **3D Extension.** Our work in this thesis is limited to developing registration and segmentation approaches in 2D+time MR images. The use of third dimension will provide valuable information for registration and segmentation. Unfortunately to the high flip angle, low TR, and flow compensating gradients used to obtain BOLD contrast with suppressed flow artefacts, breath-hold acquisition is long and we cannot acquire a full 3D volume thus current datasets are 2D(+time). Once advancements in 3D BOLD acquisition are possible with good image quality we will definitely extend our work to 3D.
- **The selection of important patches.** Patches which may be associated with the pathological changes of myocardium were selected for classification in our work. Selection of important patches were determined by using feature selection techniques. A spatial distance threshold may also be defined to control the overlap between patches. The introduction of the spatial distance threshold may not provide an optimal strategy for selecting important patches. Future work will focus

on developing more advanced patch selection approaches as this is an important improvement for our work.

- **The level of registration.** In our work, non-rigid registration with a final control point spacing of 1.25 mm was used for aligning images and the classification results show that the proposed method can work efficiently at this level of registration. The major reason is that the extracted features (intensities within patches) are well aligned for comparison when non-rigid registrations are adopted. If only affine registration is used, one patch may be extracted inside the myocardium of an image under study, while the corresponding patch at the same location in another image may be extracted far away from the myocardium. In this case, the comparison between these patches may not provide useful information for classification as it does not compare corresponding anatomical locations. Although non-rigid registration can ensure that the corresponding structures are well aligned, part of the differences due to pathology may also be removed at the same time. We believe that there is a trade-off between the level of non-rigid alignment and the amount of detectable pathological changes, which can still allow us to measure subject-specific differences for classification. The trade-off of the control point spacing and the regularization should be further investigated. The idea of using a scale-space representation of this trade-off represents another interesting approach as this may provide additionally useful information for classification.
- **Computational Time.** A natural limitation of joint registration and segmentation framework is longer processing times compared to standard segmentation techniques. Our joint registration and segmentation algorithm has an average execution time of approximately twenty minutes on a 2.4 GHz processor with an average data set ($192 \times 114 \times 30$). Most of this time is spent on the registration stage.

8.4 Future Work

In this part, we will address the possible extensions of the current techniques that are utilized in this thesis. Some of these gaps imply open research challenges in the area. Therefore, further research efforts might consider the following aspects.

One issue of interest is promoting the discrimination in our dictionary learning framework. This can be achieved by promoting more diverse atoms in our dictionary thanks to a refinement step designed for a particular task. Moreover, the dictionary learning strategies suffer from the difficulty of seeking a balance of reconstructive and discriminative abilities of the learned dictionary. One interesting approach is proposed by Wang et al. [132]. In their work they propose to learn a shared dictionary in addition to the target and background specific dictionaries for robust visual tracking. With the shared dictionary modelling the commonality between the target and background, and specific dictionaries capturing the difference, their learned dictionary is both reconstructive and discriminative which can better distinguish the target from the background. This approach is particularly interesting for our task; because most of our misclassifications are at the border areas of myocardium, where we need more ambiguity in our defined dictionaries for background and myocardium respectively.

Secondly, we would like to investigate new methodologies to enforce further spatial information into our framework. One approach we would like to consider for this challenge is to train Hough Forests [44] for implicitly adding spatial information to defined dictionaries. Another issue of interest is the smart dictionary update stage, where we propose to prone the atoms according to the contribution to the classification in Chapters 4 and 5. One interesting technique of interest is spatial constrains on dictionaries and spatial pruning of atoms. The definition of radial constraints on sparse representations is an interesting application for segmentation of the myocardium.

Another are of further investigation this thesis carved the path to is disease detection. With the available tools for registration and segmentation disease detection algorithms can be coupled. One specific idea is to use motion information alongside the motion patterns to identify ischemic regions.

