Cardiac Motion and Function Analysis using MR Imaging

Haiyan Wang

A dissertation submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy
of
Imperial College London

Februry 2015
Department of Computing
Imperial College London

Declaration of originality

I declare that the work presented in this thesis is my own, unless specically acknowledged.

Haiyan Wang

Abstract

Cardiacvascular disease (CVD) is the single leading cause of death in the world, claiming 17.3 million lives a year according to the World Health Organisation (WHO). The development of magnetic resonance (MR) imaging has provided clinicians and researchers with effective tools to detect, assess and monitor the progress of the disease and treatments. MR imaging produces images with high spatial resolution using noninvasive and non-ionising techniques. However, quantitative analysis of the cardiovascular system from MR images remains challenging.

The work presented in this thesis focuses on the utilization of cardiac motion information including motion tracking, quantification of the motion and prediction of clinical variables by incorporating motion information. The first main contributions of the thesis are approaches for sparse and dense motion tracking: a sparse set of key landmarks is detected and tracked. They are used as constraints to perform cardiac dense motion tracking using both 3D tagged and untagged image sequences from short-axis and long-axis MR views simultaneously.

In order to improve speed and accuracy of the motion tracking, we also develop an approach to identify and track a sparse set of distinctive landmarks in the presence of relatively large deformations for myocardium motion tracking without applying dense motion tracking. An integrated framework is proposed to combine entropy and SVD-based sparse landmark detection with a MRF-based motion tracking framework. In addition, the regional wall thickness systolic dyssynchrony index (SDI) derived directly from sparse motion tracking provides accurate quantification of LV motion, which agrees well with the clinical measurements.

In our last contribution, we successfully used manifold learning as a feature selection approach for a SVM-based classification and regression to analyse 209 cardiac MR image sequences. The SVM-based approaches directly operate on the manifold coordinates of the MR images without requiring any non-rigid registration or segmentation and is hence computationally efficient. We demonstrate that, by considering both inter- and intra-subject variation in the manifold learning, we are able to extract both anatomical and functional information. This can be used to construct powerful and reliable classifiers that are more predictive than global indices such as LV volume and mass. The manifold allows for investigating how much temporal information is needed improve the classification performance. The regression experiments demonstrate that there is a very strong correlation between manifold coordinates and obesity indices.

Acknowledgements

I would like to express my gratitude to my supervisor: Daniel Rueckert for all of his insightful guide, encouragement, patience and advice over the past few years.

I would also like to thank my second supervisor Dr. Philip Edwards and all of my collaborators, colleagues in the biomedical image analysis group for the companionship and the support. Especially to Dr. Wenzhe Shi, Dr. Xiahai Zhuang, Dr. Wenjia Bai, Dr. Claire Donoghue, Dr. Kanwal Bhatia, and Dr. Kai Pin (Tom) Tung.

I would also like to thank my husband, my son and my brother for support and encouragement.

Dedication

To my parents.

Acronyms

Adaboost . . adaptive boosting

ACS . . . . . acute coronary syndrome

AHA . . . . . American heart association

AV . . . . . atrioventricular

BEM . . . . . boundary element method

BMI . . . . . body mass index

BRIEF . . . . binary robust independent elementary features

BRISK . . . . binary robust invariant scalable keypoints

BSA . . . . . body surface area

CC . . . . . correlation coefficient

CCD . . . . . congenital cardiovascular defects

CHD . . . . . coronary heart disease

CI . . . . . cardiac index

CMR . . . . . cardiac magnetic resonance

CO . . . . . cardiac output

CSPAMM . . complementary spatial modulation of magnetisation imaging

CT . . . . . computed tomography

CTA . . . . . computed tomography angiography

CVDs . . . . cardiovascular diseases

DoG . . . . . difference of Gaussians

DMT . . . . . dense motion tracking

ECG . . . . . electrocardiography

ED . . . . . end diastolic

EDV . . . . . end-diastolic volume

EF . . . . . ejection fraction

EFFD . . . . extended free-form deformation

EM . . . . . expectation maximisation

ES . . . . . end systolic

ESV . . . . . end-systolic volume

FAST . . . . . features from accelerated segment test

Fat% . . . . . body fat percentage

FFD . . . . . free-form deformation

FFMI . . . . . fat free mass index

FREAK . . . fast retina keypoint

GMM . . . . Gaussian mixture models

HLA . . . . . horizontal long-axis

HOG . . . . . histogram of oriented gradients

LA . . . . . long-axis

LAD . . . . . left anterior descending artery

LBP . . . . . local binary patterns

LEE . . . . . Laplacian eigenmap embedding

LLE . . . . . local linear embedding

LM . . . . . left main coronary artery

LV . . . . . left ventricle

LVM . . . . . left ventricular mass

LVV . . . . . left ventricular volume

MDS . . . . . multidimensional scaling

MI . . . . . mutual information

MR . . . . . magnetic resonance

MRF . . . . . Markov random fields

NMI . . . . . normalized mutual information

ORB . . . . . orientated FAST and rotated BRIEF

PCA . . . . . principle component analysis

PDA . . . . . posterior descending artery

PDF . . . . . probability density function

PDM . . . . . point distribution model

PET . . . . . positron emission tomography

RCA . . . . . right coronary artery

RF . . . . . radio frequency

RMSE . . . . root mean squared error

RMSt . . . . regional myocardial strain

ROI . . . . . region of interest

RPM . . . . . robust point matching

RV . . . . . right ventricle

RWM . . . . regional wall motion

RWT . . . . . regional wall thickening

SA . . . . . short axis

SD . . . . . standard deviation

SDI . . . . . systolic dyssynchrony index

SIFT . . . . . scale invariant feature transform

SLMT . . . . sparse landmark motion tracking

SPAMM . . . spatial modulation of magnetisation imaging

SPECT . . . single-photon emission computed tomography

SSD . . . . . sum of squared differences

SSFP . . . . . steady state free precession

SURF . . . . speeded up robust feature

SUSAN . . . smallest univalue segmentation assimilating nucleus

SV . . . . . stroke volume

SVM . . . . . support vector machine

SVR . . . . . support vector regression

TDFFD . . . temporal diffeomorphic free form deformations

TFE . . . . . turbo field echo.

US . . . . . echocardiogram.

VECG . . . . vector electrocardiography.

VLA . . . . . vertical long-axis.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Cardiovascular disease

Cardiacvascular disease (CVD) is the single leading cause of death in the world, claiming 17.3 million lives a year according to the World Health Organisation (WHO) [Organization, 2013]. The WHO estimates that the death toll will reach 23.3 million per year by 2030.

Cardiovascular diseases are a group of disorders of the heart and vessels. Figure 1.1 shows the percentage breakdown of deaths due to CVDs in United States, according to the report from the American Heart Association(AHA) in 2013 [Go et al., 2013]. Among those dieases, coronary heart disease, stroke, high blood pressure, and heart failure are the leading causes of death.

Coronary Heart Disease

As shown in Figure 1.1, coronary heart disease (CHD) causes the greatest proportion of deaths in United States. Worldwide, it is also the single largest cause of death [Finegold et al., 2012].

CHD is caused by atherosclerosis (Figure 1.2), a condition in which plaques, formed by fatty deposits, calcium deposits and abnormal inflammatory cells, build up on the inner lining of the

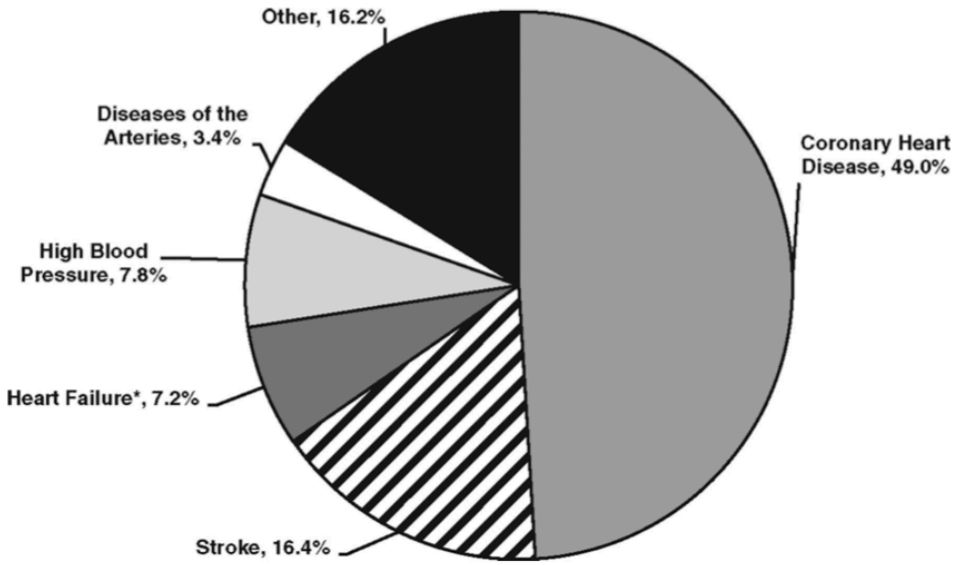


Figure 1.1 – Percentage breakdown of deaths due to CVDs in United States [Go et al., 2013].

arteries. Over time, plaques can gradually harden to narrow and obstruct the artery, or rupture to form a large blood clot which can block the artery. The risk factors for causing the CHD include aging, smoking, family history, high blood cholesterol, diabetes, high blood pressure and others.

The obstruction of a coronary artery will reduce the blood flow to the heart, hence restricting the supply of oxygen and nutrients to the muscles of the heart. This may leads to serious consequences, including angina, myocardial infarction, heart failure and in some cases arrhythmias [NIH, 2014].

Angina occurs when the heart muscle is not supplied with sufficient oxygenated blood, which is usually felt in the form of pain in the chest, shoulders, neck or arms. When a plaque becomes large enough to partially obstruct the blood flow, it will cause the above symptoms of angina when the heart demands more blood (e.g. during exercises) than can be supplied. This type of angina is termed stable angina as it occurs regularly with exertion or stress. In other situations, a partially ruptured plaque may cause a sudden worsening of the blockage in the artery, leading to an unpredictable occurrence of angina. This medical condition is referred to as unstable angina. Patients with this type of angina are at high risk of developing a myocardial infarction.

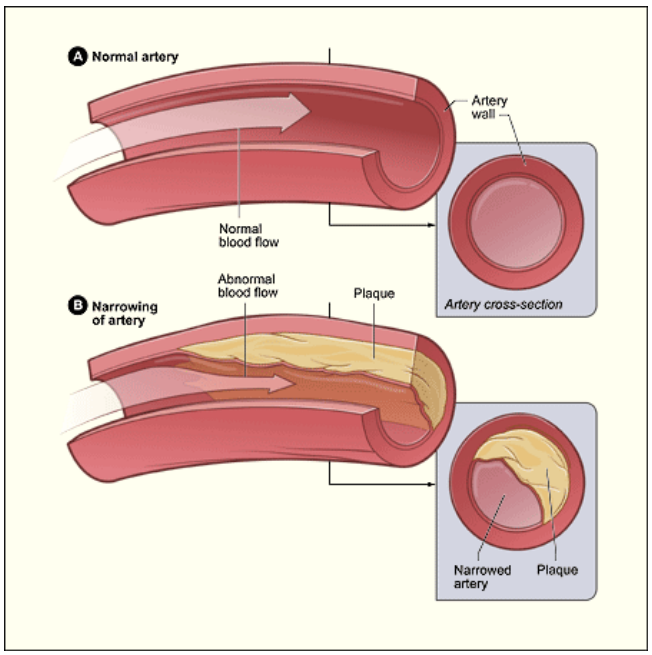


Figure 1.2 – A normal artery and an artery with the condition of atherosclerosis, from [NIH, 2014].

Myocardial infarction, commonly known as heart attack, occurs when the ruptured plaque develops a total blockage of one of the coronary artery, so that the area of the heart muscle supplied by that artery is injured or even dies. The severity of a myocardial infarction depends on how large the affected region of the heart muscle is. Most infarctions involve the left ventricle because the myocardium is especially vulnerable to coronary occlusion due to its larger size and greater demand for oxygen and nutrients.

Over time, CHD can weaken the heart muscle and can lead to changes in the cellular structure of the myocardium. This may alter the propagation of electrical impulses in the heart and cause arrhythmias. During an arrhythmia, the heart can beat with an irregular rhythm, too fast or too slow, and the heart may not be able to pump sufficient blood to the rest of the body. Some of the arrhythmias can be serious or even life threatening.

Stroke

A stroke happens when the blood supply to a part of the brain is blocked, causing the rapid loss of brain function. The two most common types of stroke are ischaemic stroke and haemorrhagic

stroke. As consequences, a stroke may cause temporary or permanent paralysis, loss of speech or vision, memory loss and pain or numbness in parts of the body.

The arteries to the brain become narrowed or blocked either by a clot that builds up on-site or a travelling blood clot formed from elsewhere, causing the ischaemic stroke. Ischaemic stroke constitutes 85% of all strokes. Haemorrhagic stroke is caused by a leaking or burst blood vessel in the brain.

Hypertension accounts for $35-50\%$ of stroke risk [Whisnant, 1996]. Other risk factors include atrial fibrillation, blood lipid, smoking, high cholesterol, diabetes and others. The diagnosis of stroke is normally carried out with the assistance of imaging techniques. Most of the strokes are preventable by managing risk factors or surgical intervention. When a stroke occurs, it is critical to get immediate treatment to prevent death and minimize the long-term consequences of a stroke.

High blood pressure

High blood pressure (hypertension) usually presents no symptoms so that it is hard to feel or notice. However, if it is not treated overtime, the heart may become enlarged which increases the chance of heart failure. High blood pressure puts persistent strain on the heart, leading to hypertensive heart disease and coronary artery disease. Hypertension can also cause heart attacks, strokes, or kidney failure ([O'Brien et al., 1995], [Singer and Kite, 2008], [Donnan et al., 2008]).

Blood pressure is usually measured using two pressure values: systolic and diastolic pressure. High blood pressure is said to be present if it is constantly higher than the recommended level (in general systolic-pressure/diastolic-pressure $> 140/90 mmHg$ for individual occasions).

The primary cause of the primary hypertension, accounting for $90-95\%$ of all cases of hypertension, remains unclear. However, aging, smoking, obesity, diabetes and a high level of salt intake are considered to be associated with the condition. High blood pressure cannot not be cured, however, medications and changes in lifestyle can help to reduce and control it.

Heart Failure

Heart failure, often called congestive heart failure (CHF), occurs when the heart is unable to pump sufficient blood flow to meet the needs of the body. As a result fluid tends to build up in the lungs or other parts of the body. This is most commonly caused by coronary artery disease (which weaken the heart muscle over time or suddenly), untreated high blood pressure (which leads to muscle stiffness and eventually weakening) or damaged heart valve and arrhythmias [McMurray and Pfeffer, 2005].

The heart failure symptoms are determined by which side of the heart fails. Left-sided failure is often accompanied by respiratory signs such as shortness of breath and increased rate of breathing. When left ventricle (LV) systolic failure occurs, the amount of blood ejected with each heart beat drops significantly. The most common cause of the LV systolic dysfunction is ischemia that results in infarction and scar formation, leading to abnormal or absent wall motion. There also exists LV diastolic failure in which the LV wall becomes too stiff to relax adequately. This causes inadequate filling of the ventricle and hence less pumped out blood. Both systolic and diastolic failures result in elevated end-diastolic pressure and cause fluid to back up into the lungs. The excess fluid in lungs impairs gas exchange.

Right-sided heart failure is most commonly caused by the left-sided heart failure. Increased fluid pressure as a result of LV failure is transferred through the lungs to the right side of the heart ultimately damaging the RV and its ability to pump. Similar to the consequences of LV failure, the RV failure also generates the fluid to back up into the veins and accumulate in the body. This causes swelling under the skin and usually affects foot and ankle first.

It is vital to identify the cause of heart failure and to reverse it if possible. Treatments include changing lifestyle and medication to help the heart muscle pump better and improve the circulation. For certain patients with heart failure, a coronary bypass surgery may help improve blood flow to the damaged heart muscle.

1.2 The anatomy of the heart

The heart is the centre of the the cardiovascular system that delivers nutrients and oxygen to the body and removes waste and carbon dioxide from the body ([Bray et al., 1994], [Thibodeau et al., 2004]). The heart (Figure 1.3) is a muscular pumping organ, consisting of four chambers in total: left atrium (LA), left ventricle (LV), right atrium (RA) and right ventricle (RV). The base of the heart connects to the great blood vessels: the aorta, vena cava, pulmonary artery and pulmonary veins.

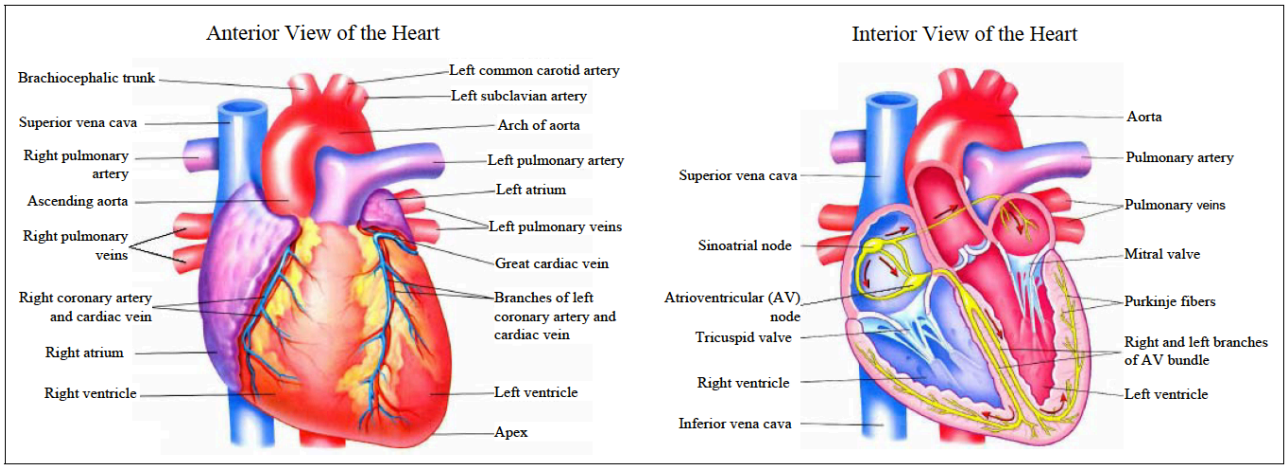


Figure 1.3 – Heart anatomy. Figures from [Thibodeau et al., 2004].

Each of the four chambers of the heart has its own separate entity, separated from each other and the two large blood vessels by valves, septum and muscular interventricle walls. The outer wall of the heart is made of three layers: the epicardium, myocardium and endocardium. The epicardium is the outer protective layer of the heart. The muscular middle layer wall is the myocardium that produces the contractions of the heart. The LV has a very thick layer of myocardium as more force is required to eject the blood to the whole body. Finally the endocardium is the inner layer of the heart that covers the inner surface of the heart valves and extends to the inner lining of blood vessels.

Circulatory loops

The cardiovascular system includes two circulatory loops: The pulmonary circulation loop and the systemic circulation loop. The cardiovascular system carries oxygen and nutrients to the cells and tissues in one direction and carries carbon dioxide and waste in the opposite direction.

The RA receives deoxygenated blood from the vena cava and passes it through the tricuspid valve into the RV where the blood is pumped through the pulmonary valve into pulmonary artery leading to the lungs for oxygenation. This process is referred to as pulmonary circulation.

For the systemic circulation, the LA receives oxygenated blood from the lungs via the pulmonary veins and pumps it through the mitral valve into the LV where the blood is pumped through the aortic valve to the aorta leading to the organs and tissues of the body.

All valves are designed to ensure that the blood only flows in one direction, allowing blood to flow either from one chamber to another, or allowing blood to flow out of the heart. The valves have two or three cusps, which are pushed open or close entirely by the pressure differences across the valve. For instance, mitral valve opens during diastole when atrial pressure increases above that of the left ventricle; and closes at the end of diastole when the pressure difference no longer exists. The papillary muscles and the chordae tendineae cause tension to hold the mitral valve when it closes.

Coronary circulation

The coronary circulatory system is the heart's own set of blood vessels that provides the myocardium with a blood supply. There are five large epicardial coronary arteries that lie in grooves between the heart's chambers, as shown in Figure 1.4. The left main coronary artery (LM) and the right coronary artery (RCA) branch off from the aorta, while the left anterior descending artery (LAD) and the left circumflex artery (CIRC) arise from the LM when it splits into two. The posterior descending artery (PDA) is a continuation of either the RCA or the CIRC. All these five major arteries and coronary capillaries deliver oxygenated

blood to all of the heart's cells. The coronary sinus, which is a vein on the posterior side of the heart, carries deoxygenated blood from the myocardium to the vena cava.

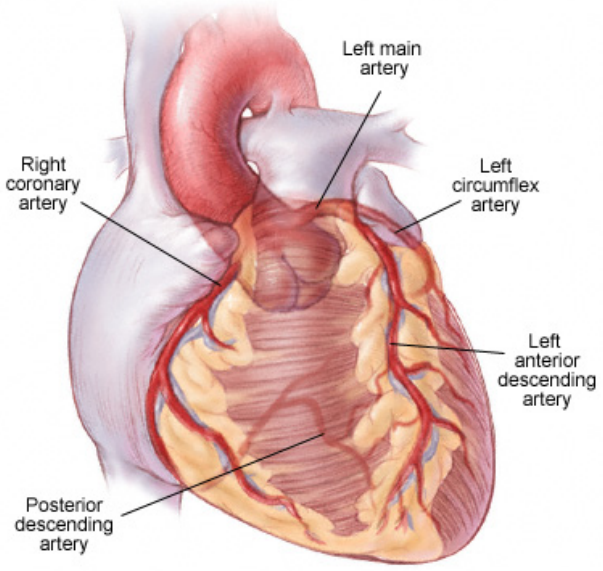


Figure 1.4 – The coronary arteries. Figure from [Scientific, 2014].

The occlusion of the coronary arteries may generally lead to an infarct, due to the fact that the blood flow from the coronary arteries supply oxygen and nutrition to the myocardium. Whenever the coronary blood flow falls below what is required to meet metabolic needs, the myocardium is considered ischemic and hence the pumping performance may be impaired. This is especially the case for the myocardium of the LV, which is a very thick muscular layer. In general, different regions of the heart are supplied by the different coronary arteries. Disease of the left main coronary artery is particularly dangerous because this artery usually supplies blood to the myocardium of the LV.

Cardiac eletrophysiology

The heart can generate and conduct electrical impulses so that the myocardium contracts and relaxes in a repetitive cycle to pump the blood throughout the body. The cardiac conduction system (Figure 1.5) consists of the sinoatrial (SA) node, the atrioventricular (AV) node, the Bundle of His and the Purkinje fiber network. The sinoatrial (SA) node, located in the upper

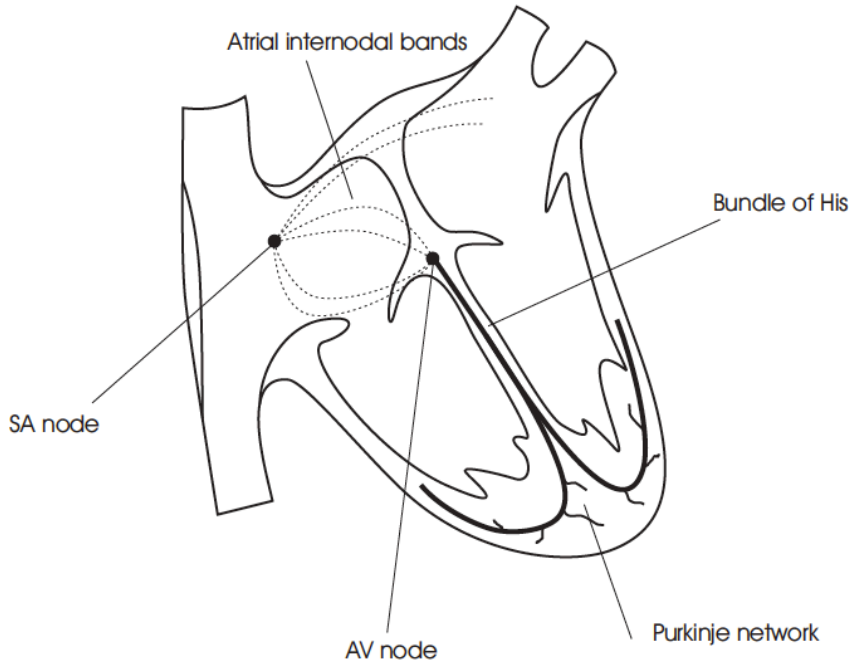


Figure 1.5 – The cardiac conduction system. Image is adapted from figure 13.5 in [Bray et al., 1994].

wall of the right atrium, emits electrical impulses at a regular rate of $60 - 100$ beats/minute in general. These electrical impulses spread through the heart wall to cause both atria to contract. The impulses then reach the AV node at the bottom of the atrium where they are delayed to allow the atria to contract and push all the blood into the ventricles. Thereafter, the electrical impulse passes through the Bundle of His which divides into two branches, one leading to the LV and the other to the RV. Then the impulse pass further into Purkinje fiber network and eventually triggers the muscle fiber in the ventricles to contract. The electrical activation spreads through the myocardium so rapidly that cells of the ventricles contract nearly simultaneously.

Cardiac cycle

The cardiac cycle is a series of events that occur in every single heartbeat. A single cycle of cardiac activity consists of two basic phases: diastole and systole.

During the diastole phase, the ventricles are relaxed. The LA receives oxygenated blood from

the lungs through pulmonary veins, while venous blood returns to the RA through the superior vena cava and inferior vena cava. The mitral and tricuspid valve are open and blood is flowing from the atria into ventricles through them. The other two valves, the aortic and pulmonic valve are closed. At the end of diastole, when the ventricles are about 80% full, both atria contract and propel more blood into the ventricles.

The systole phase comes shortly after the end of the diastole and the ventricles contract. The ventricle pressure increases rapidly and causes the mitral and tricuspid valve to close when it exceeds the atrial pressure. As a result no blood is entering the ventricles during the systole; however, the atria continue to receive blood through the vena cava and pulmonary veins. Meanwhile, the continued contraction raises the ventricular pressure quickly beyond the pressure in the aorta and the pulmonary artery. This opens the pulmonary and aortic valve to permit ejection of blood into the aorta and pulmonary artery. As the ejection of blood continues the ventricular pressure falls and eventually goes below the pressure in the aorta and the pulmonary artery, causing the pulmonary and aortic valve to close, and later the mitral and tricuspid to open, which begins a new cycle.

1.3 Imaging of the cardiovascular system

Magnetic resonance imaging (MRI) [Lauterbur et al., 1973] and other imaging techniques including ultrasound (US), computed tomography (CT), positron emission tomography (PET), and single-photon emission computed tomography (SPECT) have provided clinicians and researchers with effective tools to detect, assess and monitor the progress of heart disease and their treatments. The assessment of the cardiac function using these techniques can be quantified by cardiac functional indices, which will be discussed in Section 3.1.

Ultrasound

Echocardiography or ultrasound allows non-invasive, low-cost visualisation of the heart and hence has been routinely used for diagnosis and management of heart diseases. It uses ultrasonic

waves which travel through an object. It then detects the reflected signals and their intensities to reconstruct images of the structures in the object. An echocardiogram can provide information about the size and shape of the heart, pumping capacity and any tissue damage. Figure 1.6 shows an echocardiography image of the heart.

Echocardiography also has the advantage of enabling real-time imaging, which is extremely important for a dynamic organ such as the heart. It is safe with no side effects and widely available. However, echocardiography has a limited field of view (FOV) and thereby the whole heart cannot be imaged in a single scan. Furthermore, image noise and artefacts present a substantial challenge for automatic image processing algorithms [Grau and Noble, 2005].

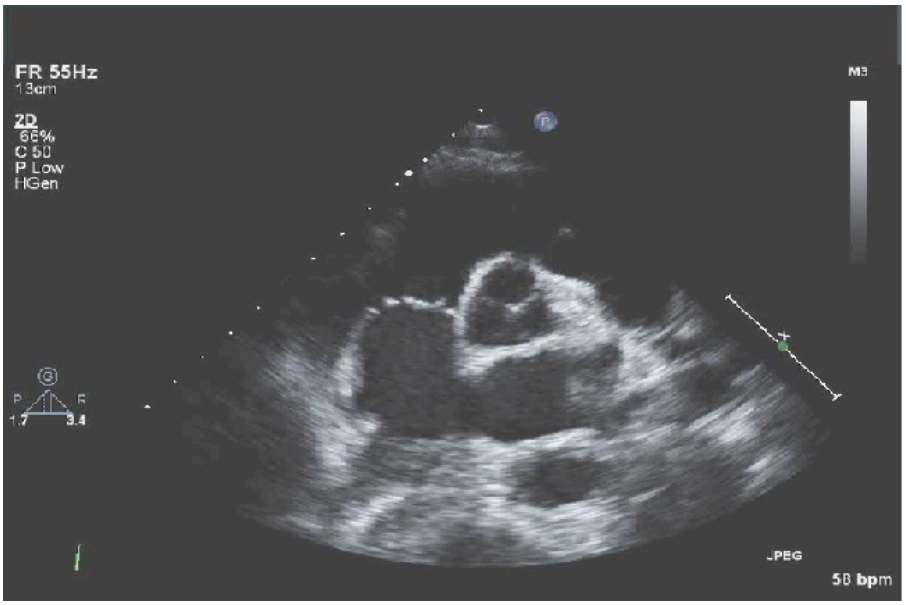


Figure 1.6 – Echocardiography image of the heart. Figure from [Ech, 2015].

Computed tomography (CT)

CT uses x-rays to produce tomographic images of the scanned object. It is best suited for viewing bone injuries, diagnosing lung and chest problems, and detecting cancers. In the area of the cardiac imaging, it can produce excellent imaging of the coronary arteries. Figure 1.7 shows example CT images of the heart.

CT is widely available and generally a quick and inexpensive imaging procedure. Its cross-

sectional images can provide high-contrast resolution to distinguish less than 1% differences of density between tissues. However, the high radiation dose used, which is 100 to 1,000 times higher than that used in conventional X-rays, can damage the tissue and organ. It has been estimated that about 1.5% $\sim$ 2.0% of all cancers in USA may be attributable to the radiation from CT [Brenner and Hall, 2007]. Additionally, injected radio-contrast agents used in around half of the CT scans in USA can cause allergic reactions.

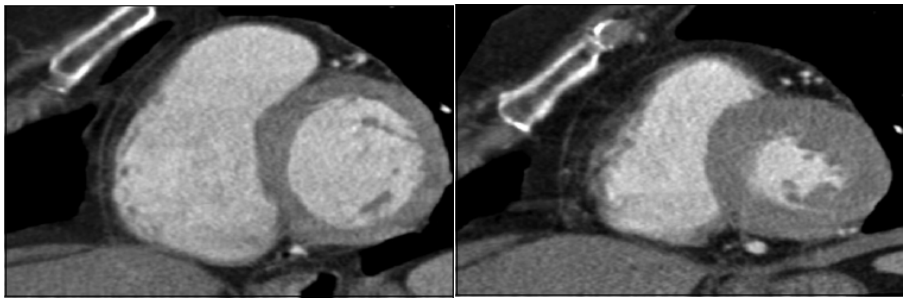


Figure 1.7 – CT images of the heart. Figure from [Bruzzi et al., 2006].

Positron emission tomography (PET)

PET [Slifstein, 2014] detects pairs of gamma rays indirectly emitted by a positron-emitting tracer and constructs three-dimensional images of tracer concentration within the detected object. The radioactive tracer is incorporated into a biologically active molecule, most commonly Fluorodeoxyglucose (FDG), before it is injected into detected object (i.e. human body). As the molecules move through the body; the cells take it up along with the radioactive tracer. Therefore the concentration of the tracer with FDG indicates tissue metabolic activity. PET can detect impaired metabolic activity, such as coronary artery disease and injured but viable myocardium. PET is also used heavily in clinical oncology and neurology. Combined with CT or MRI, PET is increasingly able to provide both anatomic and metabolic information [Chander et al., 2008].

PET comes with unique advantages due to its ability to study molecular activity, which includes identifying disease in its earliest stage before anatomic changes become apparent and distinguishing between active and non-active tumors (tumor cells are usually metabolically

active). Despite using radioactive tracers, PET is noninvasive and safe as the tracer is short-lived and low dose. Nevertheless, PET has the limitation of high cost, low spatial and temporal resolution and can only be carried out in specialised centres. Figure 1.8 shows PET images of the heart.

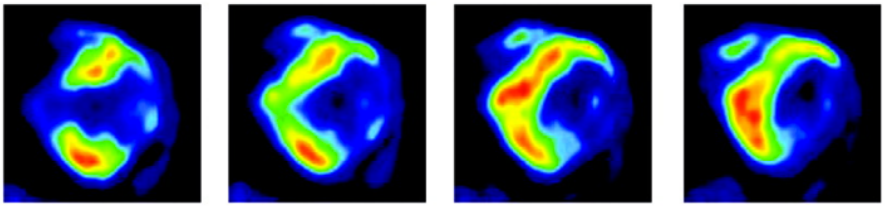


Figure 1.8 – PET images of the heart. Figures from [Hasegawa et al., 2006].

Single-photon emission computed tomography (SPECT)

SPECT, similar to PET, also relies on the emission of gamma rays, however, the tracer used in SPECT emits gamma radiation that is measured directly. The tracer is normally attached to a specific ligand, whose properties bind it to certain types of tissues in the body. SPECT can be used to produce myocardial perfusion imaging (MPI) for the diagnosis of ischemic heart disease. In MPI, tracers are taken up by the myocardium in rough proportion to perfusion; the diseased myocardium receives less blood flow thus taking less tracers. SPECT has also seen applications in neurology.

SPECT is noninvasive and safe. In addition, it is significantly less expensive than PET scans but its images have lower resolution than that of PET [Rahmim and Zaidi, 2008]. Figure 1.9 shows an example myocardial perfusion scans.