Finally, we would like to address the computational cost of the current framework. For defining faster sparse representations; one idea is to replace Orthogonal Matching Algorithm (OMP) with a more efficient algorithm. Gregor and Lecun [36] proposed a method for learning fast approximation of sparse coding. Furthermore, we will investigate the approach proposed in [104] to design stage-wise K-SVD. This algorithm designs atoms according the energy levels of the atoms in the dictionaries. Another techniques to consider is the efficient sparse coding algorithms developed by Lee et al. [54]. In their work, they propose fast algorithms for solving two general-purpose convex problems. First, L1-regularized Least Squares problem solver using the feature-sign search algorithm and second L2-constrained Least Squares problem solver using Lagrange dual. It is important to reduce the run time of the approaches to milliseconds for them to be a part of the every day clinical practice.

Bibliography

[1] Adluru et al., Automatic Segmentation of Cardiac Short Axis Slices in Perfusion, ISBI, pp. 133-136, 2006.

[2] Aharon et al., K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation, IEEE TSP 54 (11), pp. 4311-4322, 2006.

[3] Alchatzidis et al., A discrete MRF Framework for Integrated Multi-Atlas Registration and Segmentation, IJCV 121(1), pp. 169-181, 2017.

[4] Alchatzidis et al., Discrete Multi Atlas Segmentation using Agreement Constraints, BMVC, 2014.

[5] Alessandrini et al., Myocardial Motion Estimation from Medical Images Using the Monogenic Signal, IEEE TIP, 22(3):1084-1095 2013.

[6] Avants et al., Symmetric Diffeomorphic Image Registration with Cross-Correlation: Evaluating Automated Labeling of Elderly and Neurodegenerative Brain, MedIA 12(1), pp. 26-41, 2008.

[7] Avendi et al., A Combined Deep-learning and Deformable-model Approach to Fully Automatic Segmentation of the Left Ventricle in Cardiac MRI, MedIA 12(1), pp. 26-41, 2016.

[8] Bai et al., A Pobabilistic Patch-based Label Dusion Model for Multi-Atlas Segmentation with Registration Refinement: Application to Cardiac MR Images, IEEE TMI 32(7), pp. 1302-15, 2013.

[9] Bai et al., Multi-atlas Segmentation with Augmented Features for Cardiac MR Images, MedIA 19(1), pp. 98-109, 2015.
[10] Benkarim et al., Label Consistent Multiclass Discriminative Dictionary Learning for MRI Segmentation, Articulated Motion and Deformable Objects, pp. 138-147, 2014.
[11] Besbes et al., Graph-based Knowledge-driven Discrete Segmentation of the Left Ventricle, ISBI, pp. 49-52, 2009.
[12] Bevilacqua et al., Dictionary-Driven Ischemia Detection From Cardiac Phase-Resolved Myocardial BOLD MRI at Rest, IEEE TMI 35(1), pp. 282-293, 2016.
[13] Boykov et al., Fast Approximate Energy Minimization via Graph Cuts, IEEE PAMI 23(11), pp. 1222-1239, 2001.
[14] Brox et al., High Accuracy Optical Flow Estimation based on a Theory for Warping, ECCV, pp. 25-36, 2004.
[15] Cai et al., Unsupervised Freeview Groupwise Cardiac Segmentation Using Synchronized Spectral Network, IEEE TMI 35(9), pp. 2174-2188, 2016.
[16] Cardiovascular Diseases World Health Organization, Fact Sheet No.317, 2015.
[17] Cerqueira et al., Standardized Myocardial Segmentation and Nomenclature for Tomographic Imaging of the Heart, Circulation 105(4), pp. 539-542, 2002.
[18] Chan et. al., Active Contours Without Edges, IEEE TIP 10(2), pp. 266-277, 2001.
[19] Chan et. al., An Augmented Lagrangian Method for Total Variation Video Restoration, IEEE TIP 20(11), pp. 3097-3111, 2011.
[20] Chang et. al., Achievable Angles between two Compressed Sparse Vectors under Norm/Distance Constraints imposed by the Restricted Isometry Property: A Plane Geometry Approach, IEEE T. Information Theory 59(4), pp. 2059-2081, 2013.
[21] Chen et. al., Semiautomated Segmentation of Myocardial Contours for Fast Strain Analysis in Cine Displacement-Encoded MRI, IEEE TMI 27(8),pp. 1084-1094, 2008.
[22] Cicek et al., 3d u-net: Learning Dense Volumetric Segmentation from Sparse Annotation, MICCAI, pp. 424-432, 2016.
[23] Cordero-Grande et al., Unsupervised 4D Myocardium Segmentation with a Markov Random Field based Deformable Model, MedIA 15(3), pp. 283-301, 2011.