Magnetic Resonance Imaging (MRI)

Among the imaging modalities available routinely in clinics, magnetic resonance imaging (MRI) offers accurate imaging of anatomy, morphology, myocardial motion and blood flow with good contrast, which enables quantitative assessment of heart function, such as the measurement

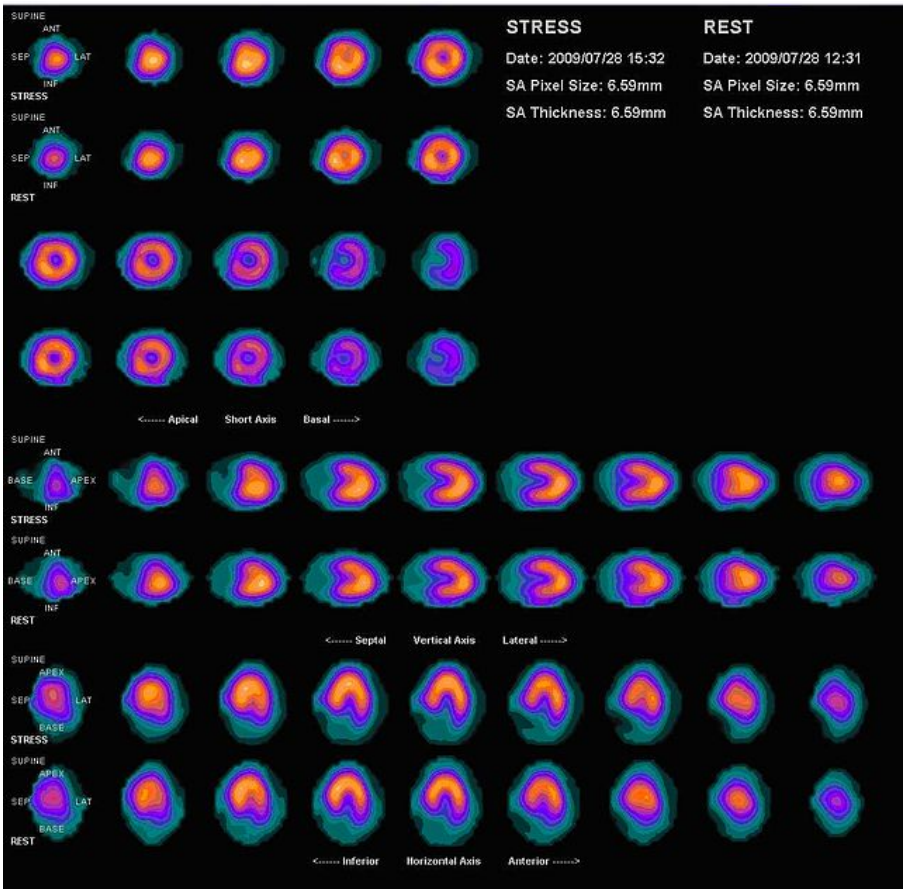


Figure 1.9 – SPECT myocardial perfusion image (MPI). Figures from [national Library of Medicine, 2011].

of ventricular volume, wall thickness and ejection fraction etc. It has been established as the gold standard for assessment of patients with cardiovascular disease. MR provides images of high spatial resolution with a non-invasive and non-ionising imaging approach and hence has been used widely. Extensive research in recent years has focused on developing algorithms for the anatomical modelling and functional modelling of the heart from MR image sequences. However it remains a challenge to accurately and reliably extract this information.

MRI uses strong magnetic fields and radio waves to form images of the body [Lauterbur et al., 1973]. Around 70% of the human body mass is water. Different tissues have different water density, longitudinal relaxation time (T1) and transverse relaxation time (T2), which are the basic phenomenons utilized in MR imaging. When the patient is positioned within an MRI scanner which forms a strong magnetic field B_ϕ, the spinning protons in water molecules will

try to align to this field and precess around the applied magnetic field B_ϕ with a frequency ω_ϕ:

$$\omega_\phi = \gamma B_\phi \tag{1.1}$$

Here γ is a constant and ω_ϕ is known as the resonance frequency. Then an electromagnetic radio-frequency (RF) pulse is applied at frequency of ω_ϕ and the RF energy is absorbed by the protons. Once the RF is removed, the absorbed RF energy is emitted by the protons at the resonance frequency ω_ϕ and the strength of the emitted signal is proportional to proton density. This signal will be received by the MRI scanner for image reconstruction. The radio signal can be made to encode position information by varying the main magnetic field using gradients.

The contrast of MRI is related to T1 and T2 relaxation. The magnetic moment generated by the proton spinning has two components, the longitudinal and transverse magnetisation. They align to B_ϕ and are tilted when RF is applied. Once the RF is removed, the recovery of the two components of a spin, are called T1 relaxation and T2 relaxation correspondingly. As mentioned above, T1 and T2 relaxation times are biological properties and differ from tissue to tissue and normal tissue to diseased tissue. This is why MRI can define the anatomical structure with high spatial resolution and good contrast between blood and tissue.

Despite all of the advantages of MRI when compared to other noninvasive imaging modalities, it requires more patient cooperation than other tests and the examination times are significantly longer. MRI is also unsuitable for patients with any metallic object inside their body.

Imaging planes

In order to make an accurate CVD diagnosis, it is essential to acquire images in multiple orientations as the heart is continuously in motion. When MRI scans are performed in clinic practices, a vertical long axis view (VLA, or two-chamber view (2CH)) is first defined using a transverse image, where the slice is positioned through apex and centre of the mitral valve [Cerqueira et al., 2002] as shown in Figure 1.10.

A second long axis view (3CH) is acquired by flipping the orientation and adjusting the plane on the vertical long axis image through the apex and the middle of the mitral valve. The main

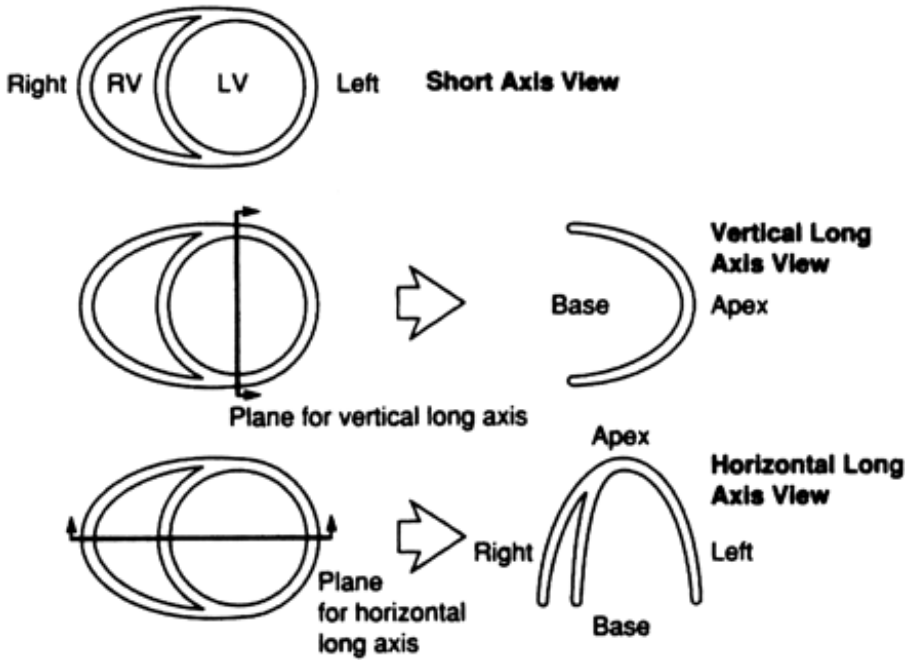


Figure 1.10 – Cardiac plane definition. Image from [Cerqueira et al., 2002].

purpose of acquiring this view is to use it to prevent any angulation errors while planning the short axis views. Slices are then defined perpendicular to the long axis of the heart representing the short axis (SA) geometry, which are parallel with the mitral valve. Four-chamber views (4CH) or so called horizontal long axis (HLA) views are thereafter planned on the equatorial short axis view to show both ventricles and atria in a single plane. Some example images are shown in Figure 1.11, and Figure 1.12 shows a stack of multiple slices of short axis view images. In general, the through-plane resolution ($8 \sim 10mm$) is much lower than the in-plane resolution ($1 \sim 2.5mm$).

Cine MR imaging

Because MRI is based on time-consuming scanning, real-time MRI is possible but only with limited image quality or low temporal resolution. Cine MR images [Uecker et al., 2010] are obtained with electrocardiography (ECG) triggered segmented imaging. Each image in the sequence is typically composed of information gathered over several heart beats within breath-holding of $10 \sim 20s$.

Although the breath hold introduces respiratory motion between images acquired during different breath holds, these images can be very helpful in studying cardiac function, valvular

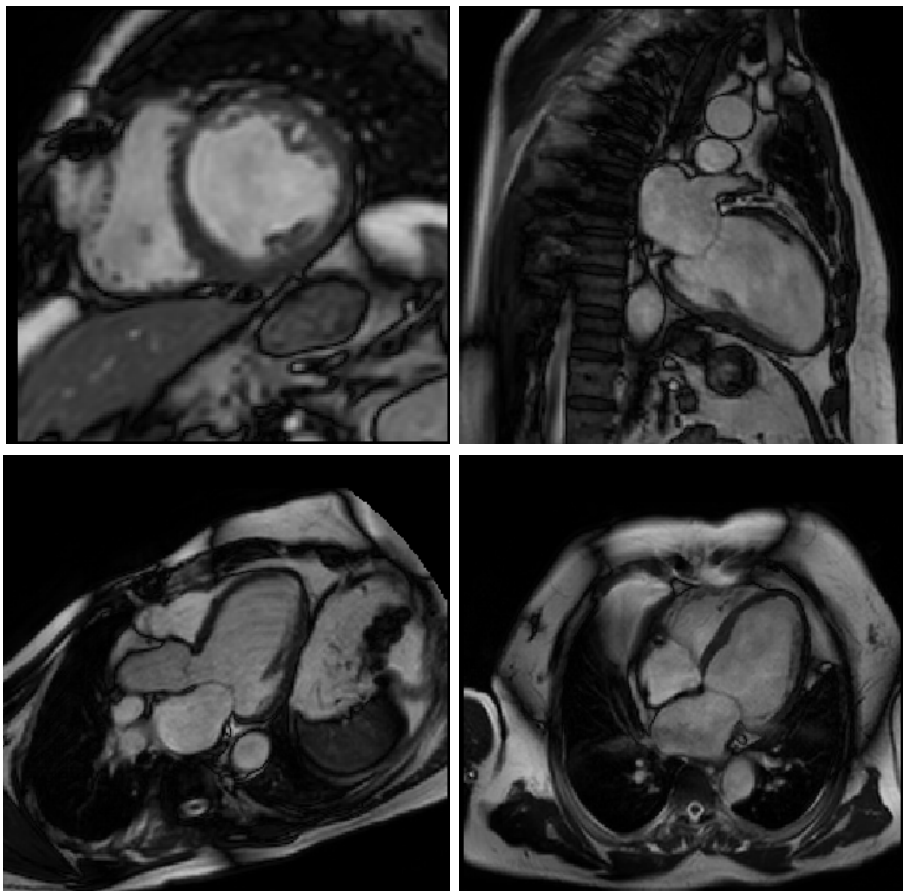


Figure 1.11 – The images from up-left to bottom-right show respectively SA, 2CH, 3CH and 4CH views of the heart.

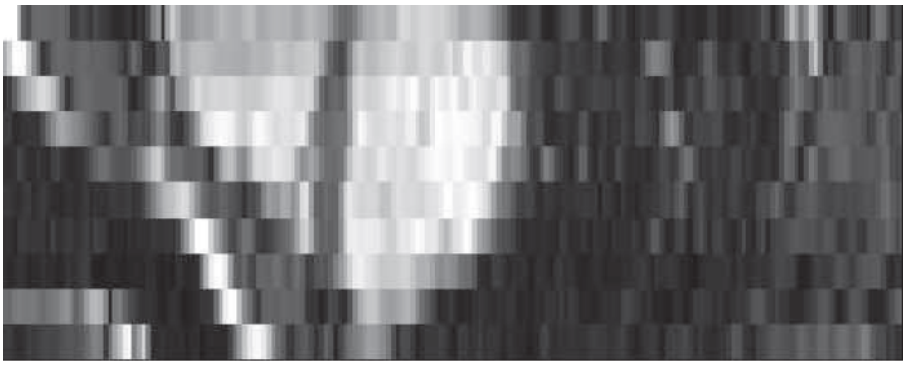


Figure 1.12 – The image shows a stack of multiple slices of short axis view images.

function, and movement of blood through the heart.

Tagged MR imaging

Among various MRI techniques, 2D and 3D tagged MRI can be used to obtain regional information about the cardiac motion; hence is useful in assessing ventricular function [Axel et al., 2005].

The 2D spatial modulation of magnetization (SPAMM) [Axel and Dougherty, 1989] technique introduces markers within the myocardium throughout the entire imaging volume. By modulating the magnetization gradient, the signal from the myocardium can be nulled in a grid pattern before the image acquisition. At imaging, this dark appearing grid moves with the tissue that was tagged. Thereby, deformation fields in the myocardium can be reconstructed by measuring the motion of the labelled tissue. However, SPAMM needs longer image acquisition time and is limited by the fading of the tag during the cardiac cycle.

In order to capture 3D cardiac motion, one approach is to acquire multiple 2D tagged slices in short-axis view and long-axis view. Another approach is to perform 3D tagged imaging [Ryf et al., 2002]. Multiple 2D tagged imaging is prone to slice mis-registration and 3D imaging can overcome this limitation. A fast 3D method has been proposed by Rutz et al. that allows measuring deformation of the whole heart in three breath-holds of 18 heartbeats duration each [Rutz et al., 2008]. Figure 1.13 shows 3D tagged MR images of the heart.

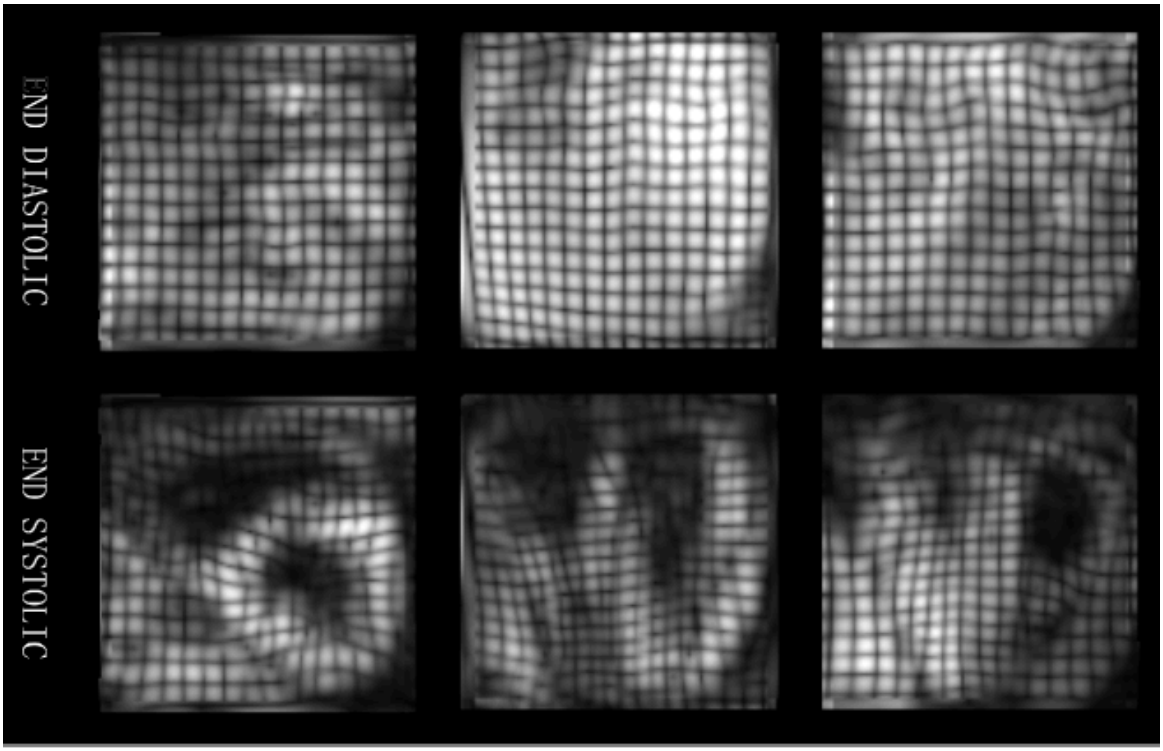


Figure 1.13 – 3D tagged MR images of the heart.

1.4 Research contributions and thesis outline

The research presented in this thesis contributes to cardiac motion and function analysis using MR imaging. The ultimate objective of cardiac image analysis is to provide useful and efficient tools for the diagnosis and treatment of patients with cardiovascular diseases. The work in this thesis has focused on utilising of cardiac motion information to help achieving this target.

Relevant concepts and methodologies from the fields of object detection and machine learning are first reviewed in *Chapter 2*. This chapter provides details of some of the most widely used object detection algorithms including feature detectors and feature descriptors, with a qualitative and quantitative evaluation of their respective strengths and weaknesses. Machine learning algorithms applied as part of the work, including SVM, Adaboost, regression and manifold learning are presented in detail. *Chapter 3* explores the state of the art techniques in the literature in cardiac motion analysis, such as sparse motion tracking, deformable models and registration based algorithm applied to both cine MR images and tagged MR images. We also discuss applications of manifold learning in the field of cardiac motion analysis. Clinical indices associated with ventricular function assessment and obesity indices are also introduced.

Chapter 4 presents a machine learning algorithm to identify and track a pair of key landmarks in three LA views throughout the cardiac cycle. Dense cardiac motion tracking is also performed using both tagged and untagged image sequences from short-axis and long-axis views simultaneously. The key advantage of the proposed method is the simultaneous use of complementary motion information contained in multiple views of cine images, and in the tagged and untagged images. By combining complementary information using a spatially adaptive weighting and valve plane constraint, we have successfully built an accurate cardiac motion analysis framework.

In order to improve speed and accuracy of the motion tracking, *Chapter 5* presents an approach to identify and track a set of sparsely distributed anatomical distinctive landmarks in the presence of relatively large deformations of the myocardium without applying dense motion tracking. An integrated framework is proposed to combine entropy and SVD-based sparse

landmark detection with a MRF-based motion tracking. The accuracy of the proposed sparse motion tracking is evaluated by tracking a group of manually annotated landmarks on the endocardial border of the left ventricle (LV). The improvement of the tracking accuracy compared to dense motion tracking method shows that the proposed sparse motion tracking approach benefits from the larger search area of the MRF. With the purpose of aiding the diagnosis, a novel method has been developed to derive regional wall thickness SDI to provide accurate quantification of left ventricular motion. The regional cardiac functional analysis has been performed using the proposed sparse landmark motion tracking and AHA LV model. Compared with the analysis generated by a state-of-the-art commercial tool, the results obtained by the proposed method are in good agreement. It shows that it is plausible to use the proposed sparse landmark motion tracking to estimate the dysynchrony index for the regional deformation.

Chaper 6 successfully used manifold learning as a feature selection approach for a SVM-based classification and regression to analyse 209 cardiac MR image sequences. The SVM-based approach directly operates on the manifold coordinates of the MR images without the need for any explicit image analysis. We demonstrated that, by considering both inter- and intra-subject variation in the manifold learning, we are able to extract both anatomical and functional information. This can be used to construct powerful and reliable classifiers that are more predictive than global indices like LV volume and mass. The manifold allows for investigating how much temporal information we need to help to improve the classification performance. The regression experiments demonstrate that there is a very strong correlation between the manifold coordinates with obesity indices. This strong correlation discovered between obesity indices and cardiac images suggests that the body constitution has a strong impact on the appearance of the heart, and obesity indices are good indicator of the appearance and motion status of the heart. In particular, to our knowledge no previous research has studied the relationship between the obesity indices and the cardiac MR images. The proposed manifold learning approach is computationally efficient as no registration or segmentation is required.

Chapter 2

Object detection and machine learning

Object detection is one of the fundamental challenges in computer vision and medical image processing. Applications for object detection include visual tracking, object categorisation, localisation, image matching and image retrieval.

Detection of anatomic landmarks from MR images provides important structural and functional information. A landmark or local feature is an image location which is distinctive from points in its immediate neighbourhood. Local features can be corners, edges, image patches or regions. Particularly in medical imaging, local feature detection has attracted more and more attention because it can provide well localised and individual anchor points, which can be used in registration, motion tracking, segmentation and other applications. Learning based object detection methods have been applied to detect objects with large variations in medical imaging. This chapter provides an overview of object detectors, feature descriptors and machine learning approaches that are relevant to the image based classification and regression analyses presented in this thesis.

In cardiac motion analysis, we want to obtain a sparse set of landmarks that can provide anatomy-specific constraints to establish correspondences between images being tracked or registered. This can be achieved by the following approach:

1. Detect a set of distinctive landmarks in the source image.

2. Compute a descriptor from a region of insterst (ROI) around each landmark.

3. Match the descriptors in the target image to find the correspondences for the landmarks.

On the other hand, some feature descriptors can be computed densely pixel by pixel over an image to provide textured regions, thereby facilitating the image precessing. For classification of textured regions, no explicit detection is required.

Accordingly the organisation of this chapter is as follows: A number of local feature detectors are reviewed in Section 2.1. In Section 2.2, feature descriptors are categorised in four classes according to the applications or the methods used. In Section 2.3, we focus on machine learning based methods for object detection, classification and regression. Machine learning can be treated as a matching procedure when it is fed with features for training and testing.

2.1 Detectors

A traditional way of utilising local features normally involves feature detection, feature description and feature matching. A feature detector locates distinctive points or interest regions and is usually followed by extracting a feature descriptor to represent the detected points or regions. In this subsection, the following detectors are reviewed: Moravec's corner detector [Moravec, 1977], the Harris corner detector [Harris and Stephens, 1988], Shi-Tomasi's corner detector [Tomasi and Shi, 1994], the smallest univalue segment assimilating nucleus (SUSAN) [Smith and Brady, 1997], the difference of Gaussians (DoG) detector [Lowe, 2004], the fast-Hessian detector [Bay et al., 2006] and the features from accelerated segment test (FAST) [Rosten and Drummond, 2005] [Rosten et al., 2010]. These detectors have played a significant role in the literature and have demonstrated very good performance in their applications.

2.1.1 Moravec's corner detector

Developed in 1977, Moravec's corner detector is one of the earliest corner detection algorithms [Moravec, 1977]. It shifts a local window by a few pixels in various directions and investigates the local change of the intensity by taking the sum of squared differences (SSD) between two patches as follows:

$$D(u,v) = \sum_{(x,y)} w(x,y)\,[I(x+u, y+v) - I(x,y)]^2 \tag{2.1}$$

Here(u,v) indicates the shift in two directions and only a set of shifts at every 45° is performed. $w(x,y)$ denotes a binary window function at (x,y):

$$w(x,y) = \begin{cases} 1, & \text{if inside window} \\ 0, & \text{if outside window} \end{cases} \tag{2.2}$$

A pixel is only considered as a corner pixel when the patches around the current location in all directions are dissimilar from each other. If the pixel is on an edge, then the nearby patches on the same side of the edge will be similar, but the nearby patches in a direction perpendicular to the edge will be dissimilar.

As pointed out in [Moravec, 1977] and [Harris and Stephens, 1988], the shortcomings of this detector are that the search for corner is not isotropic due to the consideration of shifts only at every 45° and the response is noisy due to the use of a binary window. In addition, this detector is very sensitive to edges as it considers of only the maximum intensity change.

2.1.2 Harris corner detector

Harris and Stephens [Harris and Stephens, 1988] developed the Harris corner detector to address the limitations of Moravec's corner detector. This detector is able to determine whether a pixel is on an edge, on a corner, or in a region with uniform intensity distribution. To achieve these,

the detector first computes the matrix $\mathbf{A}$ for each pixel (x, y) in a given image I:

$$\mathbf{A}(x,y) = \begin{pmatrix} \sum\limits_{x,y} w(x,y) \cdot [I_x(u,v)]^2 & \sum\limits_{x,y} w(x,y) \cdot I_x(u,v) I_y(u,v) \\ \sum\limits_{x,y} w(x,y) \cdot I_x(u,v) I_y(u,v) & \sum\limits_{x,y} w(x,y) \cdot [I_y(u,v)]^2 \end{pmatrix} \tag{2.3}$$

Here I_x and I_y are the partial derivatives of the image I in the x- and y- direction, (u, v) indicates the shift in two directions. $w(x, y)$ denotes a Gaussian window function at (x, y), which substitutes for the binary window function in Moravec's detector. $\mathbf{A}$ considers all small shifts.

The eigenvalues λ_1 and λ_2 of the matrix $\mathbf{A}$ represent the principle curvatures of the image intensity and their magnitudes can be used to characterise the local intensity patterns as follows:

1. If $\lambda_1 \approx 0$ and $\lambda_2 \approx 0$, then this pixel is in an uniform intensity region;
2. If $\lambda_1 \approx 0$ and λ_2 is large, or $\lambda_2 \approx 0$ and λ_1 is large, then the pixel is on an edge;
3. If both λ_1 and λ_2 are large, then a corner has been located.

To avoid actually computing the eigenvalue decomposition of the matrix $\mathbf{A}$, a cornerness measure $C(x, y)$ is introduced for each pixel (x, y) to replace the eigenvalues for justification:

$$C(x,y) = \det(\mathbf{A}) - k(\mathrm{tr}(\mathbf{A}))^2 \tag{2.4}$$

Here k is a constant value; $\det(\mathbf{A})$ and $\mathrm{tr}(\mathbf{A})$ indicate the determination and trace (the sum of the elements on the main diagonal) of the square matrix $\mathbf{A}$ respectively.

1. If $C(x, y) > 0$ and is small, then this pixel is in an uniform intensity region;
2. If $C(x, y) < 0$, then the pixel is on an edge;
3. If $C(x, y) > 0$ and is large, then a corner has been located.

In practice, a threshold needs to be set carefully to avoid false positives.

The Harris detector is widely used in practice because of its good detection rate and high repeatability rate. However, it is computationally expensive compared to Moravec's corner detector because it requires convolution with a Gaussian kernel. Moreover, it is sensitive to noise due to the application of gradient information. Despite its improved response, it is still not rotationally invariant because the elements of the Harris matrix are calculated using only the horizontal and vertical intensity gradients. Nevertheless, Harris detector is probably the most popular feature point detector. Its underlying idea has been extended to yield stable features under rotations, scalings, up to full affine transformations [Tomasi and Shi, 1994] [Zheng et al., 1999] [Triggs, 2004] [Kenney et al., 2005].

2.1.3 Shi-Tomasi's corner detector

Shi and Tomasi [Tomasi and Shi, 1994] proposed a similar criterion for selecting points with 'good features' to the Harris corner detector. They find the minimal eigenvalue instead of cornerness defined in Equation 2.4. A candidate point is accepted as corner when,

$$C(x, y) = \min(\lambda_1, \lambda_2) > \lambda \tag{2.5}$$

Here λ is a predefined threshold. In their work [Tomasi and Shi, 1994], Shi and Tomasi also proposed a technique for monitoring the quality of image features by measuring the feature dissimilarity during tracking. They demonstrated that using the criteria in Equation 2.5 to select corners produces stable features for tracking.

2.1.4 SUSAN

SUSAN (Smallest Univalue Segment Assimilating Nucleus) has been introduced by Smith and Brady [Smith and Brady, 1997]. It uses a morphological approach: Considering that pixels belonging to the same object share similar intensities if they are in a small enough region, SUSAN compares the intensity within a circular mask M. Given the nucleus (x_0, y_0), i.e. the

centre of M, the intensity of each pixel (x, y) within M is compared with the intensity of (x_0, y_0):

$$m(x,y) = \exp\Big(-\Big(\frac{I(x,y) - I(x_0,y_0)}{t}\Big)^6\Big) \tag{2.6}$$

Here t denotes the intensity difference threshold, which determines the minimum contrast of edges which will be picked up. The use of the sixth power of the exponent has been determined empirically, which provides a balance between good stability about the threshold and the function of counting similar pixels. The total number of pixels within M having similar intensity to the nucleus can be represented by:

$$n(M) = \sum_{(x,y)\in M} m(x,y)$$

This is referred to as the area of USAN(univalue segment assimilating nucleus).

The SUSAN operator is given by:

$$R(M) = \begin{cases} g - n(M), & \text{if } n(M) < g \\ 0, & \text{otherwise} \end{cases}$$

Here g is a fixed threshold, set to 3/4 of the total number of pixels within M. As illustrated in Figure 2.1, SUSAN filters out more or less uniform regions. The area of USAN is around 50% of the total area of M for regions near edges and when near corner it decreases to 25%. The SUSAN operator will reach local maximum for a corner pixel and relatively smaller values for edges.

SUSAN demonstrates a high repeatability for detecting feature points. It avoids the computation of derivatives, thus allows edges, lines, corners and junctions to be accurately and quickly located. SUSAN is invariant under translation and rotation. However, the feature points detected based on edge are less discriminative [Tuytelaars and Mikolajczyk, 2008].

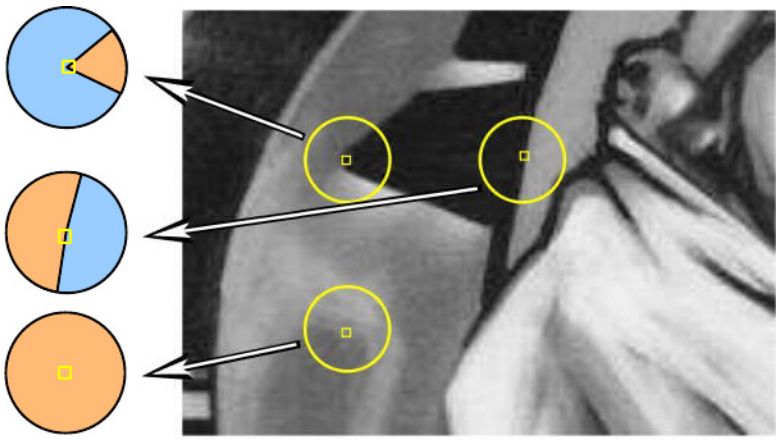


Figure 2.1 – SUSAN corners are detected by segmenting a circular neighborhood into similar (orange) and dissimilar (blue) regions. Corners are located where the relative area of the similar region (USAN) reaches a local minimum below a certain threshold [Tuytelaars and Mikolajczyk, 2008].

2.1.5 Difference of Gaussians (DoG)

The Difference of Gaussians (DoG) [Lowe, 1999] [Lowe, 2004] extracts distinctive invariant features by searching for scale space extrema with a cascaded filtering. It was introduced in conjunction with a descriptor, SIFT, which will be reviewed in Section 2.2.2.

The DoG detector first constructs a scale space by convolving the original image with a Gaussian function $G(x, y, \sigma)$, kernel at different scales σ:

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2} * I(x, y) \tag{2.7}$$

To efficiently detect local features that are invariant to the scale σ, difference-of-Gaussian images are produced by subtracting adjacent scale Gaussian images:

$$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) = L(x, y, k\sigma) - L(x, y, \sigma) \tag{2.8}$$

Lowe [Lowe, 2004] smoothed the image with Gaussian convolution mask $s + 3$ times in each octave scale, such that $k = 2^{1/s}$. Octave is a music theory term adopted by DoG, which refers to doubling the frequency. Then, the smoothed Gaussian images are combined using Equation 2.8 to produce a set of $s + 2$ DoG images. Increasing s might in principle yield more refined

feature points, but can make the selection of points unstable due to noise. This process is repeated for the next octave while the resolution of the image decreased by a factor of 2 (see Figure 2.2(a)). Increasing the number of octaves enables the searching of feature points of all possible sizes.

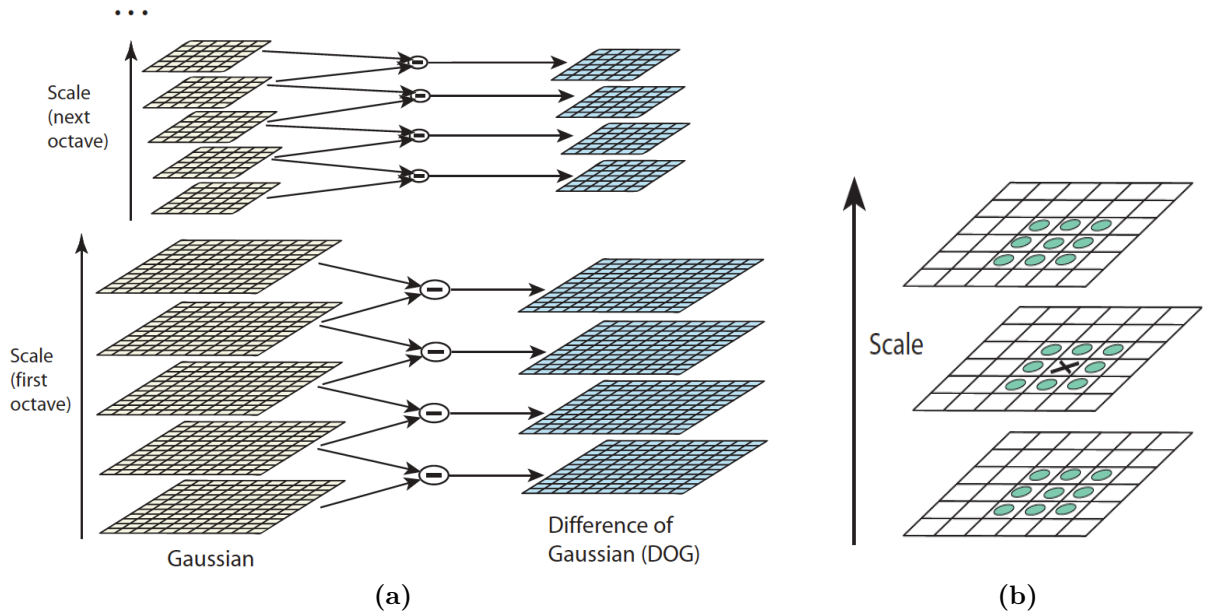


Figure 2.2 – (a) DoG produces a set of scale space images by subtracting adjacent Gaussian images. Two octaves of scale are shown while in the next octave the Gaussian images are down-scaled by a factor of 2. $s = 2$ here. (b) Maxima and minima of the DoG images are detected by comparing a pixel (marked with X) to its 8 neighbors at the current scale and 18 neighbors at the adjacent scales (all marked with circles) [Lowe, 2004].

Local maxima and minima of $D(x, y, \sigma)$ are detected by comparing each point to its eight neighbors at the current scale and nine neighbors in the scale above and below, as illustrated in Figure 2.2(b). In their initial implementation [Lowe, 1999] the corresponding feature points to these local extrema in scale space are simply identified at the location and scale of the central sample point.

Since the Difference-of-Gaussian produces strong response on edges, poorly defined maxima which are unstable need to be eliminated by evaluating their principal curvatures. The ratio of the principal curvatures across the edge and that in the perpendicular direction are estimated borrowing ideas from the cornerness map approach (see Equation 2.4 in section 2.1.2) used by Harris and Stephens [Harris and Stephens, 1988]. The filtering step eliminates feature points that have a ratio of the principal curvatures that is greater than a predefined threshold.

In general, DoG provides robust matching across affine distortion, changes in view point and the addition of noise [Lowe, 2004].

2.1.6 Fast-Hessian detector

Inspired by SIFT, Bay et al. [Bay et al., 2006] introduced a Hessian matrix-based detector, and its corresponding descriptor is SURF (speeded up robust features). This detector is also referred to as SURF detector.

The Hessian matrix describes how the normal to an iso-surface, e.g. the image intensities, changes. The Hessian matrix at point $\mathbf{x}$ at scale σ is defined as:

$$H(\mathbf{x}, \sigma) = \begin{pmatrix} L_{xx}(\mathbf{x}, \sigma) & L_{xy}(\mathbf{x}, \sigma) \\ L_{xy}(\mathbf{x}, \sigma) & L_{yy}(\mathbf{x}, \sigma) \end{pmatrix} \tag{2.9}$$

Here $L_{xx}(\mathbf{x}, \sigma)$, $L_{xy}(\mathbf{x}, \sigma)$ and $L_{yy}(\mathbf{x}, \sigma)$ denote the convolution of the second-order Gaussian derivative in the $x-$, $xy-$ and $y-$ direction with the image at point $\mathbf{x}$, which are in effect the second differentials of the smoothed images. The local maxima of the determinant of the Hessian and the trace of this matrix can be used to detect local features. DoG (discussed in Section 2.1.5) approximates the trace using multiple scale-space pyramids, while the fast-Hessian detector relies on the determinant for selecting both the location and the scale of the local feature.

Bay et al. then approximate the second-order Gaussian derivative by box filters (Figure 2.3), which can be efficiently computed using integral images (which will be discussed in section 2.2.3 [Viola and Jones, 2002]). The filter response is defined as:

$$\det(H_{approx}) = D_{xx}D_{yy} - (0.9D_{xy})^2$$

where D_{xx}, D_{yy} and D_{xy} are the convolution of the box filters with the image. The scale space is exploited by up-scaling the filter size from 9×9 as shown in Figure 2.3 to 15×15, 21×21 and

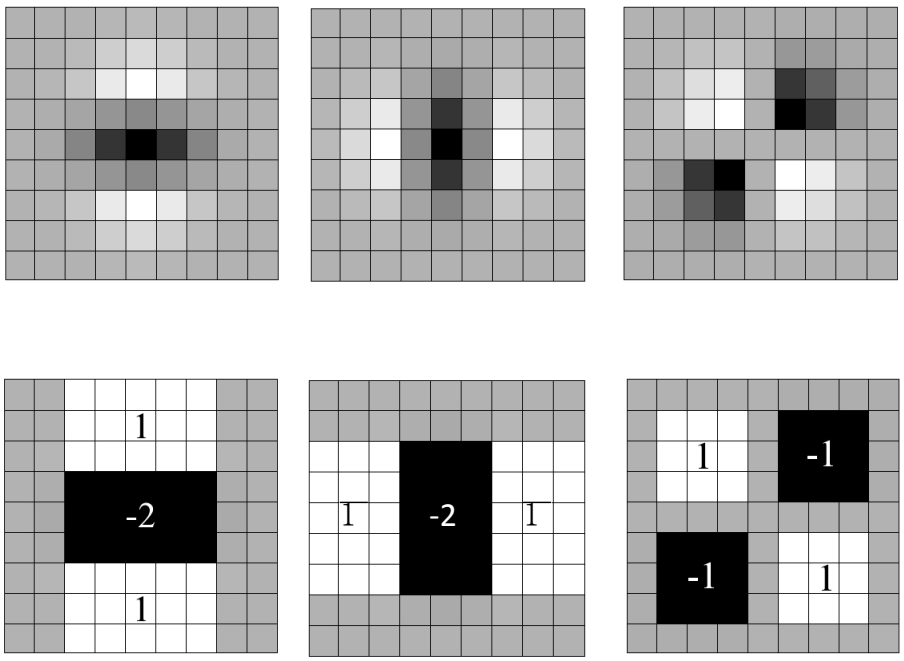


Figure 2.3 – Top line left to right: The second-order Gaussian partial derivatives in y-direction, x-direction and xy-direction. Bottom line left to right: Box filters used in fast-Hessian detector in y-direction, x-direction and xy-direction. Weights of black and white regions are shown, while grey regions are weighted between 1 and 0. Figure is adapted from [Bay et al., 2006].

27×27. A non-maximum suppression in a $3 \times 3 \times 3$ neighbourhood and sub-pixel refinement are then applied similar to the DoG detector. The fast-Hessian detector has been reported to be $3 \sim 5$ times faster than DoG [Bay et al., 2006].

2.1.7 Features from Accelerated Segment Test (FAST)

The FAST detector [Rosten and Drummond, 2005, Rosten et al., 2010] is a high-speed corner detector that does not rely on discrete image derivatives. Similar to the idea in SUSAN (discussed in Section 2.1.4) of comparing the intensity of the pixels to that of the centre pixel, FAST only compares pixels on a circle around the centre pixel p. As illustrated in Figure 2.4, the pixel p is classified as a corner if there exists a set of 12 contiguous pixels in the circle of 16 pixels around x which are brighter or darker than x by a threshold t. To speed up the process, the detector first examines pixels 1 and 9; then pixels 5 and 13. If neither pixel breaks the contiguous chain of 12, the remaining pixels on the circle will be tested. To improve the generality and speed, a machine learning approach is introduced to select the pixels which yield

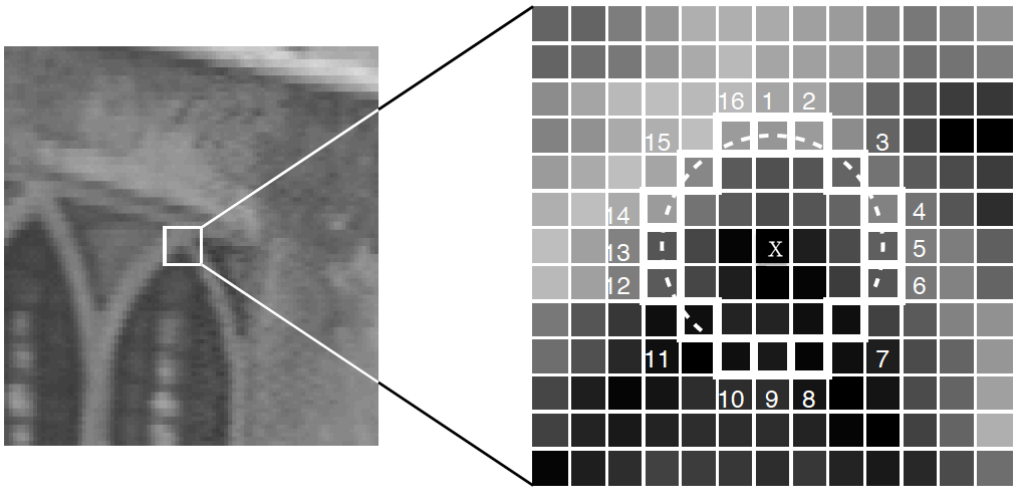


Figure 2.4 – 12 point segment test corner detection in an image patch. The highlighted squares are the pixels used in the corner detection. The pixel at x is the centre of a candidate corner. the arc is indicated by the dashed line passes through 12 contiguous pixels which are brighter than x by more than the threshold [Rosten et al., 2010].

the most information about whether the candidate pixel is a corner: The pixels are classified into three classes based on the relative intensity differences in which the pixels are either darker, brighter or similar to their centre pixel respectively. Then for each candidate pixel, a vector of size 16 with 3 values are formed with a Boolean variable as label. A decision tree is trained by recursively selecting a location in this vector to partition the whole training set into three subsets. The resulting decision tree can correctly classify all corners seen in the training set by querying location and relative intensity state of a pixel in the circle around the target point. Finally, non-maxima suppression is applied on the sum of the absolute difference between the pixels in the circle and the center pixel.

The FAST detector is indeed very fast. According to [Rosten et al., 2010] it is about 40 times faster than the DoG detector (see section 2.1.5) and 20 times faster than Harris detector (see section 2.1.2). Detailed comparisons in terms of speed with other detectors are listed in the following Table. Table 2.1 summarises the most important properties for the feature detectors discussed in this section. It is difficult to provide an objective comparison of the accuracy of the detectors as different experimental frameworks and criteria can be used for evaluation. For the accuracy in terms of localization, scale and robustness, depending on the application

scenario, some of these properties are more crucial than others. In addition, a trade-off has to be made between efficiency and accuracy. In general, the DoG, Fast-Hessian and FAST detectors provide the best overall performance among all the detectors.

Table 2.1 – Summary of feature detectors.

Detector \ Property	Types of feature			Scale invariant	Rotation invariant	Translation invariant	Learning	Pixel rate[1] (MPix/s)	Relative speed%[2]	Subpixel
	corner	edge	blob							
Moravec's	✓	✓	×	×	×	×	×	NA[4]	NA[4]	×
Harris	✓	✓	×	×	×	✓	×	8.05	58.9	✓
Shi-Tomasi's	✓	×	×	✓	✓	✓	×	6.50	72.8	✓
SUSAN	✓	✓	×	×	✓	✓	×	12.3	38.3	×
DoG	✓	×	✓	✓	✓	✓	×	4.72	100	✓
Fast-Hessian	✓	✓	✓	✓	✓	✓	×	15.3[3]	30.8	✓
FAST	✓	×	×	✓	✓	×	✓	188	2.5	×

Note: [1] Timing results for a selection of feature detectors run on frames of two video sequences. The pixel rates are given by [Rosten et al., 2010]

[2] The percentage of the rate between the speed of a detector and the speed of DoG detector based on [Rosten et al., 2010].

[3] Fast-Hessian is claimed to be three times faster than DoG [Bay et al., 2006], so the pixel rate of Fast-Hessian is estimated according to the listed DoG pixel rate.

[4] Not tested.

2.2 Feature descriptors

Feature descriptors are designed to characterise points or regions. They may encode a substantial amount of local spatial information to make it invariant to scale, rotation, change of viewpoint and brightness. A large number of feature descriptors has been proposed and can be divided into two classes: one is sparse descriptors which sample a local patch and describe its features; the other is dense descriptors which extract local features pixel by pixel over the input image [Randen and Husoy, 1999] [Chen et al., 2010]. Some descriptors can be used in a dense or sparse sampling way. Therefore we categorise feature descriptors in another way as intensity-based, distribution-based, filter-based and binary descriptors to avoid confusion.

2.2.1 Intensity-based descriptors

The intensities within a patch around a point of interest are the most straightforward description of a feature point. However, it is not invariant to scale, rotation or translation, though it is tolerant of a certain degree of blur. It is normally used for matching when the patch appearance is stable [Coupé et al., 2011].

The intensity-based descriptors are very sensitive to to possible image artifacts, consequently image denoising is often required. Most of these denoising methods restore the intensity value of each image pixel by averaging the intensities of its neighbours in some way, for example box filtering [McDonnell, 1981]. The non-local patch-based mean filter has been originally introduced for image denoising [Buades et al., 2005]. In this approach, a non-local weighting is introduced to capture the geometric similarity between the neighbouring points and the point under study, enhances the influence of the most similar neighbouring points when averaging the intensity. The restored non-local mean intensity $nl(\mathbf{x}_i)$ of pixel $\mathbf{x}_i$ is the weighted average of all the pixel $\mathbf{x}_j$ within its neighbourhood V_i as:

$$nl(\mathbf{x}_i) = \sum_{\mathbf{x}_j \in V_i} \omega(\mathbf{x}_i, \mathbf{x}_j) I(\mathbf{x}_j) \tag{2.10}$$

where $\omega(\mathbf{x}_i, \mathbf{x}_j)$ is assigned to each point $\mathbf{x}_j$ in the neighbourhood V_i around the $\mathbf{x}_i$. As illustrated in Figure 2.5, the value of the weight is defined as the Gaussian-weighted Euclidean distance between patch V_i and V_j:

$$\omega(\mathbf{x}_i, \mathbf{x}_j) = \frac{1}{Z_i}\exp(-\frac{||I(M_i) - I(M_j)||_2/\sqrt{2}\sigma^2}{h^2}) \tag{2.11}$$

Here σ^2 is the local variance of the noise and can be estimated via pseudo-residuals for each pixel $\mathbf{x}_i$, Z_i is a normalization constant ensuring that $\sum_j \omega(\mathbf{x}_i, \mathbf{x}_j) = 1$ and h is a smoothing parameter [Coupe et al., 2008]. Using this non-local weight, higher weights are assigned to pixels that are similar to the central pixel, and lower weights are associated with pixels that are dissimilar to the central pixel $\mathbf{x}_i$.

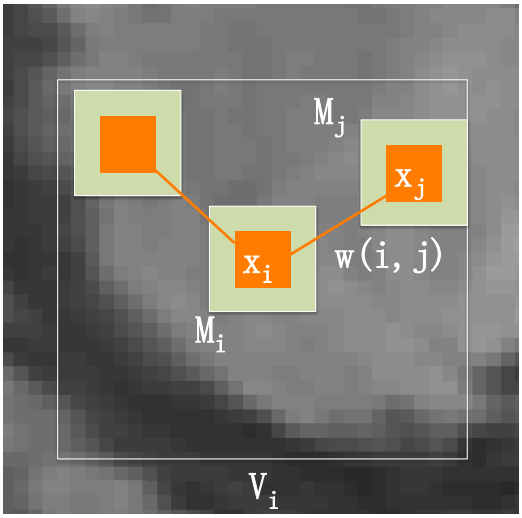


Figure 2.5 – This figure shows the estimation of the non-local weights $\omega(\mathbf{x}_i, \mathbf{x}_j)$ between patch $P(\mathbf{x}_j)$ and $P(\mathbf{x}_i)$, an extended region N_i is defined around the pixel $\mathbf{x}_i$.

The non-local mean filter has been adopted for segmentation [Rousseau et al., 2011] [Coupé et al., 2011] and registration [Heinrich et al., 2011]. In these approaches, for every pixel $\mathbf{x}_i$ of image to be segmented or registered, $\mathbf{x}_j$ is no longer the neighbour of $\mathbf{x}_i$ but the labeled sample inside the search region V_i for the N selected subjects. Consequently the label of $\mathbf{x}_i$ can be estimated based on a weighted label fusion of label of $\mathbf{x}_j$.

2.2.2 Distribution-based descriptors

Scale Invariant Feature Transform (SIFT)

The Scale Invariant Feature Transform (SIFT) was introduced in conjunction with a DoG detector presented in Section 2.1.5. However, both components can be used separately or together with other detectors/descriptors. SIFT is always regarded as state of the art and has been demonstrated to achieve the most stable and efficient results [Li and Allinson, 2008]. In order to achieve rotation invariance, the SIFT algorithm ([Lowe, 1999], [Lowe, 2004]) calculates a consistent orientation for each keypoint to produce the descriptor relative to this orientation.

Given the keypoints detected by the DoG detector, the gradient orientation $\theta(x, y)$ and magnitude $m(x, y)$ are calculated for each pixel around a keypoint $\mathbf{x}$ on the Gaussian smoothed image $L(x, y, \sigma_{\mathbf{x}})$ (see Equation 2.7), where $\sigma_{\mathbf{x}}$ is the scale at which $\mathbf{x}$ are detected.

Each of these pixels around a keypoint $\mathbf{x}$ are added to an orientation histogram of 36 bins covering 360 degrees, where each pixel is weighted by its gradient magnitude $m(x, y)$ and by a Gaussian-weighted circular window ($\sigma = 1.5 \times \sigma_{\mathbf{x}}$) related to the distance between the pixel and the keypoint. Peaks in this orientation histogram represent the dominant gradient direction of the keypoint $\mathbf{x}$. Multiple keypoints are created if multiple peaks exist.

Once the orientation of the keypoint is determined, the gradient orientation for each surrounding pixel is rotated relative to this keypoint orientation to achieve orientation invariance. The rotated samples are then smoothed by a Gaussian function and accumulated into multiple 8-bin orientation histograms over subregions as shown in Figure 2.6. In practice 4×4 subregions are used and a $4 \times 4 \times 8 = 128$-element feature vector is generated for each keypoint.

For keypoint matching, the Euclidean distance of the closest neighbour is compared to that of the second-closest neighbor. Normally all matches for which the ratio is greater than 0.8 are rejected, which eliminates 90% of the false matches while discarding less than 5% of the correct matches [Lowe, 2004].

SIFT has been shown to be invariant to image rotation, scale and change in illumination, which

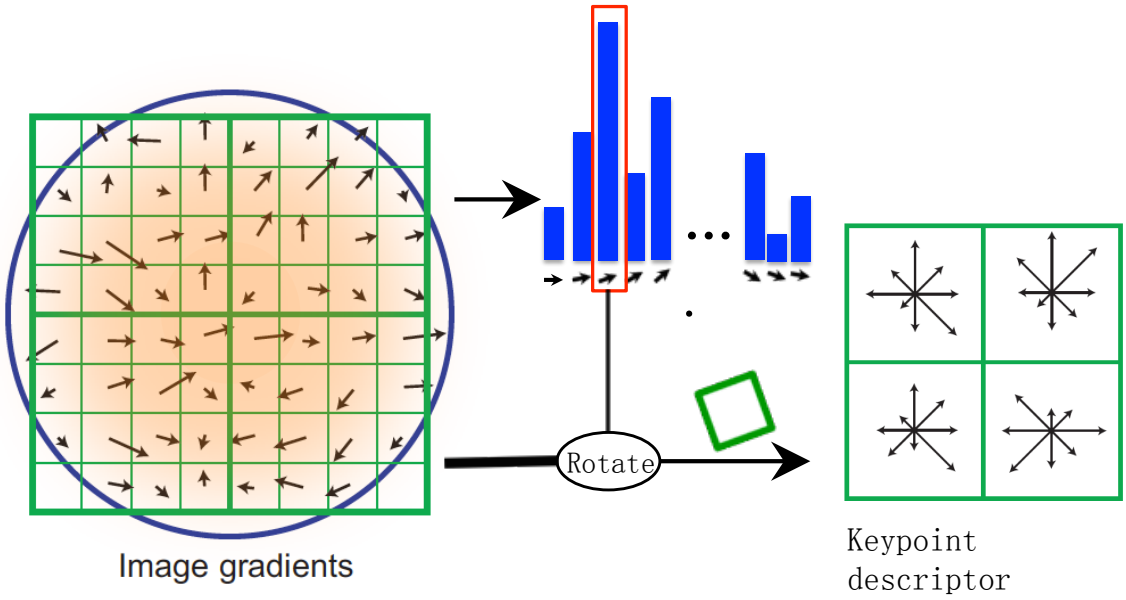


Figure 2.6 – A SIFT descriptor is created by first computing the gradient magnitude and orientation for each pixel around the keypoint, as illustrated on the the left. Each orientation added to a histogram is weighted by its magnitude and a Gaussian window, indicated by the overlaid circle. The peak (framed in red) in the histogram is used to rotate the coordinates of the descriptor and gradient orientation of each surrounding pixel. The rotated samples are then accumulated into multiple 8-bin orientation histograms over 4×4 subregions (only 2×2 subregions are shown for illustration purpose), as shown on the right. This figure is adapted from [Lowe, 2004].

is achieved by assembling a high-dimensional vector representing a local region gradients.

Speeded Up Robust Features (SURF)

SURF is designed to be a scale- and rotation-invariant local feature descriptor, for which the fast-Hessian detector serves as a keypoint detector [Bay et al., 2006].

SURF first assigns an orientation to each keypoint by first convolving two rectangular Haar wavelets with a set of pixels within a circle C around the keypoint, where the radius of the circle and size of the two wavelets as well as the sampling interval are all dependent on the scale σ at which the keypoint has been detected by the fast-Hessian detector. The wavelets responses are then weighted with a Gaussian and represented as points in a vector space, with x-response generated by the vertical Haar wavelet and y-response generated by the horizontal Haar wavelet. All points within a rotating orientation window covering an angle of 60° are

summed to form a local orientation vector for each of the all six positions of the sector (see left on Figure 2.7). The local orientation vector with the largest magnitude gives the dominant orientation of the keypoint.

As for SIFT, a local coordinate system is therefore defined by the scale and orientation of the keypoint and used to construct a square region centered around the keypoint. This square region is then split up into 4×4 subregions and the following feature vector is computed over 25 regularly spaced sample points at each subregion:

$$\left[\sum d_x, \quad \sum d_y, \quad \sum |d_x|, \quad \sum |d_y| \right]$$

Here d_x and d_y are the Haar wavelet responses in horizontal and vertical direction respectively. The SURF descriptor is then the vector consisting of $4 \times 4 \times 4 = 64$ components of all feature vectors from each subregion. For keypoint matching, the Eudlidean distance between descriptor vectors is calculated. A matching pair is detected if its distance is closer than 0.7 times the distance of the second nearest neighbour [Bay et al., 2006] [Baumberg, 2000].

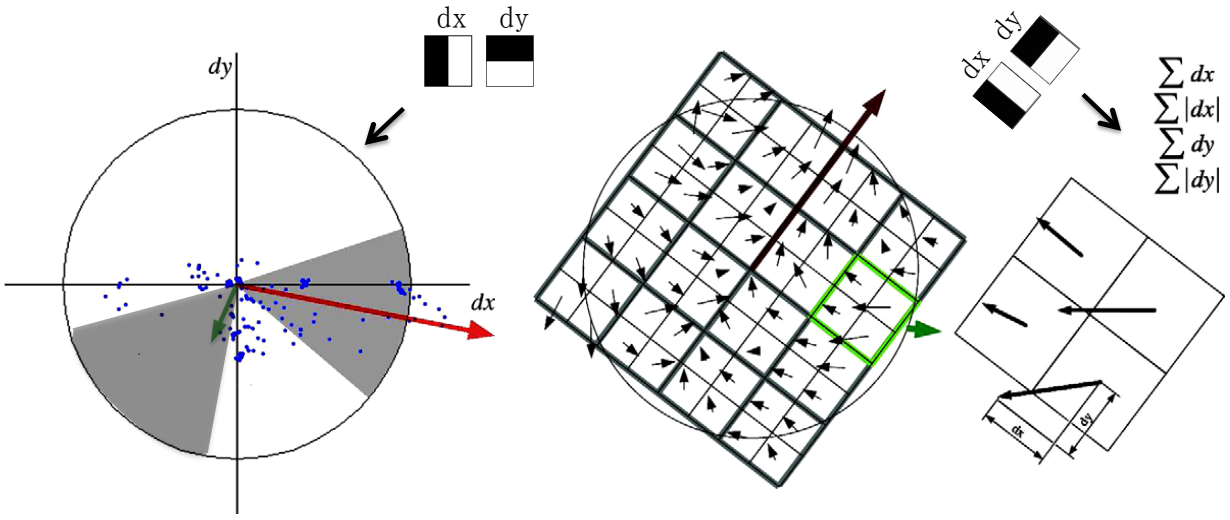


Figure 2.7 – Haar wavelets responses of all pixels within a circle around the keypoint are represented as points in a vector space. Points within a rotating sector are summed up to form a local orientation vector for each position of the section; two such vectors are shown as examples on the left. A square region centered around the keypoint is constructed under a local coordinate system. This square region is then split up regularly into 4×4 subregions and a feature vector is computed over 25 regularly spaced sample points at each subregion (only 4 sample points are shown for illustration purpose), as illustrated on the right. Two pairs of Haar wavelets used for convolution are also shown. This figure is adapted from [Bay et al., 2008].

Histograms of Oriented Gradients (HOG)

The HOG feature descriptor has been introduced by Dalal and Triggs [Dalal and Triggs, 2005] for pedestrian detection by computing histogram of oriented gradients.

It first computes the gradient orientation and magnitude for each pixel by applying 1-D derivative mask. Then a local spatial region in the shape of rectangle/square or radial (log-polar sector) is defined as a cell. Each pixel within a cell casts a vote weighted by its gradient magnitude to form an orientation-based histogram, where the histogram spans $0^\circ \sim 180^\circ$ or $0^\circ \sim 360^\circ$.

These cells are further grouped into blocks, of which two classes of block geometries are used. Rectangular blocks (R-HOG) consists of rectangular cells and circular blocks (C-HOG) partitioned into cells in log-polar shape. In order to account for the influence of local variations in illumination and foreground-background contrast on gradient strength, local contrast normalization is performed on block level. Concatenating all normalized cell histograms from all of the block regions results in a HOG descriptor. In the experiments by Dalal and Triggs, they found 3x3 cell blocks of 6x6 pixel cells with 9 histogram bins of gradient performed best for pedestrian detection. For object detection, a support vector machine (SVM) classifier is applied using the HOG descriptor as feature vector.

2.2.3 Filter-based descriptors

In image processing, filters are mainly used to reduce noise and/or extract useful image structures. Filters like Gaussian filters, Laplacian filters, and Sobel filters are widely used in image blurring/sharpening, edge detection, and feature extraction. Other filters are designed to extract specific spatial information from the image, for instance, steerable filters work well for rotations and Haar-like wavelet filters can be used to encode information.

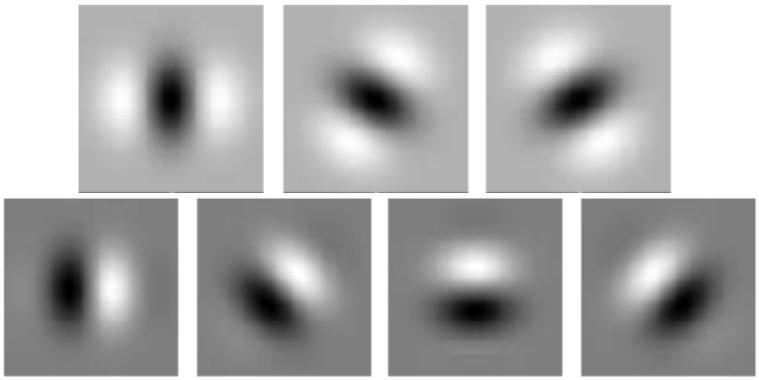


Figure 2.8 – Illustration of G_2 and H_2 quadrature pair basis filters.

Steerable Filters

Steerable filters [Freeman and Adelson, 1991] are a linear combination of filters with arbitrary orientations, which can be expressed by the following equation:

$$F^{\theta} = \sum_{i=1}^{N} k_i(\theta) F_i$$

Here F_i is the i-th basis filter and $k_i(\theta)$ is the i-th coefficeint in orientation θ. As an example, steerable filters are designed in quadrature pairs, which have same frequency response but differ in phase by 90°, to allow adaptive control over phase as well as orientation. Figure 2.8 illustrates the G_2 and H_2 quadrature pair basis filters. The formulas for steerable filter pairs (G_2 and H_2) based on the second derivative of a Gaussian are given as [Freeman and Adelson, 1991]:

$$\begin{aligned} G_2(x,y) &= 0.9213(2x^2 - 1)e^{-(x^2+y^2)} \\ H_2(x,y) &= (-2.205x + 0.9780x^3)e^{-(x^2+y^2)} \end{aligned} \tag{2.12}$$

Applications of steerable filters include edge detection, oriented texture analysis and shape from shading.

Haar-like Filters

To represent the regional intensity differences at different orientations, Haar filters are computed over the ROI (Region of Interest) in the image [Viola and Jones, 2002]. Haar-like features are effective to capture the regional information, and orientation and scale of a voxel.

Haar-like features are so called because they are computed in a similar fashion to the coefficients in Haar wavelet transforms. Several types of filters are used to capture the differences between local regions as proposed by Viola [Viola and Jones, 2002] and Kuranov et al. [Kuranov et al., 2002], which are shown in Figure 2.9.

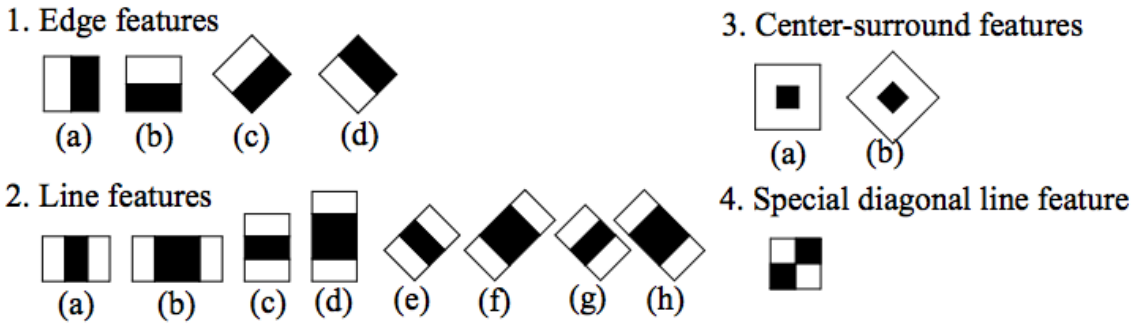


Figure 2.9 – Illustration of Haar-like features Viola [Viola and Jones, 2002].

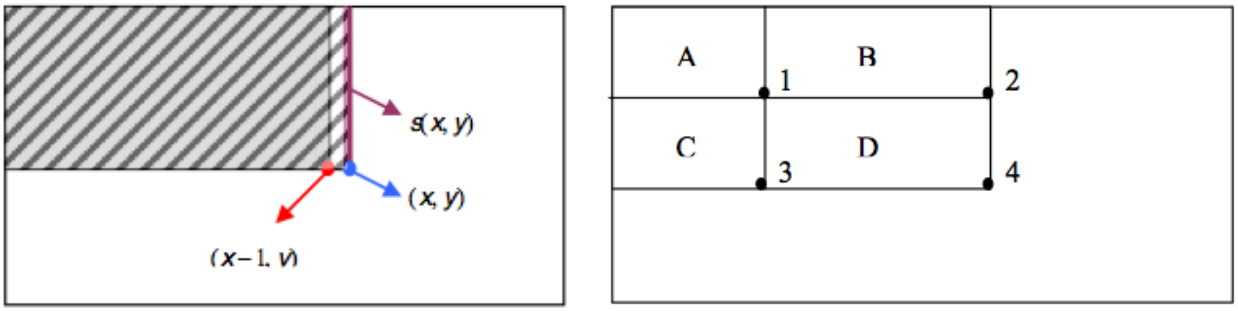
Four edge features, eight line features, two center-surround features, and a special diagonal line feature are shown in Figure 2.9. For instance, the two-rectangle features in Figure 2.9 denote the difference between the sums of voxels in two adjacent same sized rectangular regions. The three-rectangle features in Figure 2.9 represent the difference between two outside regions (white) and the inside region (black).

To calculate a rectangle feature for an area, a brute-force algorithm would have a computational complexity of $O(mn)$. Viola and Jones [Viola and Jones, 2002] borrowed an idea from an application in texture mapping [Crow, 1984] to decrease the computational complexity to $O(1)$: an intermediate representation, which is called integral image $II(x, y)$, for the source image $I(x, y)$ is generated first. The integral image at location (x, y) contains the sum of the pixels

above and to the left of point (x, y), inclusive:

$$II(x, y) = \sum_{x' \leq x, y' \leq y} I(x', y')$$

As shown in Figure 2.10(a), $II(x, y)$ can be computed as the sum of the integral image at $(x - 1, y)$ and the cumulative row sum of pixels above point (x, y).



(a) **(b)**

Figure 2.10 – (a)The value of the integral image (area in pattern) at point (x, y) is the sum of the integral image (area in grey shadow) at point $(x - 1, y)$ and the row sum $s(x, y)$ of pixels above, inclusive. (b)The sum of the pixels within D can be calculated as $D = II(4)+II(1)-II(2)+II(3)$, where $II(1)$is the integral image at location 1 that represents the sum of the pixels in area A. $II(2)$ is for $A + B$, $II(3)$ is for C and $II(4)$ is for $A + B + C + D$. Figures from [Viola and Jones, 2002].

The integral image $II(x, y)$ can then be calculated in one pass over the source image using the following equations:

$$\begin{aligned} s(x, y) &= s(x, y - 1) + I(x, y) \\ II(x, y) &= II(x - 1, y) + s(x, y) \end{aligned} \tag{2.13}$$

For the area above and to the left of the image $I(x, y)$, $s(x, y < 0) = 0$, and $II(x < 0, y) = 0$. The rectangle features in Figure 2.9 are actually the difference between the intensity sum of the rectangular areas. Once we obtained the integral image, the sum of rectangular areas can be easily computed as shown in Figure 2.10(b).

2.2.4 Binary descriptors

In the field of keypoint descriptor, SIFT has been widely used in various applications ([Lowe, 1999], [Lowe, 2004]) and has been regarded as the best-in-class for some time. However, along with the recent development of mobile devices, new vision algorithms are needed for faster performance to serve for applications such as panorama stitching, object detection, navigation and tracking. Binary descriptors have started to attract attention due to the fact that they are very fast to compute and to match. These approaches build a binary string by comparing pairs of image intensities, which can still efficiently represent a keypoint. For point matching and descriptor comparision, the Hamming distance can replace the usual Euclidean distance to significantly increase speed.

Binary Robust Independent Elementary Features (BRIEF)

BRIEF [Calonder et al., 2010] [Calonder et al., 2012] is a local binary descriptor which is very fast both to compute and to match. It delivers good recognition performance with fast matching using Hamming distance. However, it is sensitive to image distortions and transformations.

BRIEF creates a bit string out of the response by applying τ for patch P of size $S \times S$ as follows:

$$\tau(P; \mathbf{x}_a, \mathbf{x}_b) := \begin{cases} 1 & \text{if } I(\mathbf{x}_a) < I(\mathbf{x}_b) \\ 0 & \text{otherwise} \end{cases}$$

Here $I(\mathbf{x}_a)$ and $I(\mathbf{x}_b)$ are the pixel intensities in a smoothed version of P at location $\mathbf{x}_a$ and $\mathbf{x}_b$. Applying τ for n_d pairs of $\mathbf{x}_a$ and $\mathbf{x}_b$ on patch P defines a n_d-dimensional bit string. n_d is ususally set to 128, 256 or 512. There exist many options for selecting the test locations $\mathbf{x}_a$ and $\mathbf{x}_b$ in the patch. The spatial arrangement of the pair of the points, displayed in Figure 2.11, can be described as follows:

(a) $(\mathbf{x}_a, \mathbf{x}_b) \sim$ i.i.d. Uniform $(-\frac{S}{2}, \frac{S}{2})$;

(b) $(\mathbf{x}_a, \mathbf{x}_b) \sim$ i.i.d. Gaussian $(0, \frac{1}{25}S^2)$;

(c) $\mathbf{x}_a \sim$ i.i.d. Gaussian $(0, \frac{1}{25}S^2)$, $\mathbf{x}_b \sim$ i.i.d. Gaussian $(\mathbf{x}_a, \frac{1}{100}S^2)$;

(d) Randomly sampled from discrete locations of a coarse polar grid introducing a spatial quantization;

(e) $\forall i : \mathbf{x}_a(i) = (0, 0)^T$ and $\mathbf{x}_b(i)$ takes all possible values on a coarse polar grid containing n_d points.

Here S is the size of the patch P. Among them, the experiment result using the (e) strategy is worse compared to all random designs (a) to (d), while (b) shows a small advantage over others. Matching the BRIEF feature vectors mainly involves computing Hamming distances which is very fast.