[24] Cortes et. al., Support-vector Networks, Machine Learning, 20(3),pp. 273-297, 1995.
[25] Constantinides et al., Fully Automated Segmentation of the Left Ventricle Applied to Cine MR images: Description and Results on a Database of 45 Subjects, IEEE EMBC, pp. 3207-3210, 2012.
[26] De Craene et al., Temporal Diffeomorphic Free-Form Deformation: Application to Motion and Strain Estimation from 3D Echocardiography, MedIA 16(2), pp. 427-450,2012.
[27] Dharmakumar et al., Assessment of Regional Myocardial Oxygenation Changes in the Presence of Coronary Artery Stenosis with Balanced SSFP Imaging at 3.0T: Theory and Experimental Evaluation in Canines, JMRI 27(5), pp. 1037-1045, 2008.
[28] Ehrhardt et al., A Variational Approach for Combined Segmentation and Estimation of Respiratory Motion in Temporal Image Sequences, IEEE ICCV, pp. 1-7, 2007.
[29] El Aidi et al.,Cardiac Magnetic Resonance Imaging Findings and the Risk of Cardiovascular Events in Patients With Recent Myocardial Infarction or Suspected or Known Coronary Artery Disease: A Systematic Review of Prognostic Studies, JACC 63(11), pp. 1031-1045, 2014.
[30] Engan et al., Method of Optimal Directions for Frame Design, IEEE ICASSP, pp. 2443-2446.
[31] Eslami et al., Segmentation by Retrieval with Guided Random Walks: Application to Left Ventricle Segmentation in MRI, MedIA 17(2), pp. 236-253, 2013.
[32] Fonseca et al., The Cardiac Atlas Project—An Imaging Database for Computational Modeling and Statistical Atlases of the Heart, Bioinformatics, 27(16): 2288-2295, 2011.
[33] Glocker et al., Simultaneous Segmentation and Multiresolution Nonrigid Atlas Registration, IEEE TIP 23(7), pp. 2931-2943, 2014.
[34] Glocker et al., Dense Image Registration Through MRFs and Efficient Linear Programming, MedIA 12(6), pp. 731-741, 2008.
[35] Gooya et al., Joint Segmentation and Deformable Registration of Brain Scans Guided by a Tumor Growth Model, MICCAI, pp. 532-540, 2011.
[36] Gregor et al.,Learning Fast Approximations of Sparse Coding, ICML, 12(6), pp. 399-406, 2010.
[37] Gupta et al., Cardiac MR Perfusion Image Processing techniques: A Survey, MedIA 16(4), pp. 767-785, 2012.

[38] Guillemaud et al., Estimating the Bias Field of MR Images, IEEE TMI 16(3), pp. 238-251, 1997.
[39] Heinrich et al., MIND: Modality Independent Neighbourhood Descriptor for Multi-modal Deformable registration, MedIA 16(7), pp. 1423-1435, 2012.
[40] Heckemann et al, Automatic anatomical brain MRI segmentation combining label propagation and decision fusion, NeuroImage 33(1). pp. 115-26, 2006.
[41] Hu et al., Hybrid segmentation of Left Ventricle in Cardiac MRI using Gaussian-mixture Model and Region Restricted Dynamic Programming, JMRI 31(4), pp. 575-584, 2013.
[42] Huang et al., Contour Tracking in Echocardiographic Sequences via Sparse Representation and Dictionary Learning, MedIA 18(2), pp. 253-271, 2014.
[43] He et al., Deep Residual Learning for Image Recognition, CVPR, pp. 770-778, 2016.
[44] Juergen, et al., Hough Forests for Object Detection, Tracking, and Action Recognition, IEEE PAMI 33(11), pp. 2188-2202, 2011.
[45] Jolly et al., Combining Registration and Minimum Surfaces for the Segmentation of the Left Ventricle in Cardiac Cine MR Images, MICCAI, pp. 910-918, 2009.
[46] Karamitsos et al., The Role of Cardiovascular Magnetic Resonance Imaging in Heart Failure, JACC 54(15), pp. 1407-1424, 2009.
[47] Khalifa et al., A New Nonrigid Eegistration Framework for Improved Visualization of Transmural Perfusion Gradients on Cardiac Firstpass Perfusion MRI, ISBI, pp. 828-831, 2012.
[48] Klassen et al., Magnetic Resonance First Pass Perfusion Imaging for Detecting Coronary Artery Disease, European Journal of Radiology 57(3), pp. 412-416, 2006.
[49] Kong et al., A Dictionary Learning Approach for Classification: Separating the Particularity and the Commonality, ECCV, pp. 186-199, 2012.
[50] Kozerke, BOLD Myocardial Perfusion Imaging and Cardiac DTI at 3T, ISMRM, 2007.
[51] Kurkure et al., Localization and Segmentation of Left Ventricle in Cardiac Cine-MR Images, IEEE TBE 56(5), pp. 1360-1370,2009.
[52] Lecellier et al.,Statistical Region-based Active Contours for Segmentation: An overview, IRBM 35(1), pp. 3-10, 2014.
[53] LeCun et al.,Deep learning, Nature 521, pp. 436-444, 2015.