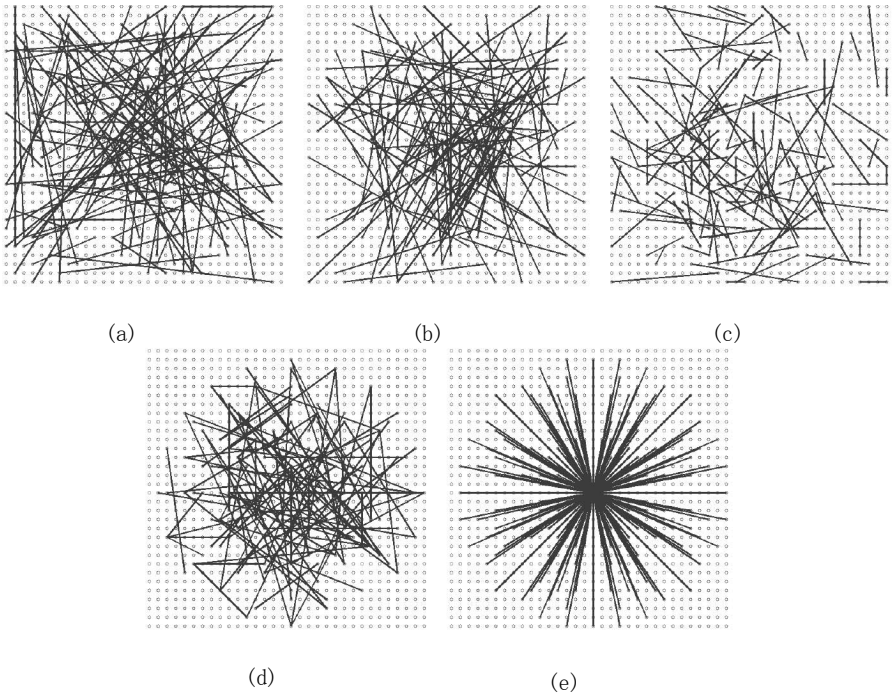


Figure 2.11 – Different approaches to choosing the test locations. All except (d) are selected by random sampling. Showing 128 tests in every image [Calonder et al., 2010]

BRIEF is scale invariant but has limited invariance to rotation. A variant of BRIEF is developed to achieve full invariance to rotation and scaling by building a database of templates, which is referred as D-BRIEF. Matching a new patch against a database then means matching against all rotated and scaled versions, which still remains much faster than using SIFT or SURF. To increase the stability and repeatability of the descriptor and increase the speed, the image

is smoothed by a box filter, which is computed using integral images (as shown in Figure 2.10(a)) [Calonder et al., 2012] [Heikkilä et al., 2009].

Orientated FAST and Rotated BRIEF (ORB)

ORB [Rublee et al., 2011] builds on the well-known FAST keypoint detector and binary descriptor BRIEF, which is rotation invariant and robust to noise.

ORB adds an orientation component to FAST for two purposes: Firstly, FAST gives strong response to edges. By employing a Harris corner measure (discussed in section 2.1.2) ORB can rank the keypoints detected by FAST and eliminate unstable points. Secondly, the orientation computed for each keypoint is used to orient the descriptor to achieve rotation invariance. An intensity centroid based on the moments of the intensity distribution within a local patch is employed to define the orientation of keypoint. The moments of the intensity distribution within a local patch surrounding a keypoint are defined as:

$$m_{ij} = \sum_{x,y} x^i y^j I(x,y)$$

The the centroid C and the orientation θ are defined as:

$$C = (\frac{m_{10}}{m_{00}}, \frac{m_{01}}{m_{00}}), \quad \theta = \tan^{-1}(\frac{m_{01}}{m_{10}})$$

The orientation of a keypoint is then used to steer BRIEF by multiplying the matrix of binary tests with the rotation matrix corresponding to θ. However, the resulting steered BRIEF descriptors have a lower variance because the oriented corner keypoints present a more uniform appearance. Therefore ORB conducts a greedy search for choosing a subset of binary tests that both have high variance and are uncorrelated. The resulting descriptor has a large variance and lower pairwise-correlation.

ORB has been reported to be up to 48 times faster than DoG in detection and up to 100 times faster than SIFT [Miksik and Mikolajczyk, 2012].

Binary Robust Invariant Scalable Keypoints (BRISK)

The BRISK descriptor [Leutenegger et al., 2011] is a combination of a DoG-like scale-space keypoint detector and a BRIEF-like binary descriptor, which is scale and rotation invariant. The BRISK feature detection, descriptor composition and keypoint matching are constructed in a modular fashion, which allows the use of the BRISK detector in combination with any other keypoint descriptor and vice versa. The BRISK detector searches the local maxima in a DoG-like scale-space (see subsection 2.1.5) using the FAST score as a measure for saliency to achieve scale invariance.

The novel concept proposed in the BRISK descriptor is to make use of a pattern to sample the neighborhood of the keypoint. As illustrated in Figure 2.12, the BRISK descriptor defines N locations equally spaced on circles concentric with the keypoint in a pattern that looks like a daisy. Gaussian smoothing with the standard deviation σ_i proportional to the radius of the respective circle are applied. The pattern is positioned and scaled accordingly for each keypoint. For estimating the dominant direction of the keypoint the local gradient for one of the $N \cdot (N-1)/2$ sampling-point pairs $(\mathbf{x}_i, \mathbf{x}_j)$ is defined as:

$$g(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_j - \mathbf{x}_i) \cdot \frac{I(\mathbf{x}_j, \sigma_j) - I(\mathbf{x}_i, \sigma_i)}{\|\mathbf{x}_j - \mathbf{x}_i\|^2}$$

A subset of short-distance pairings $\mathcal{S}$ and another subset of long-distance pairings $\mathcal{L}$ of L pairs are also defined:

$$\mathcal{S} = \{(\mathbf{x}_i, \mathbf{x}_j) \| \mathbf{x}_j - \mathbf{x}_i\| < \delta_{max}\} \quad \text{and} \quad \mathcal{L} = \{(\mathbf{x}_i, \mathbf{x}_j) \| \mathbf{x}_j - \mathbf{x}_i\| > \delta_{min}\}$$

The threshold distance δ_{max} and δ_{min} are set according to the scale of the keypoint. Assuming local gradients annihilate each other, the dominant direction of a keypoint is only computed over all long-distance pairs:

$$g = \begin{pmatrix} g_x \\ g_y \end{pmatrix} = \frac{1}{L} \cdot \sum_{(\mathbf{x}_i, \mathbf{x}_j) \in \mathcal{L}} g(\mathbf{x}_i, \mathbf{x}_j)$$

Upon the determining of the direction, BRISK rotates the sampling pattern by $\alpha = \arctan2(g_y, g_x)$ around the keypoint and extracts the descriptor by comparing the intensities between all short-distance pairs $(\mathbf{x}_i^\alpha, \mathbf{x}_j^\alpha) \in \mathcal{S}$. The bit-string contains 512 bits.

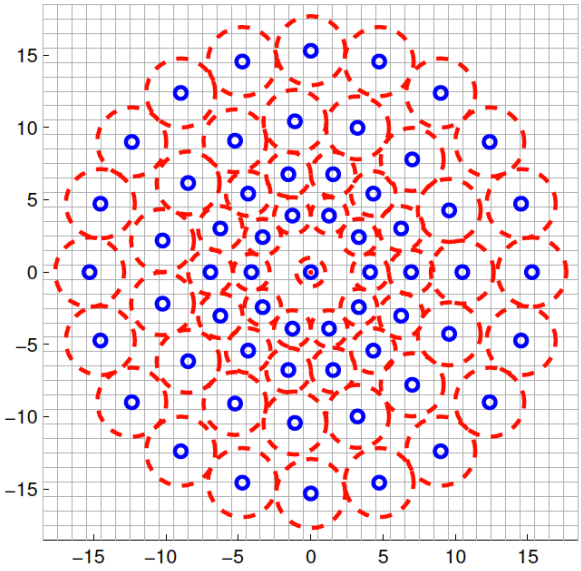


Figure 2.12 – The BRISK sampling pattern with $N = 60$ points: the small blue circles denote the sampling locations; the bigger, red dashed circles are drawn at a radius corresponding to the standard deviation of the Gaussian kernel used to smooth the intensity values at the sampling points [Leutenegger et al., 2011].

Fast Retina Keypoint (FREAK)

Inspired by the human visual system, Alahi et al. [Alahi et al., 2012] proposed a cascaded binary descriptor FREAK which, similar to BRISK, uses a sampling pattern to simulate the receptive field of the retina in the human visual system. The sampling pattern is also circular but has higher density of points near the center, as illustrated in Figure 2.13. The circles, representing the standard deviation of the Gaussian kernel applied to their corresponding centre, overlap with each other and exponentially change in size. The overlapping of the receptive field adds redundancy to the sampling that yields more discriminative power. The FREAK descriptor is extracted in the same way used by other previous discussed binary descriptors by comparing the smoothed intensities pairwise. FREAK uses 43 receptive fields which leads to potentially around $n = 1000$ pairs. In order to have discriminant and loosely correlated features, a selection

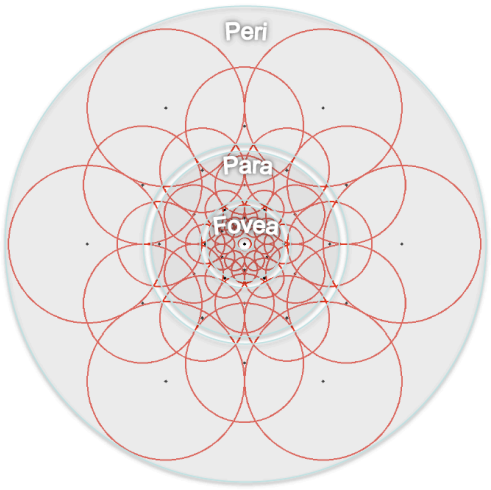


Figure 2.13 – Illustration of the FREAK sampling pattern. Each circle represents a receptive field where the image is smoothed with its corresponding Gaussian kernel [Alahi et al., 2012].

rule is learned from a training set of $k = 50,000$ keypoints. A $k \times n$ matrix of descriptors is constructed and the mean and variance of each column is computed. Then the columns are ranked according to their variance and the column with the lowest correlation with other selected ones is iteratively selected. The selected 512 pairs are further divided into four subsets and the first subset is found to involve mainly the peripheral receptive fields whereas the last subset is more central. In order to estimate the rotation of the keypoint, FREAK sums the local gradients similar to BRISK (subsection 2.2.4), but over pairs with symmetric receptive fields with respect to the center.

The matching procedure starts by searching with the first 128 bits and only carries on if the distance is smaller than a given threshold. The cascading comparisons accelerate the matching and as the result, FREAK claims to be faster than BRISK in extracting and matching feature points.

Local Binary Patterns (LBP)

Local binary patterns (LBP) [Ojala et al., 1994] is a simple yet efficient operator to describe local image pattern, and it has achieved impressive classification performance on representative texture databases.

The original LBP operator [Ojala et al., 1994] compares the intensity of each point $\mathbf{x}_i$ in 3×3 neighborhood with the intensity of the central pixel $\mathbf{x}_c$ and considers the result as a binary number. Ojala *et al.* [Ojala et al., 2002] extended the LBP operator to use a circular neighborhood with any radius and any size of the neighborhood points. Bilinearly interpolation is used for non-interger pixel coordinates of the neighborhood points. The value of the LBP code as follows:

$$LBP_P = \sum_{i=0}^{P-1} s(\mathbf{x}_i)2^i, \qquad s(\mathbf{x}) = \begin{cases} 1 & \text{if } I(\mathbf{x}) \geq I(\mathbf{x}_c) \\ 0 & \text{otherwise} \end{cases} \tag{2.14}$$

Here P is the number of pixels in the neighbourhood of $\mathbf{x}_c$. The LBP code value is computed based on the intensity differences, by assuming that differences are independent of the intensity of the central point, the LBP code approximates the joint distribution of the gray levels of P pixels and achieves gray scale (brightness) invariance.

The LBP refers to the local appearance with limited number of discontinuities in the circular presentation of the pattern as a 'uniform' pattern. Figure 2.14 illustrates a few examples of uniform patterns. These uniform patterns provide over 90% of the texture patterns and are implemented with a lookup table to help achieving rotation invariance. The discrete occurrence histogram of the uniform patterns computed over an image or a region estimates the distribution of the local structures, and has been shown to be a very powerful texture feature.

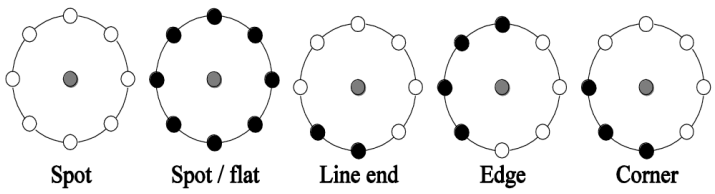


Figure 2.14 – Examples of uniform patterns of LBP. Figures from [Ojala et al., 1994].

Table 2.2 gives an overview of the descriptors we are reviewing in this section and their most important properties.

Table 2.2 – Summary of feature descriptors.

	Descriptor	Typical size	Scale invariant	Rotation invariant	Blur invariant	Brightness invariant	Sampling grid	Similarity	Speed*
Patch	non-local	various	×	$< 10°$	limited	limited	square	SSD/NCC/..	NA[1]
Distribution	SIFT	128	✓	✓	✓	✓	square	Euclidean ratio	449/2095
	SURF	64	✓	✓	✓	✓	circular	Euclidean ratio	117/390
	HOG	81	×	×	✓	✓	3 types[2]	svm	fast
Filter	Steerable	various	✓	✓	×	×	square	Mahalanobis	NA[1]
	Haar-like	various	×	×	limited	✓	rectangle	Euclidean	NA[1]
Binary	BRIEF	256bits	✓	limited	✓	✓	random pairs	Hamming	3.8/370
	ORB	256bits	✓	✓	✓	✓	random pairs	Hamming	4.2/370
	BRISK	512bits	✓	✓	✓	✓	circlular	Hamming	10.6/524
	FREAK	64bits	✓	✓	✓	✓	retinal	Hamming	7/350†
	LBP	various	×	✓	×	✓	square/circle	classification	NA[1]

Note: $*$ Computation time/matching time in *ms.* of the descriptors for 1000 SURF points, according to [Miksik and Mikolajczyk, 2012].

† FREAK takes around two third of the time for extracting and matching comparing to BRISK, according to [Alahi et al., 2012]. The listed times for FREAK are estimated according to this.

[1] Not applicable.

[2]square, rectangle and circle

2.3 Machine learning

Machine learning is a branch of artificial intelligence which uses statistical and computational methods to learn information and improve its performance automatically through learning without being explicitly programmed [Samuel, 1959]. Machine learning algorithms make predictions or decisions using the statistical model it builds based on the input data.

Machine learning usually refers to a system that learns in response to input and external feedback and then performs updating in order to improve performance. The procedure of learning and updating depends on the algorithm being used. Commonly used algorithms are supervised learning, unsupervised learning or a combination of both. Frequently used algorithms include artificial neural networks, decision tree learning, Bayesian learning, genetic algorithms and so on. Machine learning algorithms have been applied effectively for certain types of learning tasks, and a theoretical understanding of learning is beginning to emerge [Mitchell, 1997]. A well designed machine learning approach needs to choose an algorithm suitable for the target problem and data set, decide how to represent the target and include prior knowledge.

2.3.1 Support Vector Machines

The Support Vector Machine (SVM) [Vapnik, 1995] [Vapnik, 1998] is popular for classification, regression and pattern recognition because it can provide an accurate and global optimal solution for such problems. Basically, a SVM arranges the original input space into a high-dimensional feature space according to the kernel function and then define a hyperplane in the feature space to separate classes by minimising the error on the training set and maximize the margin between classes. Hence the generalization performance does not drop significantly even when the training data is scarce [Abe, 2005].

For the two-class classification problem, the hyperplane can be defined as a set of transformed points X :

$$f(x) = W^T \phi(X) + b = 0 \tag{2.15}$$

Here, W is surface normal to the hyperplane, b is a bias term, and $\phi(\cdot)$ denotes the non-linear function which maps the input points into a high dimensional space. In this space, two surfaces can be described by:

$$W^T\phi(X) + b = 1 \tag{2.16}$$

and

$$W^T\phi(X) + b = -1 \tag{2.17}$$

Both are chosen so that the perpendicular distance between them is maximized. The distances between the hyperplane and any one of the two surfaces are same, known as margin. The two surfaces are determined by a subset of the data points, which are called support vectors, as illustrated in Figure 2.15.

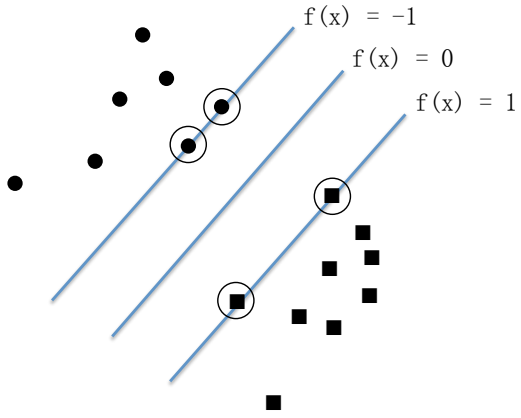


Figure 2.15 – Support vectors are indicated by the circled points, which are lying on or close to the two surfaces.

Obviously, the margin is of the magnitude of $\frac{1}{\|W\|}$. Maximizing the margin is equivalent to minimizing $\frac{1}{2}\|W\|^2$, which can be solved by using Quadratic Programming optimization.

The following constraint is added to ensure that every points i is outside the margin region,

$$y_i(W^T\phi(X) + b) \geq 1, 1 \leq i \leq N \tag{2.18}$$

Here $y_i \in \{-1, 1\})$ is the associated label of i.

A positive slack variable ζ_i is introduced to allow for misclassified points when the data cannot

be fully separated by a hyperplane,

$$y_i(W^T\phi(X)+b) \geq 1-\zeta_i, 1 \leq i \leq N \tag{2.19}$$

Therefore with the relaxing of the margin constraint, the optimization is now expressed as following:

$$\min \left\{\frac{1}{2}\|W\|^2 + C\sum_{i=1}^{N}\zeta_i\right\} \tag{2.20}$$

subject to

$$y_i(W^T\phi(X)+b) \geq 1-\zeta_i, \zeta_i \leq 0 \tag{2.21}$$

Here C is a user-defined regularization parameter controlling the trade-off between the slack variable penalty and the margin. So with the help of the Lagrange multiplier $\alpha(\alpha > 0)$ and $\mu(\mu > 0)$, the constraints can be included in equation 2.21:

$$L_P = \frac{1}{2}\|W\|^2 + C\sum_{i=1}^{N}\zeta_i - \sum_{i=1}^{N}\alpha_i\left[y_i(W\cdot X_i+b)-1+\zeta_i\right] - \sum_{i=1}^{N}\mu_i\zeta_i \tag{2.22}$$

The optimal parameter W, b and α are needed to be found in order to minimize L_P. By using Wolfe dual form, Vapnik [Vapnik, 1995] [Vapnik, 1998] coverted the minimizing problem to a maximization problem,

$$L_D = \sum_{i=1}^{N}\alpha_i - \frac{1}{2}\sum_{i,j}^{N}\alpha_i\alpha_j y_i y_j k(x_i, x_j) \tag{2.23}$$

subject to $0 \leq \alpha_i \leq C$ and $\sum_{i=1}^{N}\alpha_i y_i = 0$

Here kernel function $k(x_i, x_j) = \phi(x_i)\cdot\phi(x_j)$ and W can be calculated as $W = \sum_i^N \alpha_i y_i x_i$. Quadratic programming can then be used to find α to maximize Equation 2.23. For the kernel function $k(x_i, x_j)$, typical options are linear, polynomial, hyperbolic tangent or radial basis functions. By applying the kernel function, the hyperplane can be constructed in the transformed feature space.

SVMs have been shown to perform well in multiple applications, including fault detection,

fraud detection, handwritten character recognition, object detection and recognition, text classification and medical image classification [Smola and Schölkopf, 2004]. Powell et al. [Powell et al., 2008] included probability map values as input features for the SVM based segmentation and has shown that the machine learning methods outperformed other probability-based methods. The probability map was constructed by calculating the probability of a neuroanatomical structure being located at a voxel location in the atlas space across all subjects in a training set. However, they also pointed out that the algorithms may not generalize to regions of greater degree of variability.

Bauer et al. [Bauer et al., 2011] proposed a fully automatic method for brain tissue segmentation, which combined SVMs using multispectral intensities and textures with subsequent hierarchical regularization based on conditional random fields (CRF). They first classified the image into tumor and healthy tissues using CRF, for which the data term is computed from the label output of a RBF kernel based SVM for each voxel. In the second stage they further classified the image into different tumor and healthy subregions.

2.3.2 AdaBoost

Among the algorithms used for classification, AdaBoost [Freund and Schapire, 1997] is a weighted voting algorithm that selects weak classifiers to combine them into a strong classifier. A weak learner is any statistical classifier that classifies better than random. Initially, every training example is assigned an equal weight. If the example is classified incorrectly its associated weight will be increased in the next iteration to force the weak learner to focus on the hard examples.

AdaBoost has been a successful measure to handle a large amount of data [Morra et al., 2008] [Heisele et al., 2003] which is a very unmanageable task to any robust classifiers. It has been reported to be able to provide a fast and accurate solution in face recognition and other areas [Viola and Jones, 2002] [Papageorgiou and Poggio, 1999] [Savio et al., 2009]. The most basic property of AdaBoost is that the training error theoretically converges exponentially towards

zero [Freund and Schapire, 1997]. It has been shown that AdaBoost can boost a weak learning algorithm into a strong learning algorithm with an arbitrarily low error rate.

Initially, a weight $w_1(i) = \frac{1}{N}$ (N is the size of the dataset) is assigned to each data point x_i. A weak learner $h_j(j = 1..K)$ consists of one of the features, a threshold and a Boolean function. At each iteration t ($t = 1..T$) in the training phase, AdaBoost finds the weak learner $h_t(x)$ with the lowest error ϵ_t comparing to label y_i,

$$\epsilon_t = \min_{j=1..k}(\sum_{i=1}^{N} w_t(i)|h_j(x_i) - y_i|)$$

and assigns coefficient α_t to it. The lower the error ϵ_t, the greater the coefficient α_t:

$$\alpha_t = (1/2)\log((1 - \epsilon_t)/\epsilon_t)$$

The real power of AdaBoost is that after each iteration t the weights $W_t(i)$ for misclassified data points are increased. So, AdaBoost can focus on the examples with high weights, which seems difficult to classify for the weak learners.

$$w_{t+1}(i) = w_t(i)\exp(-\alpha_t y_e h_t(x_e))/(2\sqrt{\epsilon_t(1 - \epsilon_t)})$$

This procedure is repeated for K iterations. At last AdaBoost uses the weighted vote to combine all weak learners into a single final classifier and the final hypothesis is:

$$H(x) = \begin{cases} 1, & (1/(\exp(-\sum_{t=1}^{T}(\alpha_t h_t(x))) + 1) > 0.5) \\ 0, & (1/(\exp(-\sum_{t=1}^{T}(\alpha_t h_t(x))) + 1) < 0.5) \end{cases}$$

The convergencee of the training process is important. Freund and Schapire [Freund and Schapire, 1997] showed that the rate of misclassified samples in the training set, i.e. the

training error of the final hypothesis, has an upper bound as:

$$\prod_t \left[2\sqrt{\epsilon_t(1-\epsilon_t)}\right] = \prod_t \sqrt{1-4\gamma_t^t} \leq \left[-2\sum_t \gamma_t^2\right]$$

Therefore, if each weak hypothesis is slightly better than random, then the training process converges exponentially fast. When the learned AdaBoost classifier is used to classify unseen samples, the generalization error is shown by Freund and Schapire [Freund and Schapire, 1997] to be at most :

$$\hat{Pr}\left[margin(x,y) \leq \theta\right] + \tilde{O}\left[\sqrt{\frac{d}{m\theta^2}}\right]$$

for $\theta > 0$ with high probability. The bounds on training and generalisation error given above prove that AdaBoost can efficiently assemble weak learners into a strong classifier.

Practically, AdaBoost is simple to implement as it has very few parameter to tune. In addition, no prior knowledge is required for learning and the most representative features will be selected during the learning. AdaBoost is equipped with theoretical properties for the convergence of the training error and generalization error. It is also reported that AdaBoost usually does not overfit even after training for thousands of iterations.

AdaBoost has been tested and used in many applications, and along with its applications in different area, various versions of AdaBoost are proposed. These variants differ in the update scheme of the weights or use some other classifiers as the weak learner. However, the performance of AdaBoost is dependent on the data and weak classifiers. The consistency of the training set is crucial for the quality of the classification. Insufficient data or too complex weak hypotheses can result in failure of AdaBoost. Boosting also seems sensitive to noise. Another drawback of AdaBoost is that the training is quite time-consuming. At each iteration, all features are tested on all training samples which makes the training time proportional to the size of the features and training set. The features, the size of features and training set need to be well defined to achieve good classification within a sensible training time. Haar-like features are calculated to effectively encode domain information to represent each sample. .

Qian [Qian et al., 2006] presented a cardiac boundary tracking method for 2D tagged MRI time

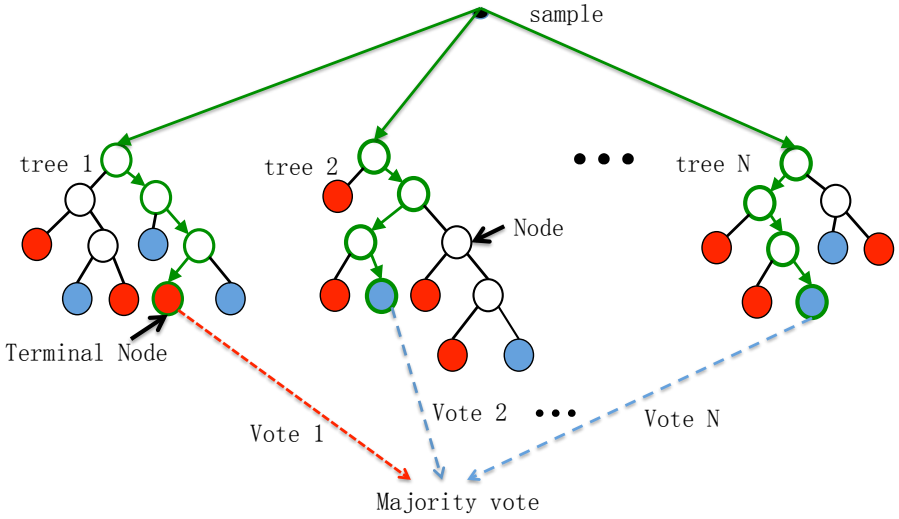


Figure 2.16 – A random forest of N trees. Color filled terminal nodes give predictions for two classes. When a test sample runs down each individual tree, marked in green, majority votes of terminal nodes are taken as the final prediction for this sample.

sequences, which integrates the motion and the static local appearance features and generates accurate boundary criteria via Adaboost. Furthermore, through boosting error analysis, the posterior probability density function of the shape model are calculated based on the confidence ratings. A principle component analysis (PCA) subspace of shape is also used constrains the shape variation and lower the dimensionality.

2.3.3 Random Forests

Random Forests were introduced by Breiman and Cutler [Breiman, 2001] in 2001 by combining Breiman's bagging approach with random selection of features [Ho, 1999] [Amit and Geman, 1997], and have been widely used in the fields of classification, clustering and regression ever since. A random forest is a classifier consisting of a collection of decision trees, for which Breiman added an additional random feature selection procedure to grow a tree. When a new sample is passed down each individual tree, majority votes of terminal nodes are taken as the final prediction, as illustrated in Figure 2.16.

To construct a random forest of N trees, draw N sets of bootstrap samples with replacement from the original data and grow a decision tree for each set. Random forest randomly selects

a subset of features to best split the nodes in the forest and the criterion for selecting such features is based on the Gini index, which measures the probability that an sample would be incorrectly classified if it was randomly labeled according to the distribution of the labels in current node. The node stops splitting when Gini index reaches zero, i.e. all samples on this node share the same label. Each tree is fully grown and not pruned in a random forest.

In their report, Breiman and Curtler defined an upper bound for the generalization error,

$$PE^* \leq \bar{\rho}(1 - s^2)/s^2$$

where s is the strength of the individual tree and $\bar{\rho}$ is the correlation between trees. Their empirical results indicated that random forests with lower generalization error have lower correlation between classifiers and higher strength. It also hints that the injecting randomness in tree construction works well for the classifier by reducing the correlation $\bar{\rho}$. In particular, the generalization error can be estimated by the out-of-bag error rate, which is obtained by aggregating the prediction of the data not in the bootstrap sample (which is so called 'out-of-bag' data). In each tree, about one-third of the samples are left out. Breiman also shows that using the out-of-bag error can also monitor strength and correlation, as well as measure the importance of the features.

The combination of randomized node optimization and bagging makes Random Forest a more accurate classifier, and more robust with respect to noise and outlier. Additionally, it is faster than bagging or boosting.

Random forests have been widely used in biomarker discovery and disease diagnosis. Criminisi and Shotton [Criminisi and Shotton, 2013] summarised the developments and applications on top of the general random forest model, ranging from kinect-based player segmentation, keypoint recognition to brain lesions diagnosis. Among these applications, Criminisi *et al.* [Criminisi et al., 2013] discussed the use of regression forests for the automatic detection and simultaneous localization of multiple anatomical regions within CT and MRI images. They formulated the anatomy localization task as a continuous multivariate parameter estimation

problem. For organ detection, a tree-based continuous model allows the position of the walls of the bounding box containing each organ to be estimated simultaneously. The multi-tree based model can also incorporate multi-atlas information and achieve anatomy localization in only a few seconds with lower error, comparing to atlas-based approaches.

Lebedev *et al.* [Lebedev et al., 2014] applied random forest to discriminate Alzheimer's disease from mild cognitive impairment and health controls on MRI images. A feature reduction approach is applied to optimize the models. They revealed the significantly higher accuracy of random forest classifier comparing to other classifiers. The performance is further improved by combing morphometric measurements and other metadata.

2.3.4 Regression

Regression analysis [Fisher, 1922] is a statistical procedure for estimating the relationships between one or more independent variables (multiple regression) and the response variable. This includes linear regression and non-linear regression. Regression analysis is widely used for prediction, forecasting, and discovering the strength of the relationship between variables.

A regression model f fits the dependent variable $\mathbf{Y}$ to a mathematical function of independent variables $\mathbf{X}$ and unknown coefficients β as follows:

$$\mathbf{Y} = f(\mathbf{X}, \beta) + \epsilon \tag{2.24}$$

Here ϵ represents errors after fitting a regression model. It is the difference between the observed value of the variable and the value predicted by the regression model. The regression aims to optimize β in order to have an minimum ϵ for a given norm space.

Once a regression model has been constructed, the goodness of the model fit is usually evaluated by R^2, the root mean square error (RMSE) or correlation coefficient (CC). The statistical significance of the overall fit can be assessed via the F-test or t-test [Tabachnick et al., 2001].

The root mean square error (RMSE) is a frequently used measure to compute the overall

regression or prediction error. It can be computed as follows:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{2.25}$$

Here $\hat{y}_i$ is the estimated response by the regression model and N is the number of subjects.

R^2 indicates how well the data points fit the model overall, and is defined as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \mu(y))^2} \tag{2.26}$$

Here $\mu(y)$ denotes the mean value of the response.

The correlation coefficient (CC) is also calculated to evaluate the correlation strength between the predicted and actual variables. It is given by:

$$CC = \frac{\sum_{i=1}^{N}(y_i) - \mu(y)) \cdot (\hat{y}_i - \mu(\hat{y}_i))}{\sqrt{\sum_{i=1}^{N}(y_i - \mu(y))^2} \cdot \sqrt{\sum_{i=1}^{N}(\hat{y}_i - \mu(\hat{y}_i))^2}} \tag{2.27}$$

A $p - value$ is usually calculated to indicate the statistical significance of the CC.

In addition, the median error is also used as accuracy measure. It is defined as the median absolute difference between the predicted and the real value. Compared to CC, the median error eliminates the influence of the standard deviation and thus is well suited for datasets with large standard deviation.

The three classification methods presented earlier also can be used for regression rather than classification. In classification, Random Forests construct decision trees and use the majority votes from the trees for the the prediction; while in the regression problem, the mean of prediction of the individual trees can be used for prediction of continuous variables. And Adaboost uses the sum of the weighted output of weak learners to solve regression problem.

SVM regression, similar to SVM classification, first projects the input $\mathbf{X}$ into a high dimensional feature space, and then constructs a linear regression model similar to Equation 2.15 as follows:

$$f(x) = W^T \phi(X) + b \tag{2.28}$$

Here ϵ intensive loss function proposed by Vapnik [Chapelle and Vapnik, 1999] is used to measure the quality of estimation:

$$L_\epsilon(y, f(x)) = \begin{cases} 0 & , \text{if } |y - f(x)| \leq \epsilon \\ |y - f(x)| - \epsilon & , otherwise \end{cases} \tag{2.29}$$

This loss function defines a ϵ intensive zone (shown in Figure 2.17) where the points inside it are ignored. The loss function is used to construct a linear regression in the high dimensional feature space and aims to reduce model complexity by minimizing ϵ. This can be achieved by introducing non-negative slack variables $\xi_i, \xi_i^*, i = 1..N$, to measure the deviation of training samples outside ϵ-insensitive zone. Figure 2.17 illustrates a simple case of one-dimensional linear regression with ϵ intensive zone. Therefore the task of the SVM regression is to minimize the following function equation:

$$\begin{aligned} & min \ \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}(\xi_i + \xi_i^*) \\ & s.t. \begin{cases} y_i - f(x_i) \leq \epsilon + \xi_i^* \\ f(x_i) - y_i \leq \epsilon + \xi_i \\ \xi_i, \xi_i^* \geq 0 \ , \ i = 1, ..., N \end{cases} \end{aligned} \tag{2.30}$$

The solution for this optimization problem is given by:

$$f(x) = \sum_{i,j=1}^{N_{sv}} (\alpha_i - \alpha_i^*) K(x_i, x), \qquad s.t. \ 0 \leq \alpha_i \leq C, \ 0 \leq \alpha_i^* \leq C \tag{2.31}$$

where N_{sv} is the number of Support Vectors and $K(x_i, x)$ is the kernel function. The cost function ignores the points inside the ϵ intensive zone and hence reduces the computational cost. This regression algorithm is also referred to as SVM ϵ regression.

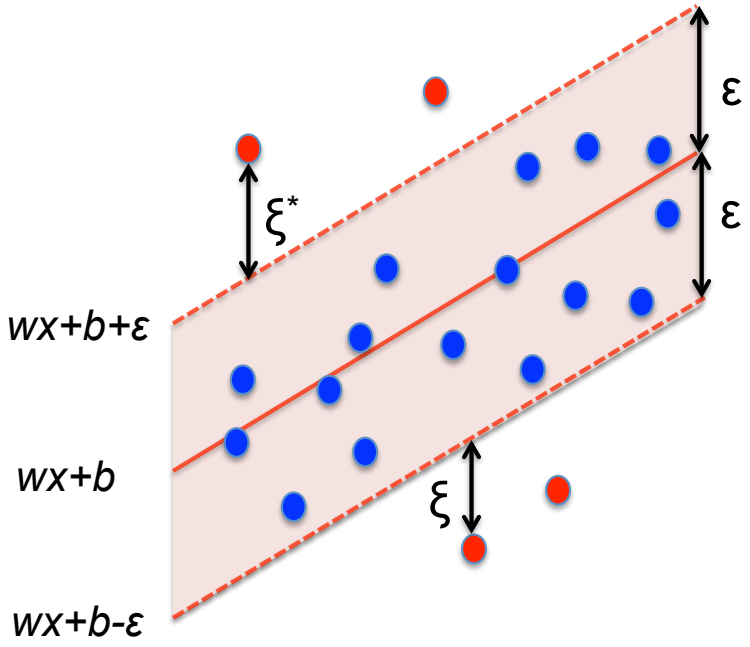


Figure 2.17 – One-dimensional SVM linear regression with ϵ intensive zone. Slack variables ξ, ξ^* measure the deviation of training samples outside ϵ-insensitive zone.

2.3.5 Manifold learning

Manifold learning is normally applied for dimensionality reduction. It preserves the underlying local structure to yield a lower dimensional embedding of the data in a high dimensional space. For a dataset $\mathbf{X} = \{x_1, x_2, ..., x_n\} \in \mathbb{R}^M$ in the original high-dimensional space, we assume there exists a lower dimension manifold which can well represent the $\mathbf{X}$. The manifold learning seeks to find a low dimensional representation $\mathbf{Y} = \{y_1, y_2, ..., y_n\} \in \mathbb{R}^m$ where $m \ll M$ [van der Maaten et al., 2009].

Principle Component Analysis (PCA) and Multidimensional Scaling (MDS) compute a global projection using a linear transformation of the data to a low dimensional space. We include PCA and MDS in this section as linear manifold algorithms in the context of dimensionality reduction. The non-linear manifold learning techniques can be divided into local and global methods: Local linear embedding (LLE) [Roweis and Saul, 2000] and Laplacian Eigenmaps embedding [Belkin and Niyogi, 2003] aim to preserve local distances between data points. Global methods like Isomap [Tenenbaum et al., 2000] consider the distance between all pairs of data points in its objective function, and hence preserve global distances.

PCA

Principle Component Analysis (PCA) is a linear dimensionality reduction algorithm, which converts a dataset of possibly correlated variables into a set of linearly uncorrelated variables, i.e. principal components [Pearson, 1901] [Jolliffe, 2002].

For a given mean centered dataset $\mathbf{X}$, PCA finds the vectors along which the dataset has maximum variance. This is achieved by solving the following eigen-problem for the eigenvalue λ and corresponding eigenvector $\mathbf{v}$:

$$\lambda \mathbf{V} = \mathbf{C}\mathbf{V} \tag{2.32}$$

where the covariance matrix $\mathbf{C}$ of $\mathbf{X}$ of size $N \times M$ is defined as follows:

$$\mathbf{C} = \frac{1}{N} \sum_{i=1}^{N} x_i x_x^T \tag{2.33}$$

The m eigenvectors associated with the largest $m (m \ll M)$ eigenvalues form a transform matrix $\mathbf{U}$, which projects the dataset $\mathbf{X}$ into a low dimensional space as: $\mathbf{Y} = \mathbf{U^T X}$.

PCA is the most popular linear dimensionality reduction algorithm. However the main drawback is the assumption that the data lies in a linear subspace. Kernel PCA is hence introduced to compute the covariance matrix after performing a kernel-based transformation [Mika et al., 1998].

MDS

Multidimensional Scaling (MDS) [Kruskal, 1964] [Cox and Cox, 2000] projects data points from high dimensional space to low dimensional embedding, while preserving their pairwise similarity or dissimilarity. Such similarity or dissimilarity can be measured by the Euclidian distance. Mathematically, MDS seeks to find a linear transformation by minimizing the objective function as follows:

$$\sum_{i=1}^{N} \sum_{j=1}^{N} (\|x_i - x_j\|^2 - \|y_i - y_j\|^2)^2 \tag{2.34}$$

Here x_i is the data point in the original high-dimentional space and y_i is the corresponding projected point in the low-dimentional space. Solving this optimisation problem means that if two data points x_i and x_j that are close to each other in the high dimensional space, then the transformed two data points y_i and y_j are also close to each other in the low dimensional space.

Isomap

Isomap [Tenenbaum et al., 2000] is a nonlinear generalization of classical MDS. It calculates pairwise geodesic distances instead of Euclidean distances. The geodesic distances are defined as the shortest paths along the curved surface of the manifold. The shortest path can be calculated by Floyd's algorithm [Floyd, 1962] or Dijkstra's algorithm [Dijkstra, 1959]. More over, with the assumption of a sufficiently smooth manifold, the geodesic distance between nearby points approximates the Euclidean distance. The Isomap embedding is computed in three steps: First, the neighbours of each data point (the k-nearest neighbours or within ϵ-radius) are identified in the high dimensional space. Then, the geodesic pairwise distances between all points are computed. The final embedding is obtained by applying MDS so as to preserve pairwise geodesic distances in the low dimensional space.

Isomap is computed considering the distances between all pairs of data points, hence preserves global distances and is regarded as a global manifold learning algorithms.

LLE

Locally Linear Embedding (LLE), first introduced by Roweis and Saul [Roweis and Saul, 2000], is a nonlinear dimensionality reduction approach which preserves the local neighbourhood in the high dimensional space. Hence it is regarded as a local manifold learning algorithm.

Similar to Isomap, there are three steps in the computation of LLE: First, the neighbours of each data point (the k-nearest neighbours or within ϵ-radius) are identified in the high dimensional space. Then, all data points are represented as a weighted combination of their neighbours.

The weights ω_{ij} are obtained by minimizing the following cost function:

$$\sum_{i=1}^{N}(\|x_i - \sum_{j=1}^{k}\omega_{ij}x_{N_i(j)}\|^2 \tag{2.35}$$

where $N_i(j)$ indicates the j^{th} neighbour of the i^{th} point.

Finally the projected points in the low dimensional space can be obtained by minimizing the following cost function with the ω_{ij} obtained in Equation 2.35:

$$\sum_{i=1}^{N}(\|y_i - \sum_{j=1}^{k}\omega_{ij}y_{N_i(j)}\|^2 \tag{2.36}$$

This minimization problem (Equation 2.36) is about solving a sparse $N \times N$ eigenvalue problem, whose bottom nonzero eigenvectors provide an orthogonal set of coordinates to construct $\mathbf{Y}$.

Laplacian Eigenmaps

Laplacian Eigenmaps [Belkin and Niyogi, 2003, Von Luxburg, 2007] finds a low dimensional representation which preserves the local properties of the data by ensuring the local neighbourhood in the high dimensional space is reflected in the low dimensional space. In manifold learning, it is common to use a similarity matrix to represent the relations between pairs of data points. The similarity matrix may also be viewed as a graph in which each vertex denotes a data point and the weight of each edge corresponds to the similarity or dissimilarity between the data pair it.

Laplacian eigenmaps build a sparsely connected graph from a pairwise similarity matrix $\mathbf{W}$ computed from the data set. The strength of the connection between each pair of data points in the graph is defined by the similarity matrix $\mathbf{W}$:

$$w_{i,j} = \begin{cases} s_{ij}, & \text{if } \mathbf{x}_i \in \psi(\mathbf{x}_j) \\ 0, & \text{otherwise} \end{cases} \tag{2.37}$$

Here $\psi(\mathbf{x}_j)$ denotes a local neighbourhood of $\mathbf{x}_j$ which can be defined by k nearest neighbour. The manifold is obtained by minimizing the objective function:

$$\begin{aligned}
\Phi(Y) &= \sum_{i,j=1}^{n} w_{ij}(\mathbf{y}_i - \mathbf{y}_j)^2 \\
&= \sum_{i=1}^{n} d_i \mathbf{y}_i^2 - 2\sum_{i,j=1}^{n} \mathbf{y}_i \mathbf{y}_j w_{i,j} + \sum_{j=1}^{n} d_j \mathbf{y}_j^2 \\
&= 2(\sum_{i=1}^{n} d_i \mathbf{y}_i^2 - \sum_{i,j=1}^{n} \mathbf{y}_i \mathbf{y}_j w_{i,j}) \\
&= 2(\mathbf{Y}^T\mathbf{D}\mathbf{Y} - \mathbf{Y}^T\mathbf{W}\mathbf{Y}) \\
&= 2\mathbf{Y}^T(\mathbf{D} - \mathbf{W})\mathbf{Y} \\
&= 2\mathbf{Y}^T\mathbf{L}\mathbf{Y}
\end{aligned} \tag{2.38}$$

where $\mathbf{y}_i, \mathbf{y}_j \in \mathbf{Y}$ is the low dimensional embedding of the input data $\mathbf{X}$; $d_i = \sum_{j=1}^{n} w_{ij}$ and $d_j = \sum_{i=1}^{n} w_{ij}$. The objective function is optimised under the constraint that $\mathbf{y}^T\mathbf{D}y = 1$, which removes arbitrary scaling factors in the embedding and prevents trivial solutions where all $\mathbf{y}_i$ are zero. The y_i that optimise the objective function are derived by solving the generalised eigenvalue problem for eigenvalue λ and its corresponding eigenvectors $\mathbf{V}$:

$$\mathbf{L}\mathbf{V} = \lambda\mathbf{D}\mathbf{V}$$

The embedding coordinates $\mathbf{Y}$ are formed by the eigenvectors corresponding to the smallest nonzero eigenvalues. Table 2.3 summarises the most important properties for the manifold learning techniques discussed in this section [Wittman, 2005].

2.4 Conclusion

This chapter has provided details of some of the most widely used object detection algorithms including feature detectors and feature descriptors, with qualitative and quantitate evaluation of their respective strengths and weaknesses. They contain valuable insights and ideas that

Table 2.3 – Summary of Manifold Learning [Wittman, 2005].

	PCA	MDS	ISOMAP	LLE	Laplacian
Global/Local	G	G	G	L	L
Linear/Non-Linear	L	L	NL	NL	NL
Speed	Extremely fast	very slow	extremely slow	fast	fast
Infers geometry	×	×	✓	✓	✓
Handle non-Convex	×	×	×	maybe	maybe
Handle non-Uniform sampling	✓	✓	✓	✓	×
Handle curvature	×	×	✓	maybe	✓
Handle corner	×	×	✓	✓	✓
Clusters	✓	✓	✓	✓	×
Handle noise	✓	✓	maybe	×	✓
Handle sparsity	✓	✓	✓	✓	✓

can be applied in medical image analysis, and can also inspire future research on object detection in medical images. Machine learning algorithms applied as part of the work, including SVM, Adaboost, regression and manifold learning are presented in detail. Machine learning approaches may be applied to images at different levels: from classifying point as part of the tissue or anatomic landmark, to classifying subjects according to their extracted feature representation. Manifold learning provides a powerful tool for finding an underlying representation of images that facilitates the classification or regression. Next chapter presents more focussed and detailed reviews of the most recent research in the field of motion analysis for cardiac MR images. The following three chapters present the main contributions of the thesis.

Chapter 3

Cardiac motion analysis

Heart motion is caused by the contraction of the four chambers and respiration. For the left ventricle, its motion may be divided into three components: longitudinal shortening, radial contraction, and circumferential twisting of the apex relative to the basal part. Motion of the left ventricle has been extensively studied for quantitative functional analysis. Cardiac image motion analysis is a challenging task because of the complex and nonrigid motion of the heart; in addition, the heart has a more ambiguous appearance with few anatomical landmarks. Wang and Amini [Wang and Amini, 2012] provide a recent review on cardiac motion analysis.

In this chapter, we focus on a review of motion tracking methods developed for cine MR imaging. The main motion analysis algorithms for cine MR images can be categorised into two groups: sparse motion tracking and dense motion tracking. In Section 3.1 we first introduce two types of indices for assessing cardiac function which are related to cardiac motion analysis. In addition, some relevant obesity indices are also included. Section 3.2 focuses on registration-based dense motion tracking and describes different measures that evaluate the similarity of the images for use of motion tracking algorithms. In Section 3.3 deformable modelling approaches are reviewed. Section 3.4 explores methods for sparse motion tracking. Finally Section 3.5 explores applications of manifold learning in the field of cardiac motion tracking.

3.1 Cardiac functional indices

Assessment of regional and global left ventricular function provides clinicians with information for the diagnosis of CVD and the prediction treatment outcomes. Global indices provide information that characterize the overall performance of the heart, whereas regional indices quantify myocardial deformation and hence are indicators for dysfunctional contraction. Among the imaging modalities available routinely in clinics, magnetic resonance imaging (MRI) offers accurate imaging of anatomy, morphology, myocardial motion and blood flow with good contrast, which enables quantitative assessment of heart function. MR imaging provides images of high spatial resolution with a noninvasive and non-ionising approach and hence is the suitable tool to assess cardiac function [Frangi et al., 2001]. Recently excess body fat has been associated with a series of heart diseases [Lauer et al., 1992] [Turkbey et al., 2010] [Aljaroudi et al., 2012], therefore we include some relevant obesity indices in this section, which will be further investigated in this work.

3.1.1 Global indices

Global indices of cardiac function include the left ventricular volume (LVV), left ventricular mass (LVM), stroke volume (SV), cardiac output (CO), cardiac index (CI) and ejection fraction (EF), which can be used to assess the performance of the left ventricle [Frangi et al., 2001].

LVV measures the volume of the blood in the left ventricle. The curve of the LVV during a cardiac cycle is usually recorded to assess the performance of the ventricle. The LVV at the end of the diastole and the systole are usually referred to as the end-diastolic volume (EDV) and the end-systolic volume (ESV). The EDV ranges from $68mL$ to 239mL, ESV has the range of $16 - 143$mL [Schlosser et al., 2005].

LVM is the product of the volume of the myocardium V_m and an estimate of myocardial density, $\rho_m = 1.05g/cm^3$ [Foppa et al., 2005]. Increase in LVM defines left ventricular hypertrophy. LVM is defined as follows:

$$LVM = V_m \rho_m$$

The SV is the volume of blood ejected from the ventricle within each heart beat and can be measured as follows:

$$SV = EDV - ESV$$

The cardiac output (CO) is the volume of the blood pumped by the ventricle per minute, which can be calculated as the SV multiplied by the heart rate.

The cardiac index (CI) is a useful marker of the performance of the heart as a pump by relating the CO to the body surface area (BSA) of the individual:

$$CI = \frac{CO}{BSA}$$

The ejection fraction (EF) refers to the percentage of blood that is ejected out of the ventricles with each contraction:

$$EF = \frac{SV}{EDV} \times 100\%$$

The EF is generally considered as one of most important indicators for the left ventricle pump ability. The EF can be calculated for both the right ventricle and the left ventricle. The normal range of the value of EF is between $55 - 70\%$ [Gatenby, 2006].

3.1.2 Regional indices

Global indices of cardiac function may not correlate with myocardial contractility, which is essential for diagnosis of ischemic heart disease. Wall motion abnormalities and regional contractile reserve are indicative of some heart diseases. Therefore several regional indices, such as regional wall thickening (RWT), regional wall motion (RWM) and myocardial strain, have been proposed to facilitate regional functional analysis [Blumenthal et al., 1983] [Schiller et al., 1988] [Cigarroa et al., 1993].

In order to conduct the regional analysis of cardiac motion, the muscle and cavity of the LV is typically divided into segments using a 17 segment model proposed by the American Society

of Echocardiography [Cerqueira et al., 2002], as shown in Figure 3.1. In this model, the LV is divided into six basal, six mid-ventricular, and five apical regions for the regional analysis of left ventricular function or myocardial perfusion. These 17 segments can also be arranged as a polar plot with the apex in the center to allow for the easy comparison of the outcome in different conditions (eg. rest/stress) or between patients.

Regional wall thickening (RWT) represents the local change in myocardial wall thickness from ED to ES, which can be assessed by [Schiller et al., 1988]:

$$RWT = (T_{ES} - T_{ED})/T_{ED} \times 100\%$$

Here T_{ES} and T_{ED} are the wall thickness measurements obtained at ES and ED respectively. The wall thickness is the distance between the endocardial and epicardial surfaces.

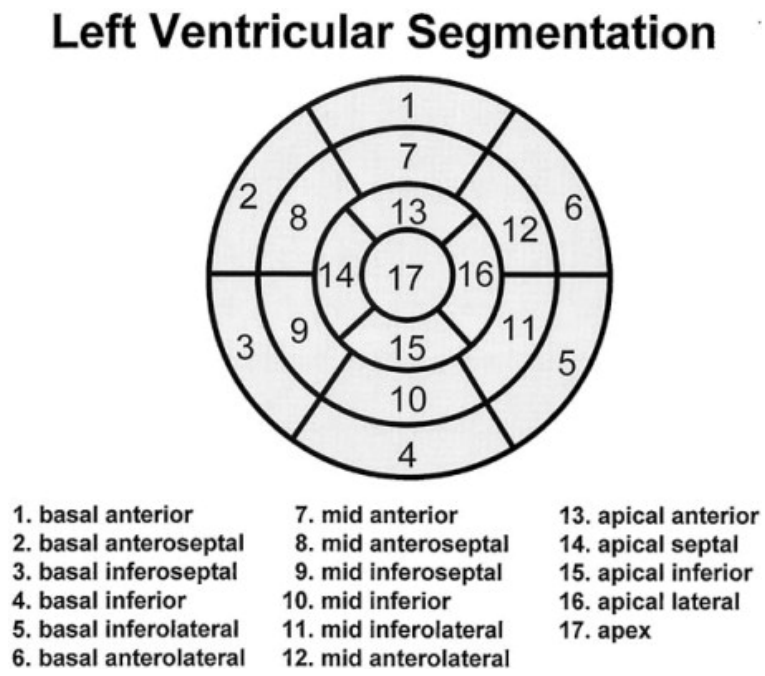
Regional wall motion (RWM) describes wall motion of a segment of the myocardium which is perpendicular to the endocardial wall. It can be calculated as the fractional change of the average radial distance from a segment to the central axis of the blood cavity,

$$RWM = (D_{ES} - D_{ED})/D_{ED} \times 100\%$$

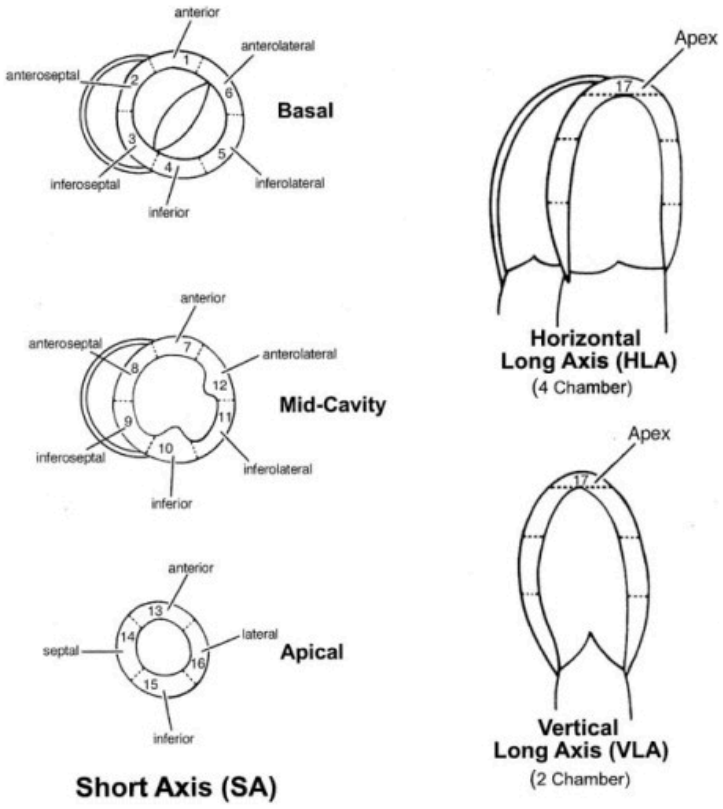
The regional wall motion score can be reduced, absent or increased, which may provide hints regarding coronary artery disease.

Azhari et al. [Azhari et al., 1990] compared two indices for distinguishing ischemic from non-ischemic myocardium. In their experiment both RWM and RWT demonstrated the ability to differentiate between the center of the ischemic zone and the remote normal zone. However, statistical analysis indicated that the RWT algorithm provided better discrimination.

Myocardial strain [Mirsky and Parmley, 1973] measures local tissue deformation to assess myocardial contractile function. It is the percentage change of the length from a rest state to a stressed state. The change of length is considered in all three directions of radial, circumferential and longitudinal and also characterized by the direction of the change.



(a)



(b)

Figure 3.1 – Diagram of the 17 myocardial segments on a circumferential polar plot on (a); and on (b) a selection of SA and LA planes [Cerqueira et al., 2002].

In clinical practice, a clinician may want to assess both global and regional indices for diagnosis. For example, patients lacking contractile reserve have lower ejection fractions and higher regional wall thickening (RWT) scores, and were more likely to have apical dyskinesis [Cigarroa et al., 1993].

3.1.3 Obesity indices

Body Mass Index (BMI)

BMI [Eknoyan, 2008] evaluates the body mass relative to the height, which provides a simple measure to assess whether a person is underweight, ideal, overweight or obese. It is defined as follows:

$$BMI = \frac{\text{weight}}{\text{height}^2}(kg/m^2) \tag{3.1}$$

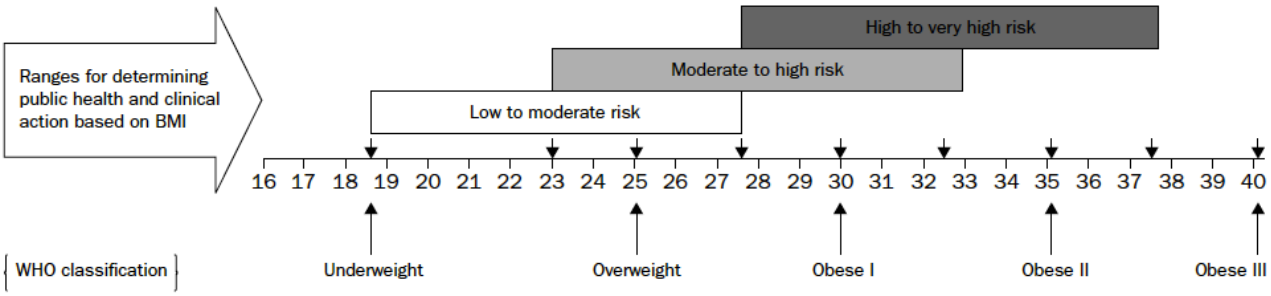


Figure 3.2 – BMI cut-off points for diabetes and heart disease by WHO. Figure from [WHO Expert Consultation, 2004].

BMI has been used extensively in research and clinical practice, in addition, it has been used by the WHO for obesity statistics and as a risk indicator for heart disease [WHO Expert Consultation, 2004], as shown in Figure 3.2. Aljaroudi et al. [Aljaroudi et al., 2012] reviewed 21666 clinical records and echocardiograms of patients with normal LVEF and concluded that BMI is independently associated with worsening diastolic dysfunction (DD).

Body fat percentage (Fat%)

Fat% is the body fat percentage of the body mass, which measures the fitness level. It varies between age and gender, as shown in Figure 3.3. In general, women have a higher fat percentage than men.

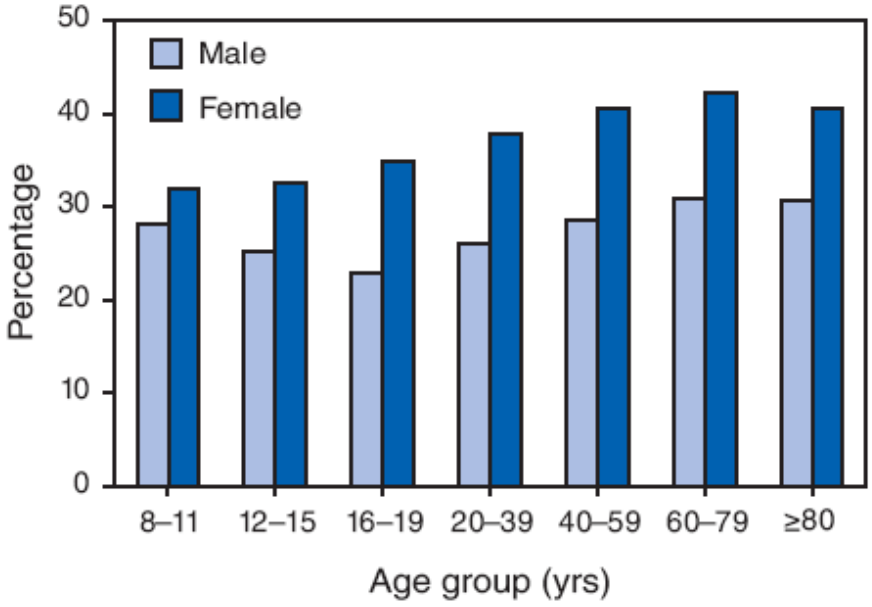


Figure 3.3 – Mean percentage body fat (Fat%) by age group and gender. Figure from [MMWR Weekly: January 2, 2009].

Due to the difference in the body composition, such as muscle mass or bone size, Fat% is considered as a more accurate measure for obesity than BMI. WHO proposed to use $> 25\%$ in men and $> 35\%$ in women to define obesity [WHO, 1995] [Romero-Corral et al., 2008]

Fat Free Mass Index (FFMI)

Fat free mass (FFM) represents body composition where FFMI expresses the fat free component of body mass relative to the body size. It is defined as:

$$FFMI = \frac{\text{fat free mass}}{\text{height}^2}(kg/m^2) \tag{3.2}$$

FFMI eliminates the differences of the Fat% associated with height by normalising the fat free mass for body size. FFMI of 19.8 in men and 16.7 in women are found to correspond to the

cut-off point of BMI of $25.0kg/m^2$ (for risk of heart disease and diabetes) [Schutz et al., 2002]. FFMI has seen increasing use to assess overweight and obesity.

3.2 Dense motion tracking

Dense motion tracking normally evaluates the alignment of the images over the whole image domain via image registration. Cardiac motion tracking can be viewed as a 4D registration problem [Makela et al., 2002]. Though this approach may better quantify the alignment through a dense representation, it comes at the cost of more expensive computations.

Image registration is a process for determining the transformation which maps points between images acquired at different times or using different imaging modalities by maximising a similarity measure. A registration algorithm therefore consists of similarity metric, transformation model and optimization process. A detailed review of cardiac image registration methods given in [Makela et al., 2002].

In image registration, the transformation model plays the most important role. It assumes the nature of the deformation to be recovered and the degree of freedom of a model determines its descriptive power and computational complexity. The transformation model can be coarsely classified into two different types: i) those that focus on interpolation and approximation strategies, and ii) those that are based on physical or deformable models.

3.2.1 Similarity metrics

Given two images, a target image I_1 and a source image I_2, a transformation T defines the correspondence between location $\mathbf{x}$ in I_1 and location $T(\mathbf{x})$ in I_2 across the image domain Ω. For cardiac motion tracking, the image at the ED phase is usually used as the target image and all other time frames can be regarded as source images. The similarity of the two images is then computed by a set of the intensity pairs $\{I_1(\mathbf{x}), I_2(T(\mathbf{x}))\}$.

The sum of squared difference (SSD) provides the simplest form of similarity metric to measure the difference between a target I_1 and a source I_2 images, which is given by:

$$SSD = \sum_{\mathbf{x}\in\Omega} (I_1(\mathbf{x}) - I_2(T(\mathbf{x})))^2$$

This similarity metric assumes that the target and the source images have similar intensity distributions, except for Gaussian noise [Viola and Wells III, 1997]. It is therefore a suitable measure for cardiac motion tracking, since the target and the source images are usually from the same sequence of images and there are few global intensity changes between different frames of the sequence. When this assumption does not hold cross-correlation may be used as the similarity measure assuming a linear relationship between corresponding intensities in the two images [Crum et al., 2014]. The normalised cross-correlation is defined by:

$$CC = \frac{\sum_{\mathbf{x}\in\Omega} (I_1(\mathbf{x}) - \mu(I_1)) \cdot (I_2(T(\mathbf{x})) - \mu(I_2))}{\sqrt{\sum_{\mathbf{x}\in\Omega} (I_1(\mathbf{x}) - \mu(I_1))^2} \cdot \sqrt{\sum_{\mathbf{x}\in\Omega} (I_2(T(\mathbf{x})) - \mu(I_2))^2}}$$

Here $\mu(I)$ is the mean intensity of image I.

In the cases where the linear relationship between the intensities of corresponding locations does not hold, entropy-based similarity metrics which use information from the whole image may be applied to measure the similarities. Shannon entropy is a measure [Shannon, 1949], which quantifies the information content of a probability distribution. For a image I, its Shannon entropy is defined by:

$$H(I) = -\sum_i p(i) \log p(i)$$

Here $p(i)$ is the probability of voxels with intensity i occurring in the whole image I. This probability is usually estimated by a histogram-based approach ([Studholme et al., 1997], [Studholme et al., 1999]) or via Parzen window [Parzen, 1962], which is a non-parametric way to estimate the probability density function of a random variable. The entropy of the image is maximized if all possible intensities are equally likely to occur, i.e. the uncertainty is highest

when all intensities are equiprobable. On the other hand, the entropy is zero (minimal) if the image consists of only one intensity level.

Similarly, the joint entropy ([Studholme et al., 1995]) can be used to estimate how well two images are aligned:

$$H(I_1, I_2) = -\sum_{i_1 \in I_1, i_2 \in I_2} p(i_1, i_2) \log p(i_1, i_2)$$

Here $p(i_1, i_2)$ denotes the joint probability distribution function of the images I_1 and I_2. It can be estimated using Parzen windows or Histograms. A low value of $H(I_1, I_2)$ is usually associated with low information content, indicating a better alignment of two images. However, a large portion of well aligned background between two images may also generate a low value of the joint entropy.

In order to overcome the limitation of the joint entropy, mutual information (MI) ([Collignon et al., 1995], [Viola and Wells III, 1997]) is used as an alternative measure. It includes the individual marginal entropies of the images and measures the information shared by two images, which is defined by:

$$MI = H(I_1) + H(I_2) - H(I_1, I_2) = -\sum_{i_1 \in I_1} \sum_{i_2 \in I_2} p(i_1, i_2) \log \frac{p(i_1, i_2)}{p(i_1) \cdot p(i_2)}$$

It is easy to see that when image I_1 and I_2 are independent, the mutual information is minimal (zero). Maximizing the MI therefore leads to better alignment of the two images.

An alternative measure, normalized mutual information (NMI) was proposed by Studholme et al. ([Studholme et al., 1999]) , which is given by:

$$NMI = \frac{H(I_1) + H(I_2)}{H(I_1, I_2)}$$

It has been shown to be more robust to changes in the overlap of the two images.

3.2.2 Registration-based motion tracking

For cardiac motion analysis, image registration based methods are widely used to extract cardiac motion [Wang and Amini, 2012]. Non-rigid registration using free-form deformations was proposed by [Rueckert et al., 1999]. In this approach, a multilevel B-splines based FFD is used to describe the location motion, while the global motion is modelled by an affine transformation. A hierarchical registration is achieved by optimising a cost function measuring the similarity between two images as well as the smoothness of the deformation required to align the images. The similarity measure used is based on NMI.

Chandrashekara et al. [Chandrashekara et al., 2004b] used a 4D B-spline motion model and nonrigid image registration proposed in [Rueckert et al., 1999] for the analysis of myocardial deformations. In this approach, the cost function represents the normalised sum of the NMI between the registered SA and LA images. In their work, tag localisation and deformation field reconstruction are performed simultaneously within the optimisation. Another advantage is that no assumptions about the nature of the tag pattern are made so that the information of other image modalities can be easily incorporated.

Perperidis *et al.* [Perperidis et al., 2005] extended Rueckert's work [Rueckert et al., 1999] to 4D B-spline based FFD for the inter-subject spatiotemporal alignment of cardiac MR image sequences. The 4D registration is resolved into decoupled temporal registration and spatial registration. The temporal transformation consists of an affine transformation for correcting scaling and translation differences between two sequences and a 1D B-Spline transformation accounting for the temporal misalignment caused by length differences of the cardiac phases. The spatial transformation also consists of an affine transformation and a 3D B-Spline transformation addressing differences in the size, orientation and translation of the hearts and differences in the shape of the hearts respectively. Two different optimisation methods were proposed: one performed both registrations simultaneously; the other performed the temporal registration followed by the spatial registration.

Sundar *et al.* [Sundar et al., 2009] proposed a 4D elastic registration algorithm, which constructs

a 4D template image by repeating the ED image and estimates the cardiac motion from this 4D template to all other time points simultaneously. Compared to other registration-based algorithms which estimate the motion by sequentially registering one frame to another, the proposed 4D framework makes the motion recovery temporally smooth and consistent. An attribute vector, consisting of intensity, boundary, and geometric moment invariants (GMIs), is defined for each voxel to reflect the underlying structure at different scales. The similarity used for registration and the distinctiveness for each voxel are calculated based on the attribute vector. The registration is performed in a multi-resolution fashion by first selecting the most distinctive points and gradually adding less-distinctive points.

Peyrat *et al.* [Peyrat et al., 2010] used a 4D spatiotemporal registration approach based on the Diffeomorphic Demons algorithm for intersubject registration. The 4D registration is decoupled into two components: temporal registration and spatial registration. The temporal registration is performed using global cardiac physiological state parameters: First, it temporally aligns two sequences with a linear transformation based on the ECG and the transformation is then refined by temporally aligning the volume curves. This temporal alignment ensures that the following intersubject spatial registration is performed between frames at corresponding states during the cardiac cycle and thus with similar geometrics. For the spatial registration illustrated in Figure 3.4, Peyrat *et al.* first applied Diffeomorphic Demons to find a dense trajectory of points in each of the two sequences; and then these trajectories are used as constraints to compute the intersubject 3D spatial transformation. So they formulated the 4D spatial registration as a multichannel registration of 3D images and solved it with Multichannel Diffeomorphic Demons.

Metz et al. [Metz et al., 2011] proposed a method based on a 3D (2D+time) or 4D (3D+time) free-form B-spline deformation model to extract cardiac motion. In order to avoid a bias towards a specifically chosen reference time point, the registration is performed directly on the dynamic image. They used a stochastic gradient descent method with adaptive step size estimation for constrained optimization. The similarity metric is computed to minimize the intensity variances over time. Optionally, cyclic motion constraints can be imposed for model building.