[54] Lee et al.,Efficient sparse coding algorithms, NIPS, 801-808, 2006.
[55] Li et al., Distance Regularized Level Set Evolution and its Application to Image Segmentation, IEEE TIP 19 (12), pp. 3243-3254, 2010.
[56] Li et al., Improved GVF Based Left Ventricle Segmentation from Cardiac MR Images Using Radial B-Snake Model, IEEE ICYCS, pp. 1000-1005, 2008.
[57] Lin et al., Automated Detection of Left Ventricle in 4D MR images: Experience from a Large Study, MICCAI, pp. 728-735, 2006.
[58] Lin et al., Localization and Atlas-Based Segmentation of the Heart from Cardiac MR Images: Validation with a Large Clinical Trial, ICBBE, pp. 2319-2322, 2008.
[59] Liu, Beyond Pixels: Exploring New Representations and Applications for Motion Analysis, Doctoral Thesis MIT, 2009.
[60] Liu et al., Accuracy Validation for Medical Image Registration Algorithms: a Review. Chinese Medical Science Review 27(3), pp. 176-181, 2012.
[61] Litjens et al., A Survey on Deep Learning in Medical Image Analysis, MedIA 42, pp. 60-88, 2017.
[62] Lombaert et al., Spatio-temporal Segmentation of the Heart in 4D MRI images using Graph cuts with Motion Cues, ISBI, pp. 492-495, 2010.
[63] Lorenzo et al., Atlas-Based Segmentation and Tracking of 3D Cardiac MR Images Using Non-rigid Registration, MICCAI, pp. 642-650, 2002.
[64] Lorenzo et al., Segmentation of Cardiac MR Images using the EM algorithm with a 4D Probabilistic Atlas and a Global Connectivity Filter, EMBC, pp. 626-629, 2003.
[65] Lorenzo et al., Segmentation of 4D cardiac MR Images Using a Probabilistic Atlas and the EM Algorithm, MICCAI, pp. 255-265, 2004.
[66] Lu et al., A Coupled Segmentation and Registration Framework for Medical Image Analysis using Robust Point Matching and Active Shape Model, MMBIA, pp. 129-136, 2012.
[67] Lynch et al., Segmentation of the Left Ventricle of the Heart in 3D+t MRI Data using an Optimised Non-rigid Temporal Model, IEEE TMI 27 (2), pp. 195–203, 2008.
[68] Mahapatra, et al., Joint Registration and Segmentation of Dynamic Cardiac Perfusion /Images using MRFs, MICCAI, 493-501, 2010.
[69] Mahapatra, et al., Integrating Segmentation Information for Improved MRF-Based Elastic Image Registration, IEEE TIP 20(1), 170-183, 2012.