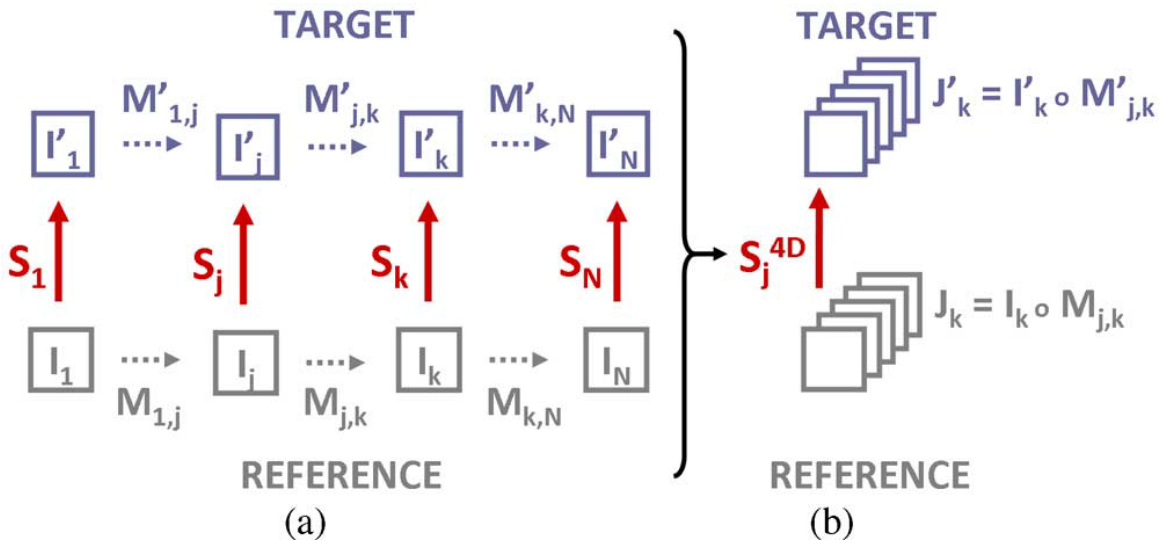


Figure 3.4 – (a) Discretization of the 4D spatial registration with the spatial transformations S_j between the sequences at time t_j, and the trajectories $M_{j,k}$ and $M'_{j,k}$ between frames at times t_j and t_k. The intersequence transformation S_j maps the reference I_j to the target I'_j at time t_j knowing the trajectories of points given by the intrasequence motion transformations $M_{j,j+1}$ and $M'_{j,j+1}$ between the times t_j and t_{j+1}. (b) Under trajectory constraints, the 4D registration can be parametrized by a single spatial transformation S_j^{4D} and thus formulated as a multichannel 3D registration problem. Once S_j^{4D} is estimated, the other transformations S_k^{4D} can be computed from S_j^{4D} with the trajectory constraints: $S_k^{4D} = M'_{j,k} \text{o } S_j^{4D} \text{ o } M_{j,k}^{-1}$ [Peyrat et al., 2010].

Yigitsoy *et al.* [Yigitsoy et al., 2011] also applied FFDs based B-splines [Rueckert et al., 1999] to model the motion, with an additional dimension along the temporal direction. This groupwise registration approach ensures the smoothness and consistency in time. The authors used accumulated pairwise estimates (APE) as a similarity measure, which was introduced as a suitable metric for groupwise framework. The APE computes the sum of the mean squared intensity differences between each of all combination pairs of warped images. To deal with the high-dimensional simultaneous registration, they combined a gradient-based optimization procedure with a stochastic sampling in the spatial domain, while only 10% of the total pixels are randomly selected in each iteration for calculating the image similarity.

The Temporal Diffeomorphic Free Form Deformation (TDFFD) [De Craene et al., 2012] extended the FFD registration technique [Rueckert et al., 1999] by summing spatiotemporal B-spline kernels to model the velocity field and hence enforced time consistency. The image similarity metric is calculated as the sum of squared differences between the intensities of each of all time frames and a reference frame. In addition, the incompressibility of myocardial tissue

is included as regularisation term in the energy function. TDFFD is shown to be more robust to noise and reduced temporal resolution.

3.3 Deformable model

Deformable models have been widely used in medical image analysis. In this approach, models are deformed to fit the image using energy minimization [McInerney and Terzopoulos, 1996]. These models are controlled by internal and external forces simultaneously. The external forces drive the model from its initial shape to match the features of the heart images; while the internal forces usually try to enforce the geometric smoothness. Many different models have been proposed to constrain the deformation. Among them, the elastic body model defines the image that deforms as an elastic body while the viscous fluid flow model is able to recover large deformations. For cardiac motion recovery in cine MR images, the external force is usually applied based on the control points extracted from myocardium or endo- and epicardial contours segmented in the reference frame. The advantages of this type of method are the easy construction of statistical models for measuring EF, RWN and RWM.

Veress *et al.* [Veress et al., 2005] used differences in image intensities between template and target images to generate a body force that deforms a finite element model. A hyper-elastic strain energy along the fiber angles is used as the regularisation force. The model is deformed to determine LV strains from mid-diastole to end-diastole using multiple SA cine images. The boundaries of the LV were manually segmented. They further applied this technique for the analysis of cardiac deformation using PET images and tagged MRI images [Veress et al., 2013]. The predicted LV circumferential and radial strains using two modalities agreed well with each other, which indicated that the methodology is relatively insensitive to alterations in image intensity, random image noise, and alterations in fiber structure.

Bistoquet *et al.* [Bistoquet et al., 2008] proposed to utilize myocardial incompressibility as a hard constraint to recover the motion of bi-ventricle. A common conclusion from earlier research shows that the total myocardial volume changes no more than 4% during a cardiac cycle. The

proposed 3D incompressible transformation model uses a divergence-free matrix-valued radial basis function to interpolate the displacement field, which enforces the incompressibility. The model allows a domain of an arbitrary topology. NMI was used as the similarity measure. The locations of the control nodes are linear weighted average of forward and backward registration.

Yu et al. [Yu et al., 2014] proposed deformable models with sparsity constraints for cardiac motion analysis. A robust deformable model is designed by integrating an L1 norm regularization with a modified Laplacian deformable model, which is able to handle outliers. A second defomable model is designed to handle both the Gaussian errors and outliers using hybrid norm regularization. The models are applied to the analysis of cardiac motion using tagged magnetic resonance imaging (tMRI) and the resulting myocardial strains are consistent with those calculated from manual labels.

Wang et al. [Wang et al., 2015] reconstructed 3D cardiac motion with a novel meshless deformable model. The volumetric deformable models are deformed by computing the dynamics function and minimizing the local Laplacian coordinates. The deformation is computed iteratively until the control points match the target positions in the consecutive image frame. The meshless model has been validated on in-vivo data of healthy subjects and patients. The experimental results show that the meshless deformable model can fully recover the myocardial motion and strain in 3D.

3.4 Sparse motion tracking

Sparse landmark-based cardiac motion tracking provides information on feature point correspondences across the heart. The LV boundary motion is quantified by finding the transformation between the matched pairs of the feature points at different time frames. This process usually starts with segmenting the endocardium and epicardium at ED phase either automatically or manually. Followed by the extraction of feature points on or within the endocardial or epicardial boundaries thereafter. Tracking the motion of feature points is often treated as an optimisation process. The correct correspondence between features at two time

points results in the lowest energy configuration of the system, which is usually achieved by the maximising a similarity measure. After establishing correspondences between feature points at two time points, the motion of the myocardium is then recovered by deforming the complete boundary using these feature point correspondences and a given deformation model. However, locating reliable distinctive landmarks is a challenge. Poorly selected landmarks can easily lead to wrong or ambiguous correspondences.

Lin *et al.* [Lin et al., 2003] proposed generalized robust point matching (GRPM) by including area-based curvature information to extend the robust point matching (RPM) [Rangarajan et al., 1997] algorithm. The GRPM is able to handle more general point-matching problems and to guide more precise motion recovery. The feature points were extracted based on thresholded area-based Gaussian curvature operator and tracked by their proposed method, which adds this curvature information to its energy function and estimates a nonrigid transformation and correspondences using soft-assign and deterministic annealing techniques.

Based on their GRPM algorithm, Lin and Duncan [Lin and Duncan, 2004] presented an Extended Free-Form Deformation (EFFD) model for tracking the motion of LV. Unlike Free-Form Deformations (FFDs) that employ paralelepipedical lattice, the algorithm first roughly segments the myocardial border in the ED phase to build an arbitrarily-shaped myocardial lattice, which is more appropriate for cardiace motion tracking. This method is claimed to be more accurate and efficient by reducing the number of control points, but it still suffers from a tradeoff between the shape smoothness and approximation accuracy depending on the size of the unit cell of the lattice [Lee et al., 1996] [Yan et al., 2007].

Yan *et al.* [Yan et al., 2007] proposed to use the boundary element (BEM)method to replace FFDs or EFFDs for regularization of the deformation model within the GRPM framework. The proposed BEM model maps the displacements associated with feature points to the boundary grid nodes and then interpolates the dense displacements on the image voxels. Because the motion of each feature point has an effect on all the boundary nodes, the size of the unit cell of the lattice does not affect the accuracy of the tracking, as shown in Figure 3.5. This model is more efficient in terms of computational complexity as it only requires discretisation of the

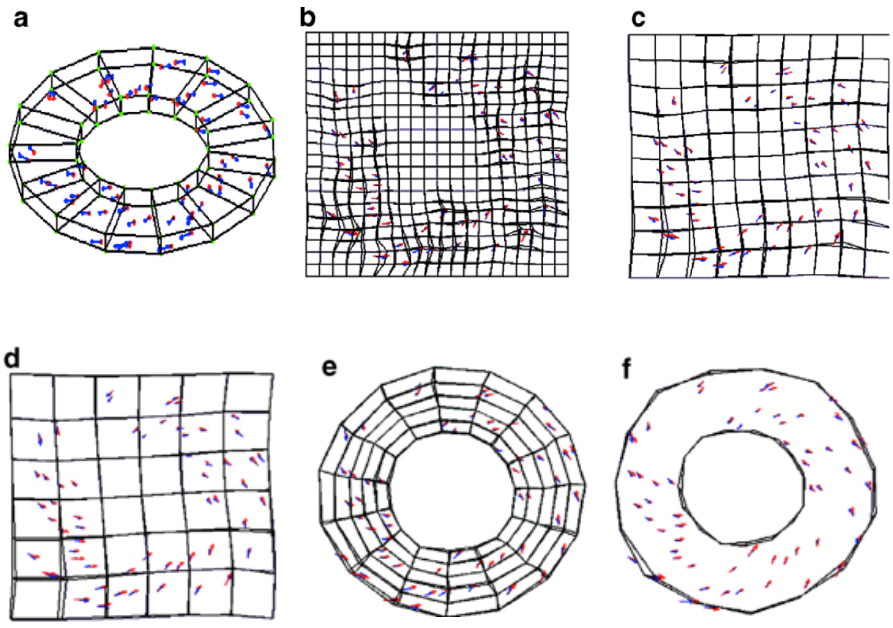


Figure 3.5 – Diagram of comparisons of BEM-based model to FFD and EFFD are evaluated on a synthetic 3D circle in (a). The scattered distributed red points are randomly chosen. The green points are the boundary nodes. FFD with different size of unit cell of the lattice of (b)one, (c)two and (d)three pixels. EFFD in (e) and BEM-based model in (f). Blue and red arrows represent the true and recovered displacements respectively. It can be seen that the red and blue arrows overlapped better in (f) than in other models. Figure from [Yan et al., 2007].

surface rather than the volume.

Remme *et al.* [Remme et al., 2005] presented a modality-independent approach for estimating LV wall motion based on manual tracking of distinct anatomical landmarks. A FE mesh was fitted to the manually segmented endo- and epicardial surfaces at the ED phase. The motion of the LV wall was then reconstructed by deforming the FE mesh to match the displacements of the sparse set of the landmarks. In order to avoid the under-constrained solution due to the sparse nature of the landmarks, a parameter distribution model (PDM) produced from tagged MR imaging was incorporated to regularize the calculated deformed geometry. While the deformation of the B-spline based FFD grid is defined in terms of control points by B-spline basis functions, the deformation induced by an FE mesh is modeled by the deformation of a physically-defined elastic object. The shape of the elements and the mesh geometry of can be adapted to the underlying image. The FE method can be considered an approximation to the physical reality, but is sensitive to the choice of parameters. Furthermore it is computationally very expensive.

3.5 Manifold learning for motion analysis

Manifold learning is a branch of machine learning techniques that learns a low-dimensional embedding of the studied data by taking into account the local information of the input data during dimensionality reduction. Manifold learning can be used to model statistical distribution of anatomical and functional features for a given population, hence clinical applications including biomarker detection, diagnosis and therapy.

Bhatia *et al.* [Bhatia et al., 2013] [Bhatia et al., 2014] extended the conventional manifold learning method by utilising local variations to produce spatially-varying manifold embeddings that characterise a given dataset. In their study, a hierarchical model is adopted to repeatedly subdivide the image into smaller patches in order to enforce spatial consistency and allow regional analysis. They applied this hierarchical manifold learning approach to cardiac MRI to learn the different motions occurring in the cardiac and respiratory cycle to determine the phase differences between different heart regions.

Duchateau *et al.* [Duchateau et al., 2012] compared the myocardial motions of individuals to a population with abnormal motion along a manifold structure. In the first step, the myocardial velocities for each individual are extracted at each time and space point using image-based registration. Then the extracted velocities of each individual are synchronized to a common reference anatomy and compared to that of an atlas of healthy volunteers to generate a 2D p-value maps of abnormalities. In the second step, a relevant manifold for a set of patients is estimated by the isomap algorithm [Tenenbaum et al., 2000]. In the last step, new subjects of patients and healthy volunteers are then mapped to the manifold by interpolation. The performance of this approach is evaluated in the context of cardiac resynchronization therapy (CRT) candidates. Their experiments demonstrated the advantage of non-linear embedding and the relevance of this technique to model a pathological pattern.

Yang *et al.* [Yang et al., 2011] presented a robust and fast prediction based collaborative 3D tracking algorithm. A prediction is introduced to generate the motion prior using motion manifold learning. The motion manifold is embedded by Isomap using 4D motion vectors from

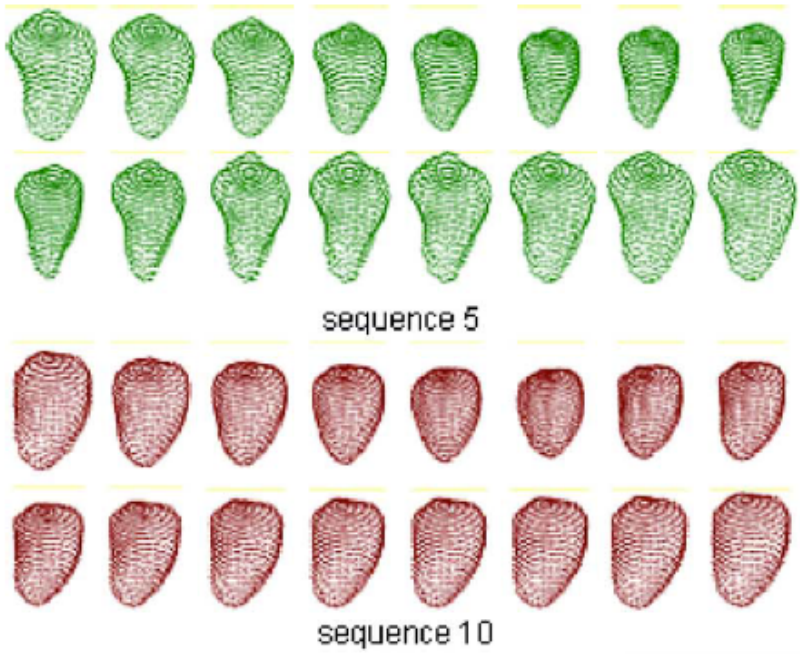
36 annotated LV motion sequences, of which each sequence has 289 annotated boundary points. Figure 3.6 shows two annotated LV motion sequences and 11 LV motion representations in a low dimensional manifold. Thereafter a hierarchical k-means clustering method is used to learn the motion modes. Collaborative trackers including a detection tracker and a template tracker are then applied to achieve both temporal consistency and failure recovery. The approach is tested on three large datasets (CT or echocardiography) for endocardium, myocardium and whole heart four chambers tracking respectively and achieved very accurate tracking results.

Taimouri *et al.* [Taimouri and Hua, 2013] presented a novel classification and visualization method based on a medial surface shape space, which considered the cardiac motion as a function defined on the 2D manifold of the surface. In their medial surface shape space, the geodesic distance connecting two points in the space measures the similarity between their corresponding medial surfaces. The experimental results on both synthetic and real imaging data shown that this method can discriminate the myopathic subjects from the healthy control subjects, and is able to detect myopathic regions.

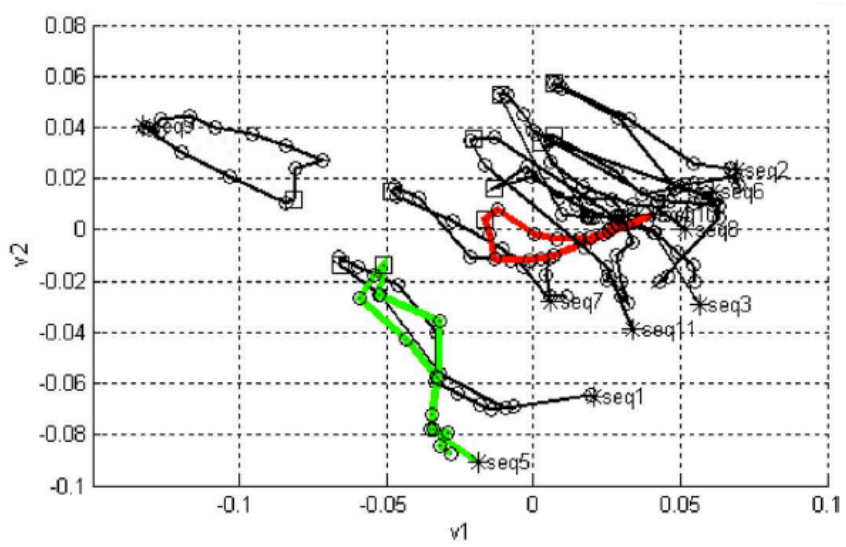
3.6 Conclusion

Cardiac MR images provide accurate imaging of anatomy, morphology, myocardial motion and blood flow with good contrast which are useful for performing global and regional function analysis. In this chapter, we have reviewed recent studies related to the cardiac motion analysis for MR images. Several cardiac motion analysis techniques such as sparse motion tracking, deformable models and registration based algorithms can be applied to both cine MR images and tagged MR images. We also discussed the applications of manifold learning in the field of cardiac motion analysis. Manifold learning is a promising approach in non-parametric dimension reduction and therefore facilitates application of machine learning for medical image analysis. Clinic indices associated with ventricular function assessment and obesity indices are also introduced.

The following three chapters present the main contributions of this thesis. The research of



(a)



(b)

Figure 3.6 – Manifold embedding for heart motion patterns. (a) Two LV surface mesh sequences. (b) 11 sequences embedded in a 2D manifold subspace. The ED phase is represented as stars and the ES phase is represented as squares, and each circle denotes a time phase of the cardiac cycle in addition to ED and ES. Two LV motion tracks are annotated in red and green. Figures from [Yang et al., 2011].

this thesis focuses on the utilization of cardiac motion information involving motion tracking, quantification of the cardiac function and prediction of clinical variables.

Chapter 4

Collaborative landmark detection and motion tracking in cardiac MR images

The work in this chapter is based on the following published papers:

- Haiyan Wang, Wenzhe Shi, Xiahai Zhuang, Simon Duckett, Kai Pin Tung, Philip Edwards, Reza Razavi, Sebastien Ourselin, and Daniel Rueckert. (2011). Automatic cardiac motion tracking using both untagged and 3D tagged MR images. Statistical Atlases and Computational Models of the Heart: Imaging and Modelling Challenges (STACOM), pages 45-56.

- Wenzhe Shi, Xiahai Zhuang, Haiyan Wang, Simon Duckett, Duy V.N. Luong, Catalina Tobon-Gomez, KaiPin Tung, Philip Edwards, Kawal Rhode, Reza Razavi, Sebastien Ourselin, Daniel Rueckert. (2012). A comprehensive cardiac motion estimation framework using both untagged and 3D tagged MR images based on non-rigid registration. IEEE Transactions on Medical Imaging, 31(6):1263-1275.

4.1 Introduction

Cardiac MR imaging provides important information for the quantitative assessment of heart function, such as the measurement of ventricular volume, wall thickness and ejection fraction. The accurate estimation of cardiac motion is becoming increasingly important in quantification of the viability and contractility of the myocardium. However, it requires accurate registration and motion tracking which is too tedious and time-consuming to perform manually.

In this chapter, we propose an automatic approach to identify and track important cardiac landmarks throughout the cardiac cycle. In particular we use a machine learning approach to detect landmarks such as the valve plane simultaneously in three different LA views and employed a collaborative similarity measure simultaneously computed in three different long-axis views to register a sequence of MR images acquired during the cardiac cycle to a reference image acquired at end-diastole. Another contribution is the combination of complementary information from tagged and untagged MR images using a spatially adaptive weighting and a valve plane constraint to construct an accurate and robust cardiac motion analysis framework. The tracked sparse anatomical landmarks in the heart help accurately modelling the cardiac motion while significantly reducing the computational complexity. The overview of the methods is illustrated in Figure 4.1.

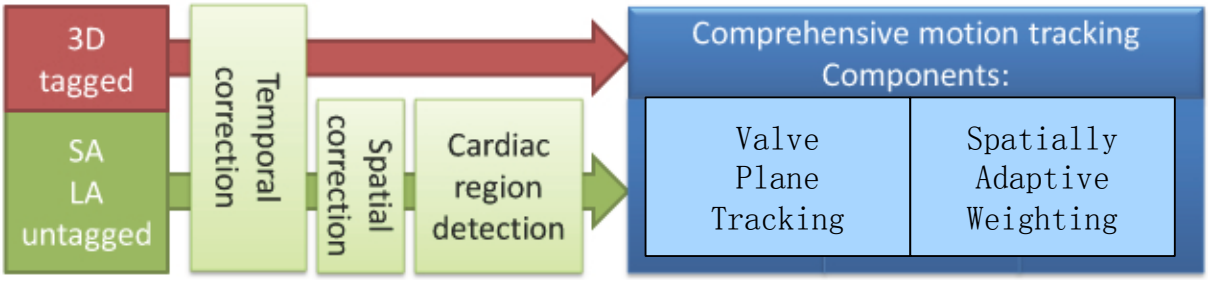


Figure 4.1 – This framework shows the overview of the methods used in this chapter.

4.2 Method and dataset

4.2.1 Material

The data used in this chapter were obtained using a 1.5T MR scanner (Achieva, Philips Healthcare) with a 32-element cardiac coil or a 5-element cardiac coil. Vector electrocardiography (VECG) is used to synchronize the data acquisition with the electrical activation of the heart. Standard views including the SA, 4CH, 2CH and 3CH views, are constructed using the following parameters: flip angle = 60^o, TR/TE = 2.9/1.5ms, resolution 1.45 x 1.45 x 10mm, 30 heart phases. Among them, single slice cine sequences are obtained for all LA views and multiple slice steady-state free precession (SSFP) sequences are acquired for the SA view.

3D tagging was implemented using three sequentially acquired 3D data sets with line tag preparation in each of the three spatial dimensions [Rutz et al., 2008]. 3D tagged images were acquired of the whole left ventricle (tag separation = 7mm, FOV = 108 x 108 x 108mm, EPI factor = 7, TFE factor = 4). The voxel size for each of the three datasets is 1.00 x 1.00 x 7.71mm, where the direction of the lower resolution is different for each of the three acquisitions. Depending on the heart rate, cardiac phases were recorded with a temporal resolution of about 30ms with 24 to 30 phases.

The dataset used for mitral valve detection and tracking consists of 50 subjects, including 47 CRT candidates and three healthy volunteers. The valve detection and tracking are performed on MR images in three LA views. The data used for the evaluation of the 4D dense cardiac motion tracking includes 12 subjects where both cine and 3D tagged images are available, six of which are healthy volunteers and six are CRT candidates.

4.2.2 Image pre-processing

The analysis of cardiac motion information from different MR images requires a common spatial and temporal reference space. However, there are three major difficulties: (i) the presence of tags in 3D tagged images obscuring the anatomy, (ii) differences in position caused by

respiratory and patient motion within sequences and across sequences and (iii) variable temporal resolution of the different image sequences. Camara et al. [Camara et al., 2009] presented a registration algorithm to correct the spatial misalignment, but temporal misalignment is not included. We extend this framework for the combination of information derived from untagged and 3D tagged MR image sequences which accounts for spatial misalignment as well as differences in temporal resolution.

Temporal alignment

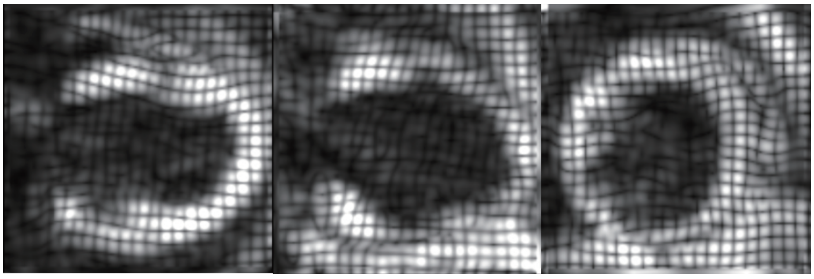
Each frame of a MR image sequence contains a DICOM meta-tag describing the trigger time, which is the reference point in the physiological signal that triggers the MR data acquisition. The trigger time defines how many milliseconds after the previous end-diastolic phase the acquisition of the current frame was triggered. We define T^s and T^e as the trigger time of the first and last phase respectively and N as number of frames. Hence each image sequence, e.g. short-axis (SA), horizontal long axis (4CH), vertical long axis (2CH), 3-chamber (3CH) untagged and 3D tagged images, has such a pair of trigger times. We define T^s_{ref} and T^e_{ref} as the maximum value of the trigger times respectively and N_{ref} as minimum value of the number of frames in each sequence. The common temporal resolution is then defined as $\Delta t = (T^e_{ref} - T^s_{ref})/N_{ref}$. All image sequences are resampled to this common temporal resolution using nearest neighbour interpolation between time frames.

Spatial alignment

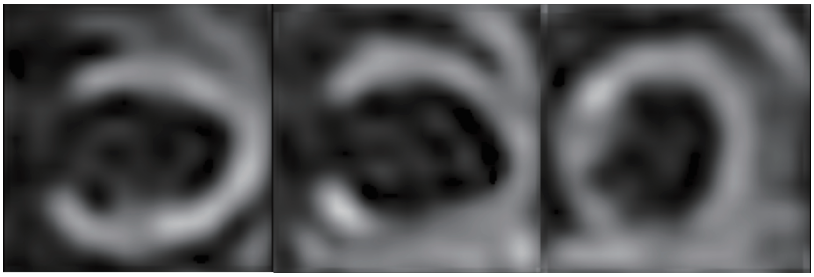
The 3D tagged MR images used in this chapter are free from respiratory motion artifacts since respiratory navigators are used during the acquisition. They contain complete 3D motion information with isotropic sampling in all three directions. Thus, they provide an ideal common spatial coordinate system. The only difficulty is the presence of tags in the image that obscure the anatomical information which is needed to align to the untagged MR images.

Several techniques for tag removal exist [Manglik et al., 2004, Qian et al., 2007, Camara et al.,

2009], yet none of these provided satisfactory results due to the fact that the 3D tagged MR images used here are dominated by tag patterns and show little of the underlying anatomy. However, one can easily extract the low frequency band by applying a FFT, band-pass filtering and inverse FFT. We have performed this naive but effective approach for tag removal and then averaged all three 3D MR images with tag removed to generate the 4D pseudo-anatomical image. Thereafter, all SA and LA untagged cine MR images are registered to the 4D pseudo-anatomical MR image using rigid registration: The registration transformation is modeled as a 3D rigid transformation between the untagged cine image sequence and the pseudo-anatomical image sequence. Additionally a 2D inplane rigid transformation is used for every cine slice to allow for misregistration between slices as the result of different breath-hold positions. Figure 4.2 shows an example of a 3D image before and after tag removed.



(a)



(b)

Figure 4.2 – This figure shows (a) an average image of three 3D tagged images, (b) an average image of three 3D images with tag removed: the 3D pseudo-anatomical image.

4.2.3 Detection of cardiac landmarks

We propose a fully automatic approach to localise and track landmarks of the heart in LA views to provide essential information for myocardial motion tracking and segmentation. This approach is illustrated in Figure 4.3 for detection and tracking of the mitral valve points across

three LA views. A-priori knowledge about the position of the landmarks is obtained based on a statistical analysis of the location of landmarks. We identify regions which are likely to contain the landmarks of interest and then apply a machine learning based landmark detector to voxels within this region.

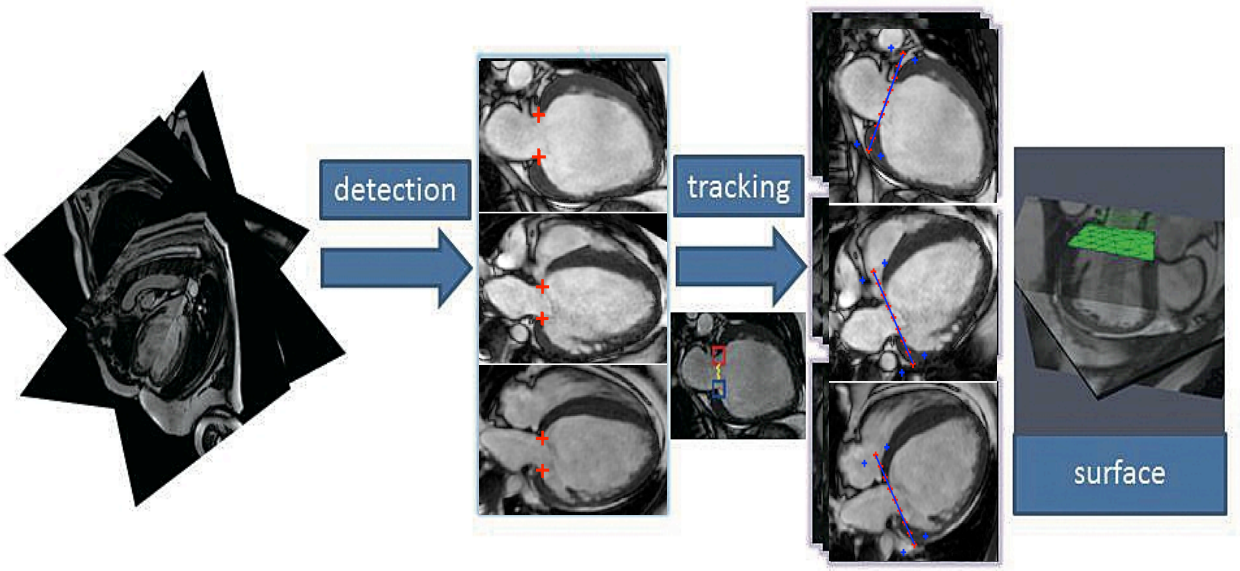


Figure 4.3 – The workflow of valve plane detection and tracking. The valve points are simultaneously detected and tracked in multiple views, and finally each of a series of valve plane surfaces is defined at each time point by the detected valve points in three LA views.

Valve point modelling

During the cardiac cycle, the mitral valve annulus undergoes mainly longitudinal motion along with a number of other types of deformations including circumferential and radial motion. In a LA view, the valve plane can be modelled by a line between two valve points in each LA view. However, due the large variability in the cardiac anatomy across subjects, its position, orientation and appearance can differ significantly across subjects. Lu et al. [Lu et al., 2010] presented a joint context method for landmark detection, but only one LA view was considered. In addition, a-priori knowledge about the position of the valve points can be exploited. As illustrated in Figure 4.4a, the line of the intersection between the 2CH view and the middle slice of the SA view, and the line of the intersection between 4CH and 2CH intersect at point **O**. This point **O** is usually approximately half way between the valve annulus and the apex

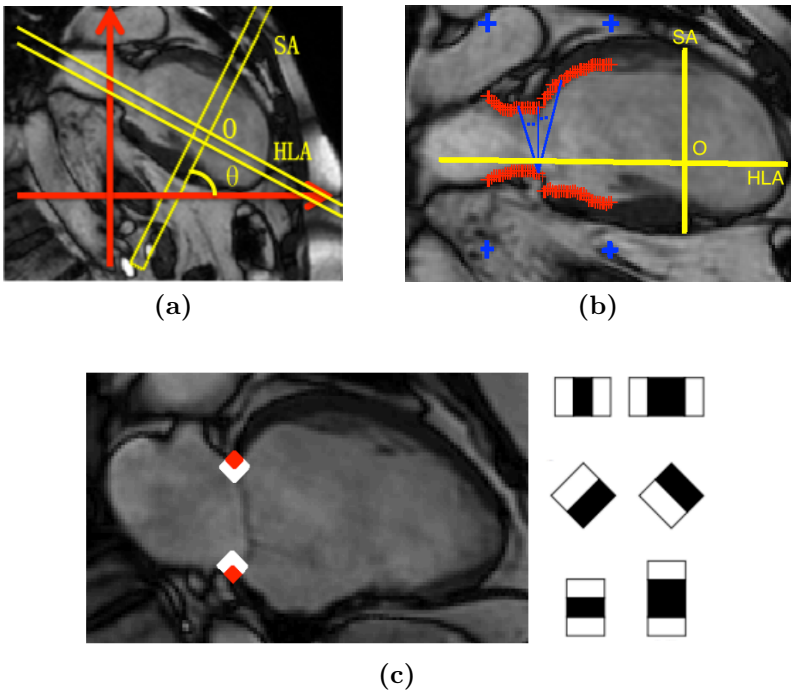


Figure 4.4 – Automatic detection of valve points. (a)A 2CH view of the heart showing the orientation of the SA view and 4CH (HLA) view. (b) An example shows a bounding box (defined by 4 blue points) for the valve plane containing the possible candidate point pairs and, (c) Some of the Haar-like features used for detection.

and therefore can be used to roughly anchor the valve plane. We refer to this point **O** as the centre of the LV in this chapter. A bounding box can be generated relative to the point **O** indicating the likely location of the valve plane, marked by 4 blue points in Figure 4.4b. A Gaussian mixture model is applied to classify the voxels in the LA images into air, soft tissue or blood. Only those voxels labelled as soft tissue are considered as candidate valve points. For each candidate valve point $\mathbf{x}_a$ its normalised distance to the border of the bounding box can be used to model the likelihood for a valve point at this location.