[70] Mahapatra, et al., Joint Segmentation and Groupwise Registration of Cardiac DCE MRI using Sparse Data Representations, IEEE TIP 20(1), 1312-1315, 2015.
[71] Mairal, et al., Online Dictionary Learning for Sparse Coding, ICML, pp. 689-696, 2009.
[72] Mairal, et al., Sparse Modeling for Image and Vision Processing, Foundations and Trends in Computer Graphics and Vision 8(2), pp. 85-283.2014.
[73] Makela et al., A review of Cardiac Image Registration methods. IEEE TMI 21(9), pp. 1011–1021, 2002.
[74] Manning et al., Cardiovascular Magnetic Resonance. 2012.
[75] Mozaffarian et al., Heart Disease and Stroke Statistics 2015 Update, Circulation, 131, pp. 1-293, 2014.
[76] Mukhopadhyay et al., Data-Driven Feature Learning for Myocardial Segmentation of CP-BOLD MRI. FIMH, pp. 189-197, 2015.
[77] Mukhopadhyay et al., Unsupervised Myocardial Segmentation for Cardiac MRI, MICCAI, pp. 12-20, 2015.
[78] Mukhopadhyay et al., Total variation random forest: Fully automatic MRI segmentation in Congenital Heart Diseases, RAMBO-HVSMR, 2016.
[79] National Institutes of Health [online], March 2008. URL: `http://www.nhlbi.nih.gov/health/dci/Diseases/HeartAttack/HeartAttack_WhatIs.html`
[80] Ngo et al., Combining Deep Learning and Level Set for the Automated Segmentation of the Left Ventricle of the Heart from Cardiac Cine Magnetic Resonance, MedIA, 35, pp. 159-171, 2017.
[81] Oksuz et al., Dictionary Learning based Image Descriptor for Myocardial Registration of CP-BOLD MR , MICCAI, pp. 205-213, 2015.
[82] Oksuz et al., Unsupervised Myocardial Segmentation for Cardiac BOLD, IEEE TMI, 2017.
[83] Oksuz et al., Joint Myocardial Registration and Segmentation of Cardiac BOLD MRI, STACOM, 2017.
[84] Oktay et al., Anatomically Constrained Neural Networks (ACNN): Application to Cardiac Image Enhancement and Segmentation, arXiv preprint, 2017
[85] Olshausen et al., Emergence of Simple-cell Receptive Field Properties by Learning a Sparse Code for Natural Images, Nature 381(6583), pp. 607-609, 1996

[86] Otsu et al., A Threshold Selection Method from Gray-level Histograms, Automatica 11, pp. 23-27,
[87] Ou et al., DRAMMS: Deformable Registration via Attribute Matching and Mutual-saliency Weighting, MedIA 15(4), pp. 622-639, 2011.
[88] Ou et al., Validation of DRAMMS among 12 popular methods in cross-subject cardiac MRI registration, WBIR, pp. 209-219, 2012.
[89] Papazoglou et al., Fast Object segmentation in Unconstrained Video, ICCV, pp. 1777-1784, 2013.
[90] Paragios et al., A Variational Approach for the Segmentation of the Left Ventricle in Cardiac Image Analysis, IJCV 50(3), pp. 345-362, 2002.
[91] Parisot et al., Joint Tumor Segmentation and Dense Deformable Registration of Brain MR Images, MICCAI, 15(2), 651-658, 2012.
[92] Parisot et al., Concurrent Tumor Segmentation and Registration with Uncertainty-based Sparse Non-uniform Graphs, MedIA 18(4), 647-659, 2014.
[93] Pednekar et al., Automated Left Ventricular Segmentation in Cardiac MRI, IEE TBME 53(7), pp. 1425-1428, 2006.
[94] Peng et al., A Review of Heart Chamber Segmentation for Structural and Functional Snalysis using Cardiac Magnetic Resonance Imaging, MAGMA 29(2), pp. 155-195, 2016.
[95] Petitjean et al., A Review of Segmentation Methods in Short Axis Cardiac MR Images, MedIA 15(2), pp. 169-184, 2011.
[96] Pluim et al., Mutual-information-based Registration of Medical Images: a Survey, IEEE TMI 22(8), 986-1004, 2003.
[97] Querios et al., Fast Automatic Myocardial Segmentation in 4D cine CMR Datasets, MEdIA 18(7), pp. 1115-1131, 2014.
[98] Radau et al., Evaluation Framework for Algorithms Segmenting Short Axis Cardiac MRI, MIDAS Journal, 2009.
[99] Ramirez et. al., Classification and Clustering via Dictionary Learning with Structured Incoherence and Shared Features, CVPR, pp. 3501-3508, 2010.
[100] Ronneberger et. al., U-net: Convolutional networks for biomedical image segmentation, MICCAI, pp. 234-241, 2015.
[101] Rubinstein et al., Dictionaries for Sparse Representation Modeling, Proceedings of the IEEE 98(6), pp. 1045-1057, 2010.
[102] Rudin et. al.,Nonlinear Total Variation based Noise Removal Algorithms, Physica D: Nonlinear Phenomena 60, pp. 259-268, 1992.