Ideally, a second valve point is located opposite to valve point $\mathbf{x}_a$ on the other side of the intersection line of two LA views. As the two valve points can provide relative location information, we combine the two valve points as one object. The line segment connecting two valve points is parallel to SA. To exploit this direction information, LA views is rotated to be horizontal so that the searching of $\mathbf{x}_b$ for $\mathbf{x}_a$ is more computationally efficient. In practice, the valve annulus deforms, we therefore include k points in the neighbourhood of $\mathbf{x}_b$ to pair up with every $\mathbf{x}_a$ so that we have a set of candidate valve point pairs

$\{(\mathbf{x}_a, \mathbf{x}_b(1)), (\mathbf{x}_a, \mathbf{x}_b(2)), .., (\mathbf{x}_a, \mathbf{x}_b(k))\}$ for each candidate location.

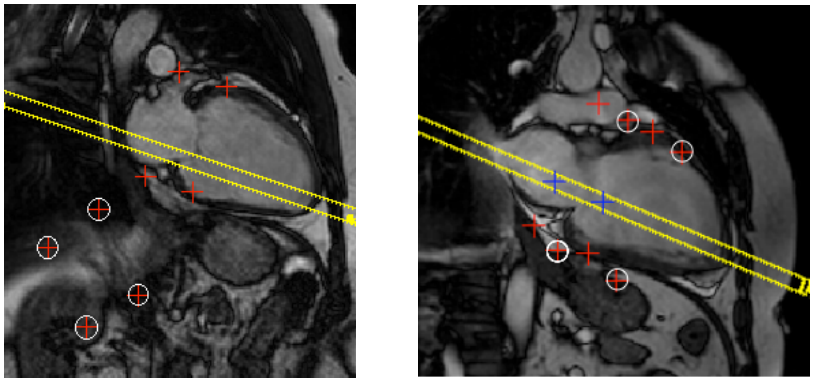
Simultaneous valve plane detection in multiple views

Most clinical cardiac MR acquistions include multiple LA views such as 2CH, 4CH and 3CH views. All three views can provide useful complementary information to help reduce the ambiguity and complexity in landmark detection. We therefore construct three different detectors for the three LA views to detect a pair of valve points in 2CH, 3CH and 4CH views.

AdaBoost [Viola and Jones, 2002]is a weighted voting algorithm that selects weak classifiers to combine into a strong classifier. In our proposed method, two layers of Adaboost are cascaded for each detector to avoid the training to be biased by negative samples of which there are 10 times more than positive samples. As different feature sets are used for the two layers, the hypotheses from the first layer are maintained to be combined with the result from the last layer. Then these pairs of points are ranked by the likelihood tested from the Adaboost classification. More detailed information about Adaboost can be found in section 2.3.2. For any machine learning based approach, a meaningful feature set which can distinguish the landmark or structure from other anatomical structures efficiently is crucial for the success of the detector. In this work a set of Haar-like features [Viola and Jones, 2002] are used, including the typical edge feature, line features and centre-surrounding features. In addition, we designed a special type of feature for the anatomic structure of valve point, as illustrated in Figure 4.4c, to represent the the difference between surrounding region in white (for blood-pool) and the region in red (for valve points).

Two further procedures are applied to combine the detection results from the three views. The first step is to check whether the intersection line of LA plane and SA plane goes through the detected regions in the middle. As the intersection line is the long axis of the LV, the region surrounding the valve annulus should be distributed symmetrically around this line. The region in Figure 4.5a which is marked by the circled landmarks will be ignored by this procedure. Next, the ROIs identified in the 2CH and 3CH views are projected into the 4CH view. The region detected in the 4CH view with the smallest distance to these projected ROIs is regarded as the

correct one. A second detected region marked by the circled landmarks is further downstream of the long-axis of the LV in Figure 4.5b.



(a) (b)

Figure 4.5 – Two further steps are applied to combine the detection results from the multiple views: (a) Check whether the intersection line of LA plane and SA plane goes through the detected regions in the middle. The region marked by circled landmarks will be screened out. (b) Center of the ROIs identified in 2CH and 3CH are projected into 4CH, marked by the blue cross-hair. The region detected in 4CH with the smallest distance to these projected ROIs is regarded as the right one. A second detected region marked by circled landmarks is further downstream of the long-axis of the LV.

4.2.4 Valve plane motion tracking

To track the motion of the valve annulus we conduct 2D collaborative tracking in three LA views. It is reasonable to make the assumption that the valve annulus mainly undergoes deformation in the long-axis direction during the cardiac cycle and is accompanied by contraction and expansion perpendicular to the long-axis direction. Hence the tracking is initially performed along the long axis and then performed in the direction towards or away from the centre of LV. When accuracy of the tracking is affected by noise or sudden motion in one view, the tracking in other views may be less affected and hence produces good overall performance

The valve annulus are tracked simultaneously in three LA views using template matching. Two patches of size of 20×30 are defined in each of the three LA view, encompassing the valve end points. In total six patches are tracked simultaneously in three LA views. The cross-correlation

is used as the similarity metric, which is computed as follows:

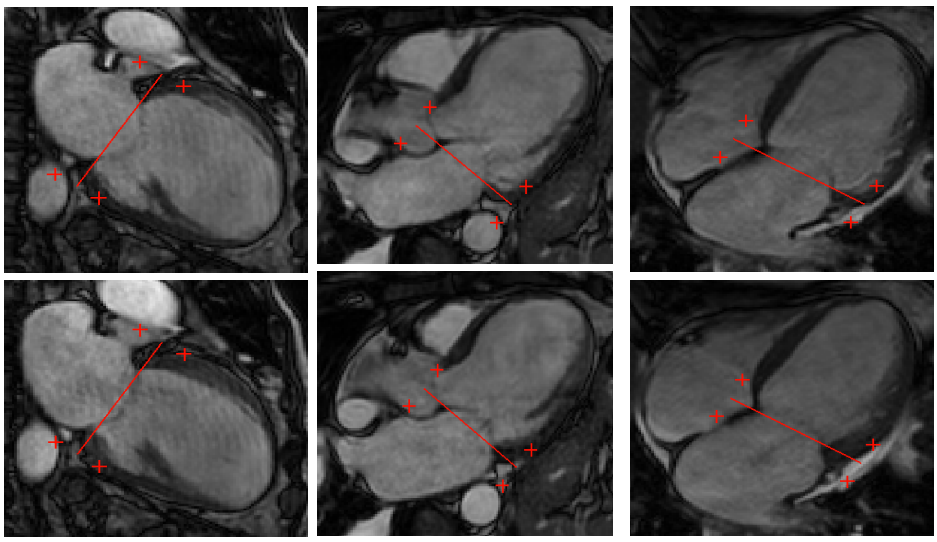
$$C(I_{t_0}, I_t) = \frac{\sum_{x,y}(I_{t_0}(x,y) - \bar{I}_{t_0})(I_t(\mathbf{T}(x,y)) - \bar{I}_t)}{\sqrt{\sum_{x,y}(I_{t_0}(x,y) - \bar{I}_{t_0})^2}\sqrt{\sum_{x,y}(I_t(\mathbf{T}(x,y)) - \bar{I}_t)^2}} \tag{4.1}$$

Here I_{t_0} corresponds to the first time frame and I_t corresponds to any other frames of image. The collaborative similarity is computed by combining the similarities in three LA views:

$$C_{all} = \omega_{2CH}C_{2CH} + \omega_{4CH}C_{4CH} + \omega_{3CH}C_{3CH} \tag{4.2}$$

Here ω_{2CH} and ω_{4CH} are set to 1 and ω_{3CH} is set to 0.8. These parameters are set empirically. The ω_{3CH} is set to be lower due to the fact that the valve points in the 3CH view are more ambiguous to be identified. This observation is verified by the valve point detection results presented in the result section 4.3. The similarity is calculated between each frame image and the first image in the sequence (ED phase image) to avoid the accumulation of motion errors.

Figure 4.6 shows the example of the valve points tracking in the three LA views. Images in ED and ES phases are shown.



(a) **(b)** **(c)**

Figure 4.6 – Example for tracking in the (a)2CH view, (b) 3CH view and (c)4CH view. Images in ED and ES phases are shown.

4.2.5 Comprehensive motion tracking

During the cardiac cycle, the left ventricle undergoes a number of different deformations including circumferential, radial and longitudinal motion. While the 3D tagged MR images provide good information about all aspects of the motion, the SA images may provide more information of radial motion and the LA images may provide some extra information about the radial and longitudinal motion.

Consider a material point in the myocardium at a position $\mathbf{x} = (x, y, z)^T$ at time $t_0 = 0$ that moves to another position $\mathbf{x}' = (x', y', z')^T$ at time $t_i = i\Delta t$ where Δt is the time interval between two consecutive phases and i corresponds to the time frame. The goal of the motion tracking is to find the transformation $\mathbf{T}$ for all time phases i such that $\mathbf{T}(\mathbf{x}, t_i) = \mathbf{x}'$. In this work we represent $\mathbf{T}$ using a series of free-form deformations [Rueckert et al., 1999] as described in [Chandrashekara et al., 2004a].

The estimation of the deformation field $\mathbf{T}$ proceeds in a sequence of registration steps. Using the spatial and temporal alignment described in section 4.2.2, all image sequences have been mapped into a common spatial and temporal coordinate system. We label the myocardium of the left ventricle at ED phase of the untagged MR images using an automatic segmentation tool [Shi et al., 2011].

In addition, a gradient detector is used to highlight the epicardial and endocardial contours on untagged MR images. The information from both segmentation and the gradient-detector is combined into a spatially varying weighting function which moderates the influence of the tagged and untagged images on the motion tracking. We then register the images taken at time t_1 to the reference image at time t_0 and obtain a transformation representing the motion of the myocardium at time t_1 [Rueckert et al., 1999]. We use the resulting transformation as an input for the next time frame and continue this process until all the time frames in the sequence are registered to the first phase [Chandrashekara et al., 2004a].

Weighted similarity measure

To exploit the complementary nature of the tagged and untagged MR images we have developed a spatially adaptive weighting function that accounts for the different types of information available: The 3D tagged images characterise well the motion inside the myocardium while the untagged SA and LA images characterize the motion well at the epi- and endocardial borders of the myocardium. Outside of the myocardium there is no useful information for cardiac motion tracking apart from papillary muscles. Thus, we would like to generate a weighting function that is zero outside the myocardial region, that maximizes the weighting of the tagged images within the myocardium and increases the influence of the untagged images at the myocardial border. The spatial weights for the tagged and untagged images are generated for the reference image used for the registration.

The weighting for the untagged images, $W^u(\mathbf{x})$, is generated by multiplying the gradient of the segmentation with the gradient of the image intensity. Higher weighting indicates a higher probability for the presence of a myocardial border which is critical to the estimation of radial motion. Let L denote the segmentation of the short-axis MR image I. This segmentation assigns a label $\Lambda = \{L_{bg}, L_{myo}, L_{blood}\}$ to every voxel as background, myocardium and blood pool. The probability for myocardium $P(\mathbf{x}, L_{myo})$ can be derived from myocardium segmentation's intensity distribution from applying Gaussian smoothing. Furthermore assume that ∇I_σ denotes the gradient of image I after convolution with a Gaussian kernel G with standard deviation σ. The weights for the untagged MR image are defined as follows:

$$W^u(\mathbf{x}) = \frac{||\nabla I_\sigma(\mathbf{x})||||\nabla P_\sigma(\mathbf{x}, L_{myo})||}{\max\left(||\nabla I_\sigma(\mathbf{x})||\right)\max\left(||\nabla P_\sigma(\mathbf{x}, L_{myo})||\right)} \tag{4.3}$$

Here $||\nabla I_\sigma(\mathbf{x})||$ and $||\nabla P_\sigma(\mathbf{x}, L_{myo})||$ are the gradient of intensity and the gradient of myocardium probability at location $\mathbf{x}$ after convolution with a Gaussian kernel G with standard deviation $\sigma = 10mm$ respectively.

The weights for the 3D tagged image are defined as follows and Figure 4.7 illustrates an example

of the resulting spatial weight map.

$$W^t(\mathbf{x}) = \begin{cases} 1 - W^u(\mathbf{x}), & \text{if } L(\mathbf{x}) = L_{myo} \\ 0, & \text{otherwise} \end{cases} \tag{4.4}$$

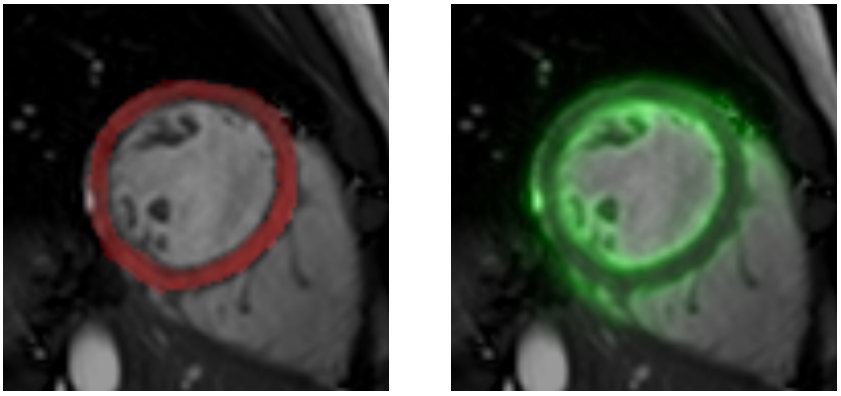


Figure 4.7 – This figure shows a spatial weight map indicated by the colour overlay. Red and green colours indicate the weights for tagged images and untagged images respectively. The transparency of the colour indicates the magnitude of weight.

Given a weight map, we define the similarity between two images I_A, I_B as the weighted normalized cross-correlation between the image intensities:

$$S(I_A; I_B; W, \mathbf{T}) = \frac{\sum W(\mathbf{x})(I_A(\mathbf{x}) - \mu_A)(I_B(\mathbf{T}(\mathbf{x})) - \mu_B)}{\sqrt{\sum W^2(\mathbf{x})(I_A(\mathbf{x}) - \mu_A)^2(I_B(\mathbf{T}(\mathbf{x})) - \mu_B)^2}} \tag{4.5}$$

For simultaneous registration of the untagged and 3D tagged images, the correlation is computed separately across tagged images and untagged images and combined into a single similarity measure.

Constrained myocardial motion tracking using tagged and untagged MR images

The valve plane is an important landmark for accurate cardiac motion estimation but is not clearly visible in untagged SA images and tagged SA images. From the LA views, we have tracked the valve annulus using the collabrative tracker described in section 4.2.4 and construct

a valve plane surface V_t for each time frame. Figure 4.8 shows a constructed valve plane used to constrain myocardial motion tracking.

To estimate the motion within the myocardium, a weighted similarity measure described in section 4.2.5 is used. The constructed valve plan surface V is applied as one of the constraints in non-rigid image registration to correlate the myocardial motion to the tracked valve plane in form of a penalty term:

$$C_v = -\frac{\sum_{\mathbf{x}\in V_{t_0}} D(\mathbf{T}(\mathbf{x}), V_{t_i})}{|V_{t_0}|} \tag{4.6}$$

where V_{t_i} denotes the valve plane surface at time t_i and D is the surface distance operator (symmetric closest points).

Therefore, the cost function for the comprehensive dense motion tracking of myocardium can be defined as follows:

$$C = \{\sum_{s\in T} |\Omega_s| \left[S(I_{t_0}^{t,s}, I_{t_i}^{t,s}, W^{t,s}, \mathbf{T})\right] + \sum_{s\in U} |\Omega_s| \left[S(I_{t_0}^{u,s}, I_{t_i}^{u,s}, W^{u,s}, \mathbf{T})\right]\} / \sum_{s\in U,T} |\Omega_s| + \lambda C_v \tag{4.7}$$

Here $|\Omega_s|$ denote the sum of weights in the image. $\mathbf{T}$ and $\mathbf{U}$ denote tagged and untagged image correspondingly. λ is the coefficient for valve plane constraint C_v.

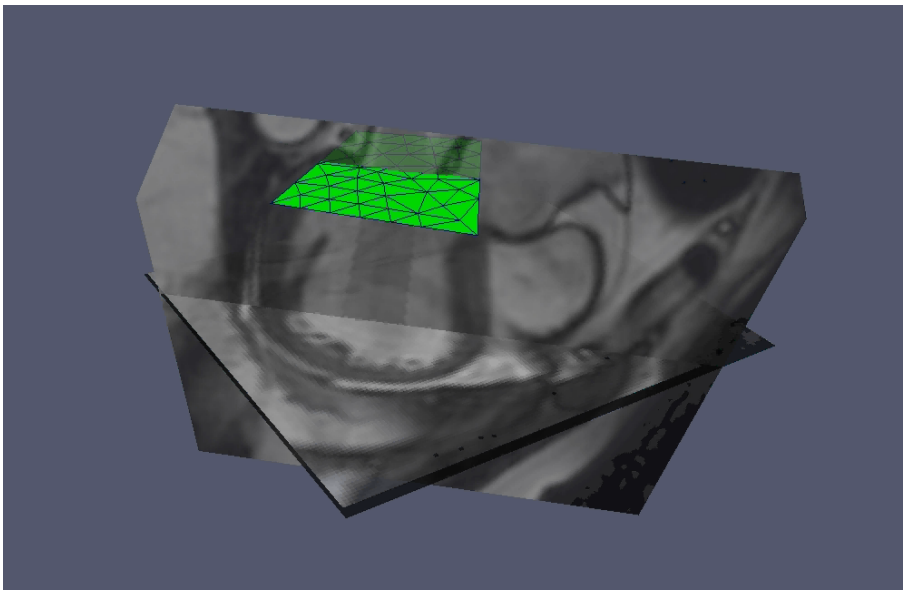


Figure 4.8 – This figure shows the detected valve motion used to construct a valve plane to constrain myocardial motion tracking. The valve points are simultaneously detected and tracked in multiple views, and a mitral valve plane surface is reconstructed at each time point from the tracked valve end-points via triangulation.

4.3 Experiments and results

4.3.1 Results for valve detection and tracking

In valve annulus detection, the proposed collaborative cascaded Adaboost based detector is applied on images in 2CH, 3CH and 4CH views in ED phase for each of the 50 subjects. Five-fold cross-validation is used in order to evaluate the detection performance. The mitral valve position is manually annotated in ED phase under the guide of clinician. The accuracy of the detection error for each pair of mitral valve point is computed as the average of Euclidean distance between the detected valve points position and their corresponding ground truth.

We refer to the obtained valve points as mis-detected if their detection error is larger than 15mm. We empirically set this value (15mm) to balance the accuracy and false detection. The valve points with the detection error marginally below this threshold are observed to be tracked with reasonable accuracies and can still provide useful constraints if the tracking results are to be used for dense myocardial motion tracking. This threshold can vary depending on how we use the valve position. In the MICCAI 2012 STACOM LV landmark detection challenge, Lu and Jolly [Lu and Jolly, 2013] rejected detected valve points only when their detection errors are larger than 30mm.

The number of validly detected pair of valve points in each of the LA views and detection performance are summarised in Table 4.1 and Figure 4.9. In addition, mitral valve points are detected for 47 subjects in at least one of the LA views.

Table 4.1 – Detection errors of the detected valve points from the ground truth positions in three LA views.

	2CH	3CH	4CH
Number of detected pairs of valve points	42	33	44
Mean (mm)	5.69	4.06	6.46
STD(mm)	4.15	3.00	3.76
Median(mm)	4.21	3.10	6.48

We evaluate the performance of our tracking algorithm by measuring the average Euclidean

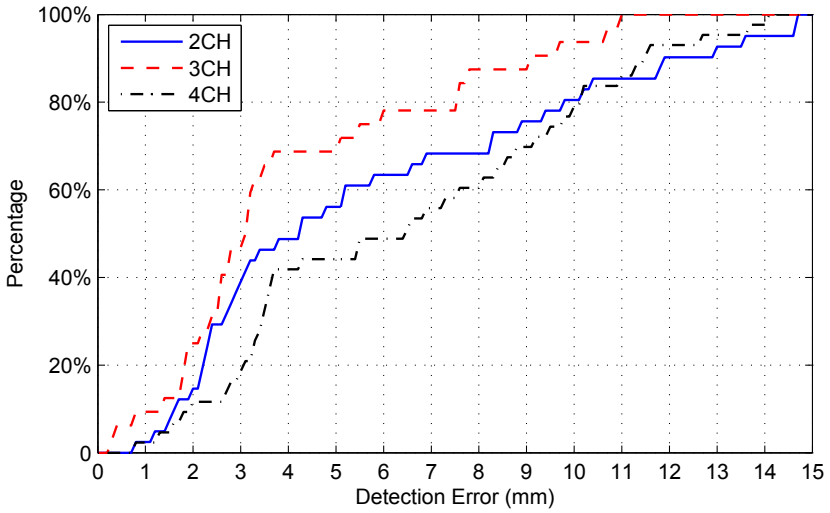


Figure 4.9 – Mitral valve detection accuracy. A point on the curve denotes the percentage of subjects of which the detection error from ground truth is less than its corresponding error value in x-axis. For example, for the 3CH view, the detection error for 60% subjects is less than 3.2mm.

distance between tracked the valve points and their corresponding ground truth in ES phase image for 50 subjects and the average error is $2.10 \pm 1.42mm$.

We calculate the displacement of the mitral valve as the Euclidean distance between the middle points of the line segment connecting two valve points in the ED phase and other time phase. The mitral valve annulus displays an average of $0.87cm$ maximum longitudinal shortening during the cardiac cycle in 2CH view. Comparing this to the $1.3cm$ displacement observed by Marcus *et al.* [Marcus et al., 1999] for 57 healthy volunteers, it can be speculated that the cardiac diseases may have a significant impact on the motion of mitral valve. Figure 4.10 shows that the maximum magnitude of displacements of mitral valve detected in the 2CH view and in the 4CH are well correlated to each other (Pearson correlation coefficient $CC = 0.64$).

As shown in the valve point detection results in Table 4.1, not all valve points are detected. Only 32 subjects have valve points detected in all the three LA views. Figure 4.11 shows some examples of applying proposed tracking approach on 10 such subjects, for whom the tracking starts with the detected valve points. The valve points positions at each frame are also manually traced as ground truth. In some cases the results from automatic detection and tracking are perfectly in line with that of ground truth (Figure 4.11 (a)(b)). Whereas, in other cases the tracking started with points that are not perfectly located at the true valve positions, but the

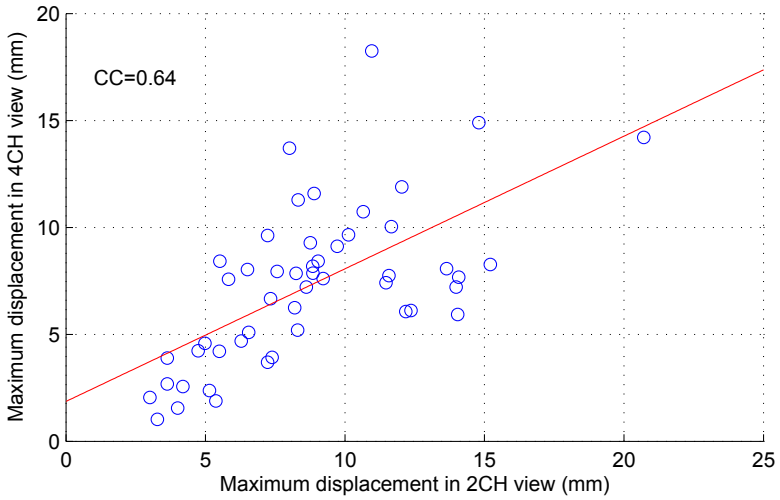


Figure 4.10 – Scatter plot of the maximum magnitude of displacement of the mitral valve in the 2CH view vs in the 4CH view. The the magnitude of displacements in two views are well correlated to each other.

tracking still yielded reasonable accuracy (Figure 4.11 (c)(d)). Across the whole cardiac cycle, the average tracking accuracy of $2.03 \pm 1.36mm$ per phase is obtained. The average tracking error in the ES phase is $5.71 \pm 5.31mm$, which not surprisingly is larger than that of the tracking starting with manually annotated valve points in ED phase. The tracking accuracies of these 10 cases are reported in Figure 4.12.

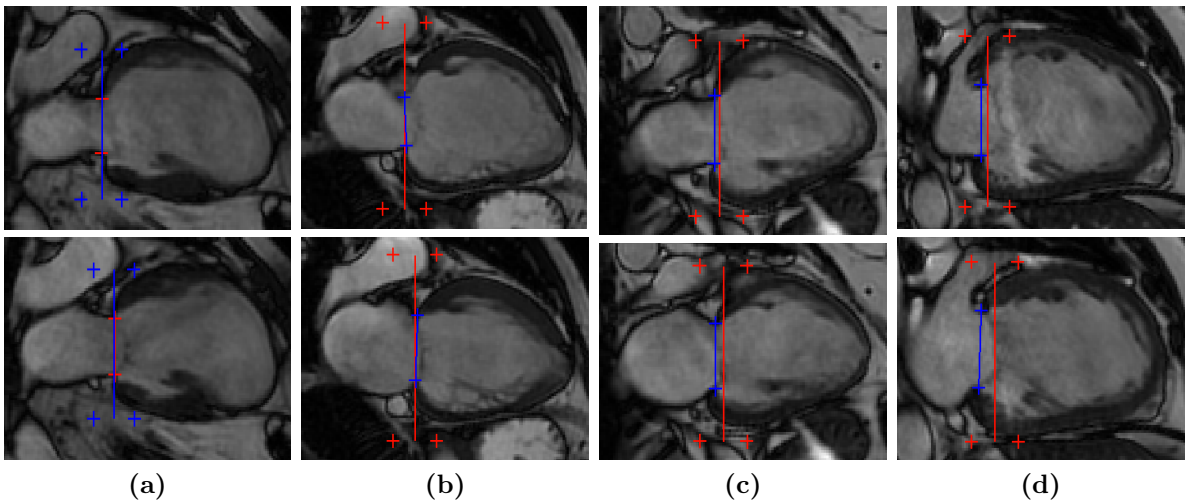


Figure 4.11 – Examples for automatic valve detection and tracking results (in red) and ground truth (in blue). Images in ED (upper row) and ES (bottom row) phases are shown. (a)(b) show that the results from automatic detection and tracking are perfectly in line with that of ground truth. (c)(d) show that the errors exit in detection and tracking results.

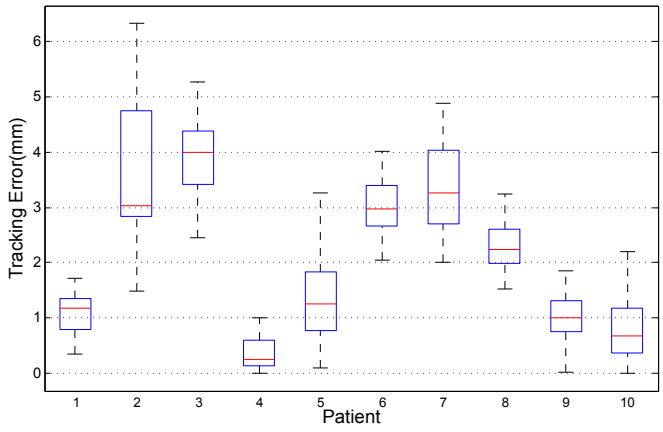


Figure 4.12 – Automatic detection and motion tracking results. Tracking errors are estimated as distances between tracked valve positions and ground truth positions at each frame across the cardiac cycle. The edges of the blue box are $25th$ and $75th$ percentiles.

4.3.2 Results for dense myocardial motion tracking

The data used for the evaluation of the 4D cardiac motion tracking includes 12 subjects where both cine and 3D tagged images are available. The cardiac motion estimation should include radial, circumferential and longitudinal motion. An accurate tracking of the endocardial and epicardial boundaries on short-axis MR images indicates good radial motion estimation. Similarly, the accurate tracking of the endocardial and epicardial boundaries on the long-axis MR images indicates good longitudinal motion estimation.

In order to evaluate the proposed motion tracking method, we have manually tracked 16 landmarks in 3D in each dataset. We select one landmark close to the center of each AHA segment excluding the apex. We then compare the position of landmarks predicted by the automatic motion tracking with that of landmarks manually traced. In addition, we have analysed the relative inter-observer variability on a subset of 3 datasets (one patient and two normal subjects). The relative inter-observer landmark tracking error is $7.8\% \pm 4.4\%$. The results of relative inter-observer surface tracking error are shown in Table 4.2. Figure 4.13 shows the relative landmark error when comparing the results of registration-based motion tracking with that of manual tag tracking. Four situations include motion tracking: using the tagged image only, using the untagged images only, using both tagged and untagged images and using both tagged and untaged images with the valve plane constraints. The valve plane

is obtained by tracking the manually annotated valve points in ED phase. The results indicate that the combined registration using tagged and untagged images performs better than the registration using either tagged or untagged images alone. Moreover, the valve plane tracking helps longitudinal motion estimation so the error is reduce. Overall, the comprehensive method performs best. The median error of healthy controls over all frames is 1.44mm for landmarks tracking, 1.04mm for endocardial surface tracking and 0.76mm for epicardial surface tracking using the comprehensive method.

Table 4.2 – Inter-observer variance of the relative error for the surface tracking of the Endo-(ED) and Epicardial (EP) contours for the different SA and LA views.

	SA	2CH	4CH
EP	$4.3\% \pm 1.2\%$	$3.4\% \pm 0.6\%$	$3.5\% \pm 1.8\%$
ED	$7.1\% \pm 2.1\%$	$4.3\% \pm 1.4\%$	$3.9\% \pm 0.8\%$

Figure 4.14 illustrates three types of derived myocardial motion fields in both short-axis and long-axis views, using only untagged images, only tagged images and both untagged and tagged images respectively. Visualisation of the radial, circumferential and longitudinal motion in different direction from an example subject is shown in Figure 4.15.

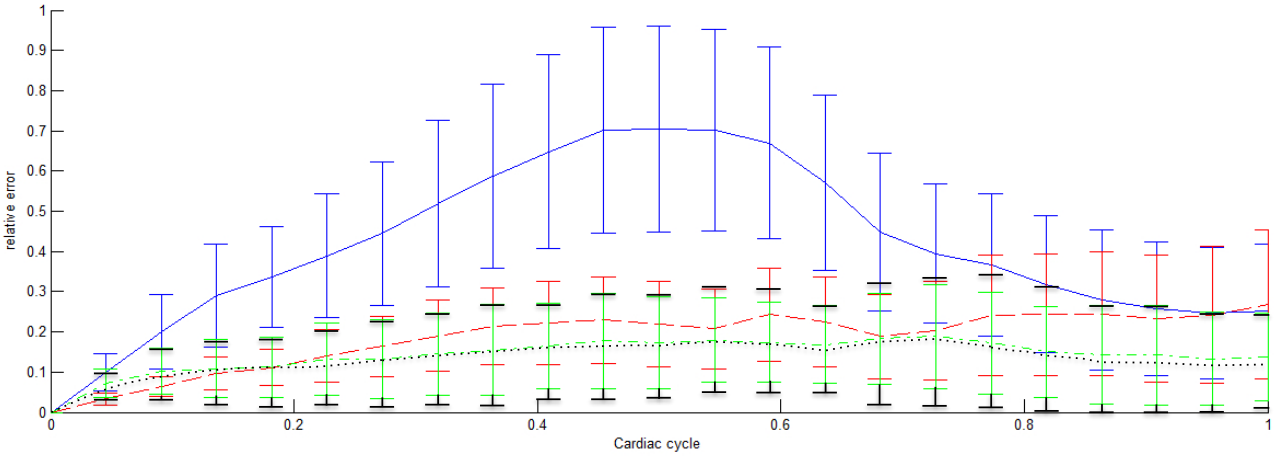
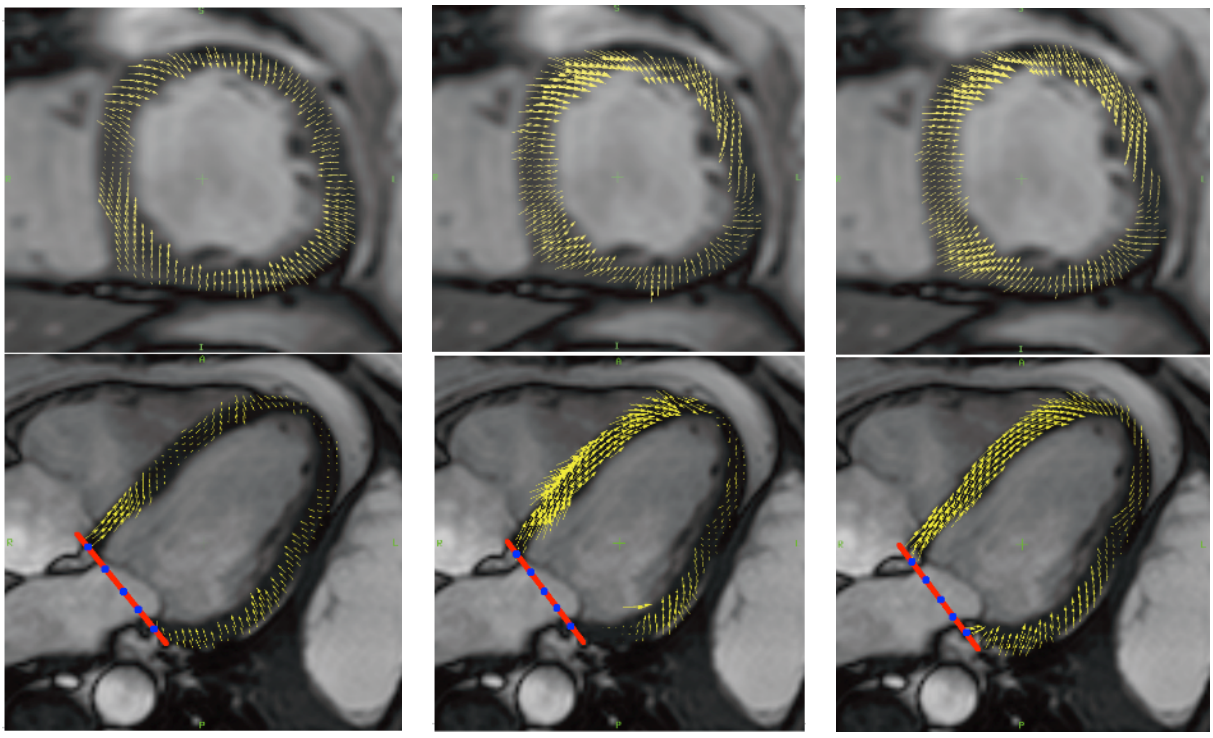
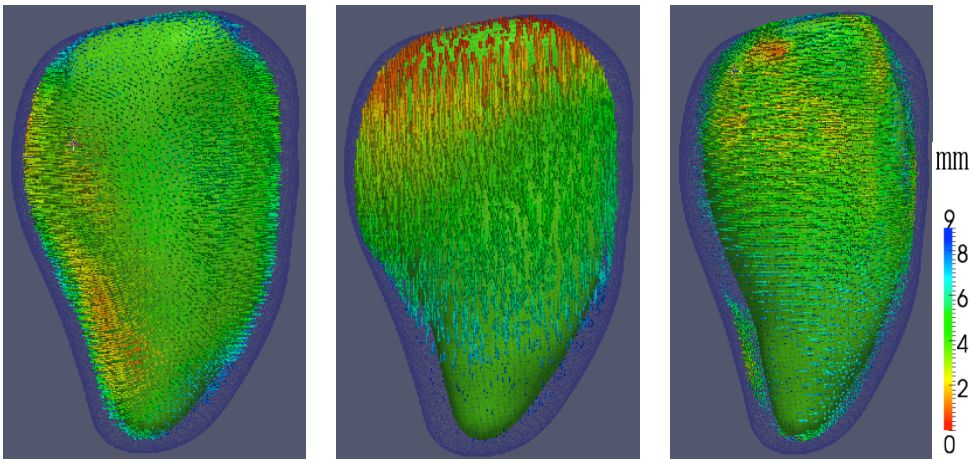


Figure 4.13 – This figure shows the relative landmark error when comparing the results of registration-based motion tracking with that of manual tag tracking. The blue solid line indicates the results using untagged images only, the red dash line shows the results using 3D tagged images only, the green dash-dot line shows the results using the combined motion tracking using both the tagged and untagged MR images and the black round dot line shows the error using both the tagged and untagged images with the valve plane constrains. The lines are the mean error and the bars indicate the variance [Shi et al., 2012].