[103] Rueckert et al., Nonrigid Registration using Free-form Deformations: Spplication to Breast MR images, TMI 18(8), pp. 712-721, 1999.
[104] Rusu et al., Stagewise K-SVD to Design Efficient Dictionaries for Sparse Representations, IEEE Signal Processing Letters, pp. 631-634, 2012.
[105] Rusu et al., Structured Dictionaries for Ischemia Estimation in Cardiac BOLD MRI at Rest. MICCAI, pp. 562-569, 2014.
[106] Rusu et. al. Synthetic Generation of Myocardial Blood–Oxygen-Level-Dependent MRI Time Series via Structural Sparse Decomposition Modeling. IEEE TMI 7(33), pp. 1422-1433, 2014.
[107] Scholkopf et al., Estimating the Support of a High-dimensional Distribution, Neural computation 13(7), pp. 1443-1471, 2001.
[108] Shakeri et al., Prior-Based Coregistration and Cosegmentation, MICCAI, 529-537, 2016.
[109] Singh et al., Pathogenesis of Atherosclerosis: A Multifactorial Process, Experimental Clinical Cardiology 7(1), 2002.
[110] Spottiswoode et al., Motion-guided Segmentation for Cine DENSE MRI, MedIA 13(1), pp. 105-115, 2009.
[111] Sotiras et al., Deformable Medical Image Registration: A Survey, IEEE TMI 7(32), pp. 1153-1190, 2013.
[112] Studholme et al., Deformation-based Mapping of Volume Change from Serial Brain MRI in the Presence of Local Tissue Contrast Change, IEEE TMI 5(25), pp. 626-639, 2006.
[113] Tan et al., Convolutional Neural Network Regression for Short-axis Left Ventricle Segmentation in Cardiac Cine MR Sequences, MedIA 39, pp. 78-86, 2017.
[114] Tavakoli et al., A Survey of Shape-based Registration and Segmentation Techniques for Cardiac Images, CVIU 117(9), pp. 966-989, 2013.
[115] Tibshirani, Regression Shrinkage and Selection Via the Lasso, Journal of the Royal Statistical Society 58(1), pp. 267-288, 1996.
[116] Tobon-Gomez et al., Benchmarking Framework for Myocardial Tracking and Deformation Algorithms: An Open Access Database, MedIA 17(6), pp. 632-648, 2013.
[117] Tong et. al., Segmentation of MR images via Discriminative Dictionary Learning and Sparse Coding: Application to Hippocampus Labeling, NeuroImage 76, pp. 11-23, 2013.
[118] Tong et. al., Discriminative Dictionary Learning for Abdominal Multi-organ Segmentation, MedIA 23, pp. 92-104, 2015.

[119] Tran et al., A Fully Convolutional Neural network for Cardiac Segmentation in Short-axis MRI, arXiv preprint, 2016.
[120] Tropp et. al., Signal Recovery from Random Measurements via Orthogonal Matching Pursuit, IEEE T Information Theory, 53(12), 4655-4666, 2007.
[121] Tsaftaris et al., A Dynamic Programming Solution to Tracking and Elastically Matching Left Ventricular Walls in Cardiac CINE MRI, ICIP, pp. 2980-2983, 2008
[122] Tsaftaris et al., Ischemic Extent as a biomarker for Characterizing Severity of Coronary Artery Stenosis with Blood Oxygen-sensitive MRI, JMRI, pp. 1338-1348, 2012.
[123] Tsaftaris et. al., Detecting Myocardial Ischemia at Rest With Cardiac Phase–Resolved Blood Oxygen Level–Dependent Cardiovascular Magnetic Resonance. Circ.: Card. Imag., 6(2), pp. 311-319, 2013.
[124] Vercauteren et al., Non-parametric Diffeomorphic Image Registration with the Demons Algorithm, MICCAI, pp. 319-326, 2007.
[125] Vlasveld; Introduction to One-class Support Vector Machines, [online], July 2013. URL: http://rvlasveld.github.io/blog/2013/07/12/introduction-to-one-class-support-vector-machines/
[126] Vlodaver et al., Coronary Heart Disease, 2012.
[127] Vu et al., DFDL: Discriminative Feature-oriented Dictionary Learning for Histopathological Image Classification, IEEE ISBI, pp. 990-994, 2015.
[128] Vu et al.,Learning a Low-rank Shared Dictionary for Object Classification, IEEE ICIP, pp. 4428-4432, 2016.
[129] Vu et al.,Histopathological Image Classification Using Discriminative Feature-Oriented Dictionary Learning, IEEE TMI 35(3), pp. 738-751, 2016.
[130] Wang et al., Joint Registration and Segmentation of Neuroanatomic Structures from Brain MRI, Acad. Radiol., 13(9), 1031-1044, 2006.
[131] Wang et al., Cardiac Motion and Deformation Recovery From MRI: A Review, IEEE TMI 31(2), pp. 487-503, 2012.
[132] Wang et al., Online Shared Dictionary Learning for Visual Tracking, ICIMCS, pp. 22-27, 2015.
[133] Weinsaft et al., BOLD New Directions in Myocardial Ischemia Imaging Myocardial Oxygenation Assessment by Cardiac Magnetic Resonance, JACC 13(7), pp. 1965-1967, 2012.

[134] Wright et. al., Robust Face Recognition via Sparse Representation, IEEE PAMI 31(2), pp. 210-227, 2009.
[135] Wright et. al., Sparse representation for Computer Vision and Pattern recognition, Proc. IEEE 98(6), pp. 1031-1044, 2010.
[136] Wu et. al., Snake Model-Based Automatic Segmentation of the Left Ventricle from Cardiac MR Images, IEEE BMEI, pp. 1-5, 2009.
[137] Wyatt et. al., MAP MRF Joint Segmentation and Registration of Medical images, MICCAI, pp. 539-552, 2003.
[138] Xiaohua et. al., Simultaneous Segmentation and Registration for Medical Images, MICCAI, pp. 663-670, 2004.
[139] Xiaohua et al.,Simultaneous Segmentation and Registration for Contrast-enhanced Breast MRI, IPMI, pp. 126-137, 2005.
[140] Xu et al., Snakes, Shapes, and Gradient Vector Flow, IEEE TIP 7(3), pp. 359-369, 1998.
[141] Yang et al., Fisher Discrimination Dictionary Learning for Sparse Representation, IEEE ICCV, pp. 543-550, 2011.
[142] Yezzi et al.,A Variational Framework for Integrating Segmentation and Registration through active contours, MedIA 7(2), 171-185, 2003.
[143] Yin et. al., Compressibility of Perfused Passive Myocardium, American Journal of Physiology 271(40), pp. 1864-1870, 1996.
[144] Zhen et al., Direct Estimation of Cardiac Bi-ventricular Volumes with Regression forests, MICCAI, 586-593, 2014.
[145] Zhou et al., Artifact-reduced Two-dimensional Cine Steady State Free Precession for Myocardial Blood-Oxygen-Level-Dependent Imaging, JMRI 31(4), pp. 863-871, 2010.
[146] Zhu et. al., Cardiac MR Image Segmentation with Incompressibility Constraint, ISBI, pp. 185-188, 2007.
[147] Zhuang et al., Registration-Based Propagation Framework for Automatic Whole Heart Segmentation of Cardiac MRI, IEEE TMI 29(9), pp. 1612-1625, 2010.
[148] Zhuang et al., Automatic Whole Heart Segmentation based on Image Registration, Diss. UCL (University College London), 2010.
[149] Zhuang et al., A Nonrigid Registration Framework using Spatially Encoded Mutual Information and Free-Form Deformations, IEEE TMI 10(30),pp. 1819-1828, 2011.
[150] Zhuolin et al., Label Consistent K-SVD: Learning a Discriminative Dictionary for Recognition, IEEE PAMI 35(11), pp. 2651-2664, 2013.