(a) (b) (c)

Figure 4.14 – Visualization of the myocardial motion field. (a)motion field derived from untagged images only, (b) shows the motion field derived from tagged images only and (c) shows the motion field derived from the combined motion tracking.



(a) (b) (c)

Figure 4.15 – This figure shows respectively (a)radial, (b) longitudinal and (c) circumferential motion at end systolic phase of an example subject.

As shown in Figure 4.13, the combined motion tracking using tagged and untagged images performs better than using either tagged or untagged images alone. By applying the valve plane constrains, the tracking results have an reasonable improvement.

4.4 Conclusions

In this chapter we have presented a machine learning method to identify and track a set of anatomically important landmarks in three different views throughout the cardiac cycle. These landmarks are used to significantly reduce the number of degrees of freedom for the registration algorithm and the associated computational complexity. The three LA image sequences provide heart chamber shape information and anchor some important landmarks. The information from the three LA views were also applied to complement each other in cardiac structure detection and tracking, which improved the detection accuracy. The displacements of the mitral valve tracked in the LA planes during the cardiac cycle are well correlated. For mitral valve detection, more information can be used to improve the detection accuracy. For example, temporal coherence presented in a long axis slice sequence can be exploited to help with mitral valve point detection accuracy. In addition, due to the large variation in heart appearance and anatomic structure (such as mitral valve location, scale), large set of cardiac MR images are needed for classifier training to fully cover these variations. Figure 4.16 presents examples of those cases with large detection errors, which cause the mis-detection on these cases.

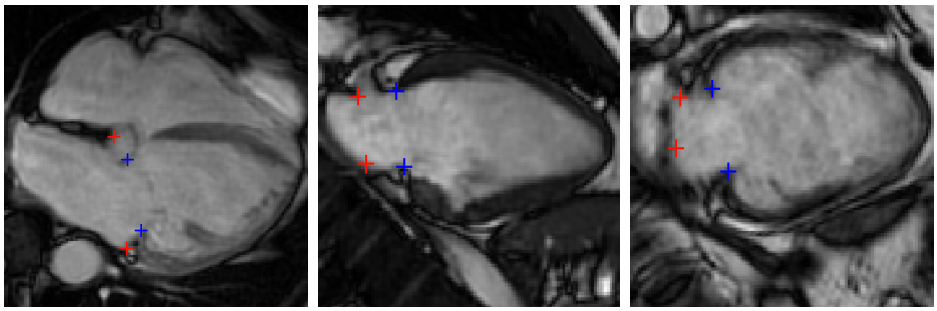


Figure 4.16 – Examples of images with large detection errors (large than 15mm in our experiment) due to the large variation of anatomic structure.

Additionally, we have shown how the proposed approach can be used to perform dense cardiac motion tracking using both 3D tagged and untagged image sequences from short-axis and long-axis views simultaneously. The key advantage of the proposed method is the simultaneous use of complementary motion information contained in the tagged and untagged images. The automatic detection and tracking of a larger number of cardiac landmarks will enable better

hybrid sparse and dense motion tracking with a view to improve speed and accuracy of the motion tracking. In the next chapter we propose an automatic framework to select and track a set of sparse distinctive landmarks in the presence of relatively large deformations in order to capture the myocardial motion in cardiac MR sequences.

Chapter 5

Sparse cardiac motion tracking and functional analysis

The work in this chapter is based on the following published papers:

- Haiyan Wang, Wenzhe Shi, Xiahai Zhuang, Xianliang Wu, Kai-Pin Tung, Sebastien Ourselin, Philip Edwards, and Daniel Rueckert. (2013). Landmark detection and coupled patch registration for cardiac motion tracking. Proc. SPIE 8669, Medical Imaging 2013: Image Processing, 86690J, pages 86690-J1-J6.
- Haiyan Wang, Wenzhe Shi, Wenjia Bai, Philip Edwards, and Daniel Rueckert. (2013). 4D Sparse Landmark Cardiac Motion Tracking and Regional Function Analysis. Proc. MIUA 2013. Pages 33-38

5.1 Introduction

Cardiac vascular disease (CVD) is the leading cause of death in the world, claiming 17.3 million lives a year according to the World Health Organisation (WHO) [Organization et al., 2012]. The accurate estimation of cardiac motion is becoming increasingly important for the quantification of the viability and contractility of the myocardium.

Extensive research on cardiac motion tracking has focused on approaches based on image registration and deformable model fitting techniques [Rueckert et al., 1999, Vercauteren et al., 2009, Chandrashekara et al., 2004a]. Methods of this class are very sensitive to the initialization of the motion tracking, especially in the presence of large motions. In addition, these dense motion tracking techniques are often computationally expensive. Alternatively, sparse landmarks can provide anatomy-specific constraints to establish correspondences between images being tracked or registered [Joshi and Miller, 2000, Shen and Davatzikos, 2002, Cootes et al., 2010, Zhu et al., 2010]. On the other hand, poorly selected landmarks can easily lead to poor or ambiguous correspondences. Hence it is vital to identify a set of accurate correspondences across a series of images which can be used to build a variety of models for registration and motion tracking.

Increasing attention has been drawn to the estimation of the deformation of the myocardium to aid the diagnosis of cardiac malfunction. Landmark tracking can provide sparse, anatomically relevant constraints to help establish correspondences between images being tracked or registered. However, landmarks on the myocardium are often characterized by ambiguous appearance in cardiac MR images which makes the extraction and tracking of these landmarks problematic.

In this chapter we propose an automatic framework to select and track a sparse set of distinctive landmarks in the presence of relatively large deformations in order to capture the myocardium motion in cardiac MR sequences. To achieve this, a sparse set of the landmarks is identified by using entropy and singular value decomposition (SVD). The resulting sparse set of landmarks is then tracked simultaneously using a multi-label Markov Random Field (MRF) to model the relationship between the neighbouring landmarks. The MRF formulation has the advantage of being able to search across a large search space to account for large deformations without huge computational cost. The novelty of our approach is the use of the two-stage model to capture different forms of deformations of the myocardium. This sparse landmark tracking can then be used to estimate cardiac motion, or to initialise dense motion tracking in cardiac image sequences. To study the clinical usefulness of the approach we assess the regional systolic dyssynchrony index (SDI) in CRT patients and healthy volunteers. The derived regional wall

thickness SDI for each of 47 subjects are compared with the clinical regional volume SDI measurements obtained using a commercial TomTec system [Kühl et al., 2004].

5.2 Methods

First, a set of sparse landmarks are identified by combining entropy and singular value decomposition (SVD). For landmark tracking, a two-stage multi-label MRF is applied to model the relationship between the neighbouring landmarks and deformation pattern. Two tracking approaches are proposed. First, the tracking of landmarks are performed frame by frame that all other frames of image are compared to the first frame to avoid accumulated error. Moreover, we further extend the approach to model more dependencies between landmarks and track them in 4D fashion. The overview of the methods is illustrated in Figure 5.1.

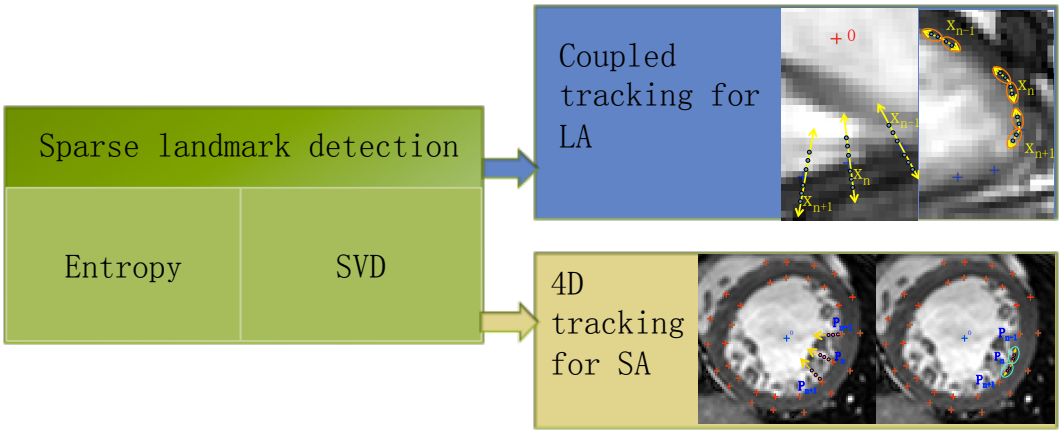


Figure 5.1 – This framework shows the overview of the methods used in this chapter.

5.2.1 Automatic detection of sparse landmarks

Increasing attention has been focused on the estimation of myocardium border deformation. In our proposed tracking approach we aim to select a group of landmarks on the myocardium border which can be best tracked across the cardiac cycle. Some points such as the mitral valve points are distinctive, but many other points on the myocardium boundary share similar

appearance and shape features, which leads to ambiguities when these points are being tracked. We propose an entropy-based landmark detector, which identifies landmarks that are recognisable in all frames throughout the cardiac cycle.

For the pixel $\mathbf{x}$ in frame f_1, the detector defines a similarity measure based on sum of square differences (SSD) between a patch centered at $\mathbf{x}$ and a patch centered at the pixel $\mathbf{y}_i$ in frame $f_n(n = 2..N)$, which are denoted by $\Omega_{\mathbf{x}}$ and $\Omega_{\mathbf{y}_i}$ accordingly in Figure 5.2. The size of the patches are $5mm \times 5mm$ and $\mathbf{y}_i$ is inside the search region R_n of the size of $40mm \times 20mm$ for $\mathbf{x}$ in frame f_n. The search region is defined according to the analysis of the distribution of the maximum displacement of the valve points in Chapter 4 (see Figure 4.10). Therefore a self-similarity map SM_n in frame f_n is generated for $\mathbf{x}$ as:

$$SM_n = \{SSD_i, \mathbf{y}_i \in R_n\}$$

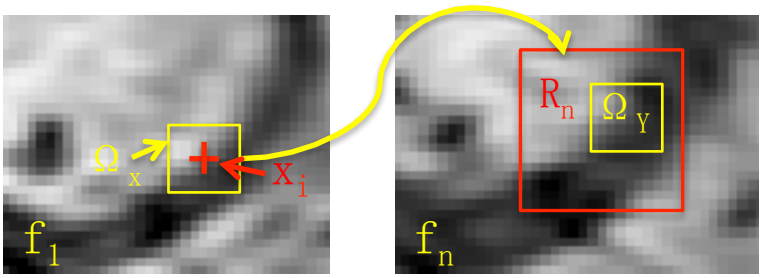


Figure 5.2 – This figures shows the search region R_n in frame f_n for pixel $\mathbf{x}$ when calculate the self-similarity map SM_n. The similarity between $\mathbf{x}$ and $\mathbf{y}_i$ is calculated based on patches $\Omega_{\mathbf{x}}$ and $\Omega_{\mathbf{y}_i}$.

Interestingly, when analysing different regions locally, we found that the corresponding distribution of the similarities were different. As shown in Figure 5.3a, the self-similarity map in the blood pool are more diffuse than those in the valve and apex regions. This is due to the fact that points in the valve and apex regions have more distinctive features which means that one is able to find good correspondences with high similarity. On the other hand, points in the blood pool have a more ambiguous appearance and thus less informative self-similarity maps will be found for those points.

Because we intend to discard the landmarks with less informative self similarity maps and

Shannon entropy can quantify information content in a variable, so we then compute the entropy H_n of this self-similarity map SM_n as follows:

$$H_n = -\sum_{k \in SM_n} p(k) \log p(k)$$

Here $p(k)$ denotes the probability mass function (PMF) of k which gives the probability that a random similarity value in SM_n is exactly equal to k. We then obtain the sum of the entropies for pixel $\mathbf{x}$:

$$H = \sum_{n=2..N} H_n$$

Here n denotes the frames other than the ED phase. The entropy is maximized if all possible self similarities are equally likely to occur. On the other hand, a low value of H_n is usually associated with less informative content (the value in the self-similarity map is either high or low), indicating a better alignment of the point $\mathbf{x}$ with only a small region in the searching region.

We hence assume that a low entropy of the similarity map corresponds to a more discriminative feature point. We choose four frames with an equal time interval from the whole image sequence to avoid unnecessary computations, and use these four frames to find a set of distinctive landmarks on the manually segmented contour of myocardium. This speeds up the landmark identification process substantially without losing generality.

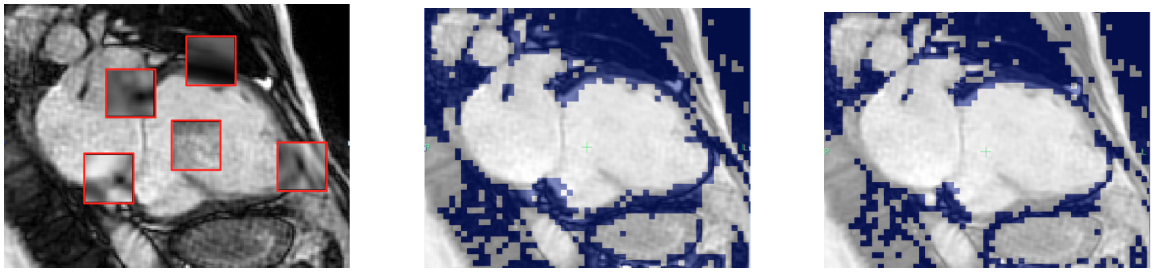
However, we noticed that the entropy approach detects a number of landmarks that have distinctive appearance but are static and hence are less interesting. Points that undergo large deformation need to be selected too in order to well represent the deformation of the endocardium. In order to pick the landmarks with larger deformation, we analyze the dynamic information across the cardiac cycle by applying SVD to the cardiac cine MR images. SVD seeks to find a low rank approximation for a matrix and has a wide range of applications, such as compression and noise reduction in image processing [Andrews and Patterson III, 1976, S Wack and D Badgaiyan, 2011].

Let a $M \times N$ matrix $\mathbf{X}$ represent a cine MR image sequence where each column of the matrix

corresponds to a frame in the sequence while each row of the matrix corresponds to a material pixel in the image. The SVD of $\mathbf{X}$ is defined as $\mathbf{X} = \mathbf{USV}^T$, where $\mathbf{U}$ is a $M \times M$ matrix, and $\mathbf{V}^T$ is a $N \times N$ matrix. The non-zero elements of the $M \times N$ diagonal matrix S are called singular values and can be sorted in descending order as $S = diag(s_1, ..., s_N)$. Then, $\mathbf{X}$ can be approximated by a matrix $\mathbf{X}^{(l)}$ of rank l,

$$\mathbf{X}^{(l)} = \sum_{k=1}^{l} u_k s_k v_k^T \tag{5.1}$$

The quality of the approximation can be measured by the ratio $\omega = \sum_{j=1}^{l} s_j / \sum_{i=1}^{N} s_i$. To achieve a given quality of the approximation, we may only need a smaller number of singular values for a less complex image. In other words, if we specify a certain rank of decomposition l, with $l < N$, a higher approximation quality can be obtained for a less complex image compared to a more complex image. Instead of performing the SVD on the whole image, the same approach can be applied to patches of the image. For decomposition with rank l, different patches may have different approximation quality due to the fact that the complexity of patch is different from each other. In the case of a cine cardiac MR sequence, higher approximation ratios correspond to patches with no or little motion and lower approximation ratios corresponding to patches with larger motion, such as mitral valve point and apex (see Figure 5.3).



(a) **(b)** **(c)**

Figure 5.3 – This figure shows (a) self similarity maps at different locations, including valve points, apex, blood pool and a random point at the myocardium boarder. And a cine image overlaid with SVD image of approximation ratio set to (b)<= 80% (c) <= 70%.

Using the sum of the entropy and SVD approximation ratio, we select a set of sparse landmarks with the lowest values. A certain space is set between landmarks and edge conditions are checked in order to ensure that the selected landmarks can fully cover the endocardium and

epicardium.

5.2.2 Coupled tracking of multiple landmarks

Given a set of distinctive landmarks located in the first frame, we aim to find their corresponding positions in the next and all subsequent frames. A Markov Random Field (MRF) can be used to model this problem to establish the dependencies between landmarks that are nearby in the image. Markov Random Field (MRF) [Kindermann et al., 1980] [Blake et al., 2011] was generalized from Markov processes to replace a time index by a space index and has been applied to a wide variety of tasks of computer vision, medical image analysis and pattern recognition etc.. Applying MRFs for vision analysis is to build a joint probabilistic model over the pixel values and hidden variables as labels associated with the pixels. In this work, We convert the landmark tracking into a discrete labelling problem and formulate it in a MRF framework to find the optimal joint label assignment.

Labelling

For a given landmark $\mathbf{x}$, we expect its position in the next frame to be within the neighbourhood of its current position. So we assign label $l_{\mathbf{x}}$ to associate with displacement from the pixel site p. The maximum spatial displacement is set to $\pm 10mm$ along the radial direction and $\pm 7mm$ toward its neighbouring landmarks along the endocardial boundary, based on the statistical analysis from the results in Chapter 4. We will discuss these two directions later, which are used to guide the tracking. Thus the label set for each landmark is $l_{\mathbf{x}} = \{1, 2, ...21\}$ in the direction towards the LV centre, corresponding to the displacement from $-10mm$ to $10mm$ with $1mm$ as a step (including zero displacement) and $l_{\mathbf{x}} = \{1, 2, ..15\}$ for circumferential direction. Figure 5.4 (b) and (c) illustrate the labelling and the displacement along the two directions.

Cost Function

We then formulate the multiple landmark tracking in LA views in a MRF framework in which we minimise the following energy function:

$$E = \sum_{i \in \chi} V_i(\mathbf{x}_i) + \sum_{(i,j) \in \chi} \omega_{ij} V_{ij}(\mathbf{x}_i, \mathbf{x}_j) \tag{5.2}$$

Here χ is the landmark set. $V_i(\cdot)$ (unary potential) encodes the data term of the energy. The pairwise potential $V_{ij}(\cdot)$ represents the smoothness term.

Data Term

For the landmark tracking, the unary potential is defined by the weighted SSD similarity metric between patches centered at the landmark positions to capture the geometric and intensity similarity between the landmark under study and its candidate matching point.

$$V_i(\mathbf{x}_i) = \sum_{\Omega_{\mathbf{x}}, \Omega_{\mathbf{y}}} w_i (\mathbf{x}_i - \mathbf{y}_i)^2 \tag{5.3}$$

Here $\Omega_{\mathbf{x}}$ and Ω_y are the local patches of size 11×11 around point $\mathbf{x}_i$ and its candidate matching point $\mathbf{y}_i$ respectively. Because the left ventricle undergoes a number of deformations including thickening of the myocardium during the cardiac cycle, the appearance of the patch surrounding a landmark may change significantly. Thus, we modify the similarity metric to include a weight w_i that is (a) zero outside the epicardium and (b) to increase the influence of the blood pool near the endocardium. The spatial weighting used in the patch-based SSD is defined as follows:

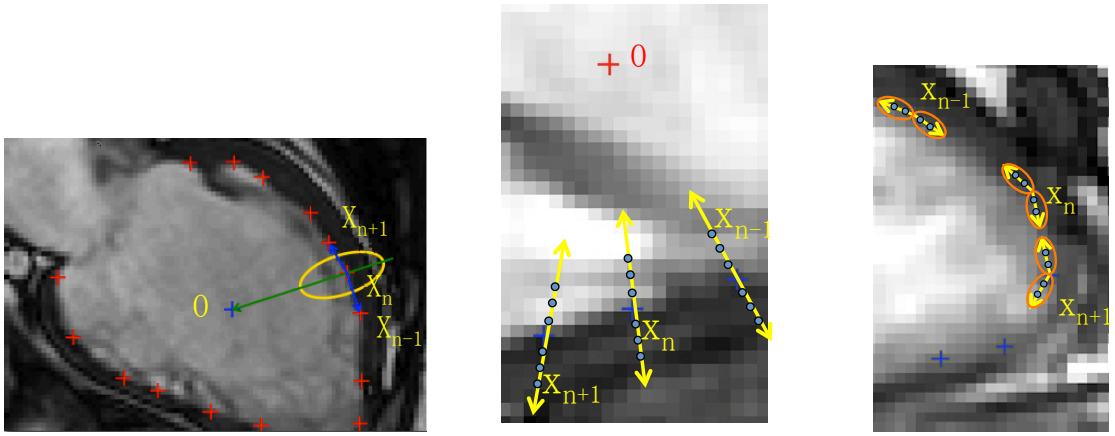
$$w_i = \begin{cases} c_1 & \text{if } L(\mathbf{x}_i) = L_{myo} \\ c_2 & \text{if } L(\mathbf{x}_i) = L_{blood} \\ 0 & \text{otherwise} \end{cases} \tag{5.4}$$

Here c_1 is set to 1 and c_2 is set to 2.

Smoothness Term

The pairwise potential $V_{ij}(\cdot)$ in the energy function models the interaction between neighboring landmarks to enforce smoothness in our application. It is defined by the Euclidean distance of the displacements $D(x)$ of a pair of neighboring points. This term imposes the assumption that other landmarks in the vicinity of a given landmark should undergo a deformation that is similar to this landmark; in other words, the intuition of this term is to maintain a coherent motion between points close to each other. This term is defined as follows:

$$V_{ij}(\mathbf{x}_i, \mathbf{x}_j) = |D(\mathbf{x}_i) - D(\mathbf{x}_j)| \tag{5.5}$$



(a) **(b)** **(c)**

Figure 5.4 – a) shows that the searching region for tracking a point $\mathbf{x}_n$ is inside a yellow oval, the model is further simplified into two stages of searching: b) towards/away from centre O and c) towards/away from neighbouring points $\mathbf{x}_{n+1}$ and $\mathbf{x}_{n-1}$.

Optimization and implementation

We initially implemented the tracking strategy for landmark tracking in the LA view. We further extend the strategy and apply it in SA views, which will be discussed in the next section. During the cardiac cycle, the LA images provide useful information about the radial and longitudinal deformation of the left ventricle. We expect the location of a landmark in the following frame to be within its neighbourhood as shown in Figure 5.4.

To optimise the search procedure, the tracking is conducted in two stages to reduce the search space. At the first stage of tracking, the main deformations, i.e. the radial and longitudinal motion of the myocardium, are best captured along the direction towards the centre. The centre of the left ventricle is defined as the intersection point of the middle slice of the short axis (SA) image and two LA images. At the second stage of tracking, we track points along the direction towards their neighbouring points in order to take into account of the circumferential twist and the motion along the border. The penalty increases when two neighbouring points move towards or away from each other and decreases when the two points move in the same direction. The tracking is conducted frame by frame, using the position of the point at the first frame as the starting position for the current frame to avoid accumulated error. Values of the search range at each stage are empirical.

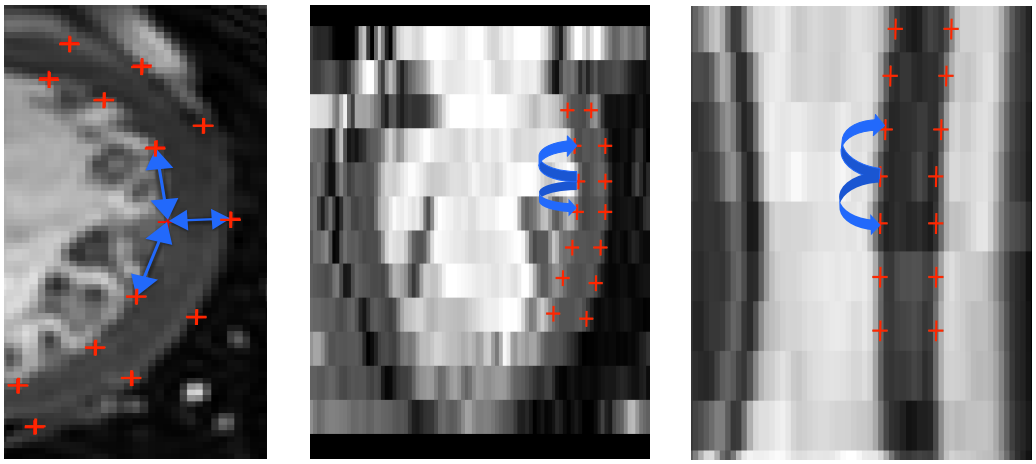
At both stages, Fast-PD, a graph cut based algorithm is used for optimization of the MRF [Komodakis and Tziritas, 2007, Komodakis et al., 2008]. This discrete optimization approach utilizes primal-dual schema to provide an efficient approach in multi-label MRF optimization [Komodakis and Tziritas, 2007]. Fast-PD has been used in image registration, stereo disparity estimation and optical flow estimation.

5.2.3 4D sparse motion tracking

We further extend our approach to model the whole sequence of the images by a 4D MRF. As the neighbourhood of nodes are defined not only within the same slice but also across slices and frames, so the motion tracking can be performed no longer frame by frame but on the whole 4D image.

The neighbourhood of each node in slice k and frame t includes not only the neighbouring nodes in the same slice (Figure 5.5a), but also those in slices $k+1$ and $k-1$ (Figure 5.5b). In addition, the neighbourhood also includes temporal neighbours, i.e. the corresponding voxels at frame $t+1$ and $t-1$ respectively (Figure 5.5c). We refer to these neighbouring relationship as endo-endo, epi-epi, endo-epi, slice-slice and frame-frame edges. In our implementation, a total of seven neighbours are used for each landmark with the exceptions of landmarks in

the first or last slice (or frame) of the short-axis image stack. Similar to the 2D+t tracking



(a) (b) (c)

Figure 5.5 – The neighborhood of a landmark at the endocardium in slice k and frame t which includes a) two neighbouring endocardial nodes and one epicardial node in the same slice, b) one neighboring node in each of the slice $k+1$ and $k-1$, and c) the corresponding voxels at frame $t+1$ and $t-1$ respectively.

method proposed in section 5.2.2, we associate each label of a point with the displacement from its original position and formulate the multiple landmark tracking in a multi-label MRF framework in Equation 5.2.

The unary potential $V_i(\cdot)$ is defined as before in Equation 5.3. The weight for endocardial landmarks is also defined as before in Equation 5.4. For epicardial landmarks, the weighting function w_i is slightly modified to ignore the influence of wall thickening, the definition for w_i is as follows:

$$w_i = \begin{cases} c & \text{if } L(\mathbf{x}_i) = L_{myo} \\ 0 & \text{otherwise} \end{cases} \tag{5.6}$$

Here c is set to 1.

For the pairwise smoothness term $V_{ij}(\cdot)$ of the energy function, as more neighboring relationships are modelled in this MRF framework, w_{ij} is introduced to weight the magnitude of the constraints between neighbouring points in Equation 5.5. Hence the smoothness term is defined as follows:

$$V_{ij}(\mathbf{x}_i, \mathbf{x}_j) = w_{ij}|D(\mathbf{x}_i) - D(\mathbf{x}_j)| \tag{5.7}$$

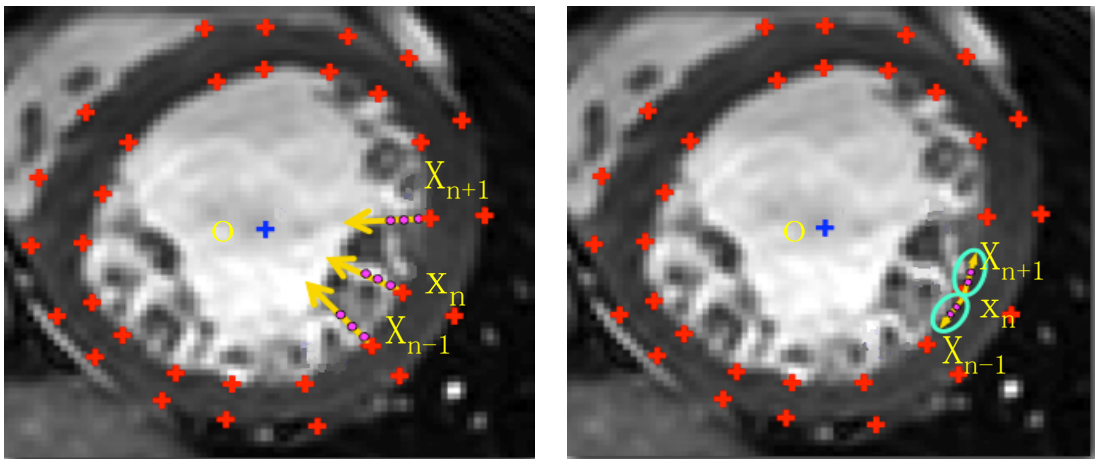


Figure 5.6 – The tracking of a landmark P_n in SA view is modelled by a two-stage searching: a) towards centre O and b) towards/away from neighbouring points $\mathbf{x}_{n+1}$ and $\mathbf{x}_{n-1}$.

w_{ij} varies according to the type of edges. For instance, the weight for the endo-endo edge is larger than that for the endo-epi edge because the motion of a landmark at endocardium correlates much stronger with that of its endocardial neighbours than that of its epidardial neighbours. All these weights are set empirically. As illustrated in Figure 5.6, the tracking is also performed in two stages as in the last section 5.2.2 to account for different types of deformation: towards/away from center and towards/away from neighbouring landmarks. At the first stage of tracking, as the displacements of the landmark in other frames are all defined relative to the ED phase, so the search range is set from $-2mm$ to $20mm$ to account for the contraction of the heart.

5.2.4 Estimation of LV dyssynchrony

Dyssynchronous myocardial contraction is believed to decrease cardiac function, which occurs in myocardial diseases such as dilated cardiomyopathy, ischemic disease and so on. The dyssynchrony of the regional deformation can be represented by systolic dyssynchrony index (SDI). The American Heart Association (AHA) model [Cerqueira et al., 2002] (see Figure 3.1) of the left ventricle is commonly used for regional cardiac functional analysis which divides the LV into 17 segments. The SDI is defined as the standard deviation of the peak phases for these 17 segments as follows:

$$\text{SDI} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(f_i - \bar{f})^2} \tag{5.8}$$

where f_i is the peak phase that segment i reach its minimum or maximum regional function and N is 17. The regional function may refer to volume, strain or myocardial wall thickness. The SDI shows how much variation from the average regional function exists and a high index value indicates more dyssynchrony [Nesser et al., 2007] or worse cardiac global functionality [Sugihara et al., 2005]. SDI has been previously reported to be a good indicator for selecting patients who respond to CRT [Bax et al., 2004].

To compute the myocardial wall thickness SDI, we first define the 16 segments of the left ventricular myocardium according to the AHA model [Cerqueira et al., 2002] for each subject; the apex segment is not defined in this work. This is performed by fitting a pre-constructed myocardium model with 16 segments to the segmented myocardium using non-rigid image registration [Rueckert et al., 1999]. After registration, a 16 segment division of the myocardium at the ED phase is obtained for each subject.

In the proposed landmark detecting and tracking approach, the landmarks are automatically selected in pairs on both the endocardium and epicardium along the radial line towards the LV centre. Hence the Euclidean distance between a pair is regarded as the wall thickness at that position. In each time frame t, the wall thickness WT_i^t of the ith segment S_i is obtained by averaging the distances between all pairs of landmarks located in this segment:

$$WT_i^t = \text{mean}(\|\mathbf{x}_{endo}(k) - \mathbf{x}_{epi}(k)\|), \{\mathbf{x}_{endo}(k), \mathbf{x}_{epi}(k)\} \in S_i, i = 1..16$$

Here $\{\mathbf{x}_{endo}(k), \mathbf{x}_{epi}(k)\}$ is the pair of landmarks on the endocardium and epicardium in segment S_i. Thereby, we can find out the peak phase f_i at which the maximum wall thickness is reached for segment S_i and then compute the SDI using Equation 5.8.

5.3 Evaluation and results

We have acquired 51 sequences from 48 CRT patients and three normal volunteers using a 1.5T MR-scanner (Achieva, Philips Healthcare) with a 32-element cardiac coil or a 5-element cardiac coil. Vector electrocardiography (VECG) is used to synchronize the data acquisition with the cardiac cycle. Standard views including the SA, LA views are obtained using the following parameters: flip angle = 60^o, TR/TE = 2.9/1.5ms, resolution 1.45 x 1.45 x 10mm, 30 heart phases. Among them, single slices are obtained for all LA views and multiple slice steady-state free precession (SSFP) sequences are acquired for the SA view.

5.3.1 Tracking results for the LA view sequences

In order to evaluate the 2D+t sparse landmark motion tracking algorithm proposed in section 5.2.2, we applied it on the LA view sequences. For comparison, five landmarks are manually annotated in the ED phase; their corresponding positions are also annotated in the end of the systolic (ES). Figure 5.7 shows an example of the landmarks annotated in different frames.

Using the landmarks tracked by the sparse landmark motion tracking (SLMT), an non-rigid transformation can be defined between ED and ES phase, which can be used to estimate the positions of the five manually annotated landmarks in the ES phase. We also predicted the position of the landmarks using dense motion tracking (DMT) (using non-rigid registration [Rueckert et al., 1999]) with and without being initialised by the result from sparse landmark motion tracking. The tracking errors in terms of root mean square (RMS) error of the manually and automatically tracked landmarks using the three approaches are shown in Figure 5.8(a) and Table 5.1. In order to illustrate the case-wised improvement by initializing the DMT with SLMT, the improvements of landmark tracking error in percentage are also shown in Figure 5.8(b). A substantial improvement of tracking accuracy is obtained for each of the 51 cases.

It can be seen that both the SLMT and the DMT initialised by the results of SLMT outperformed the DMT without initialisation substantially with improvement of 20.8% and

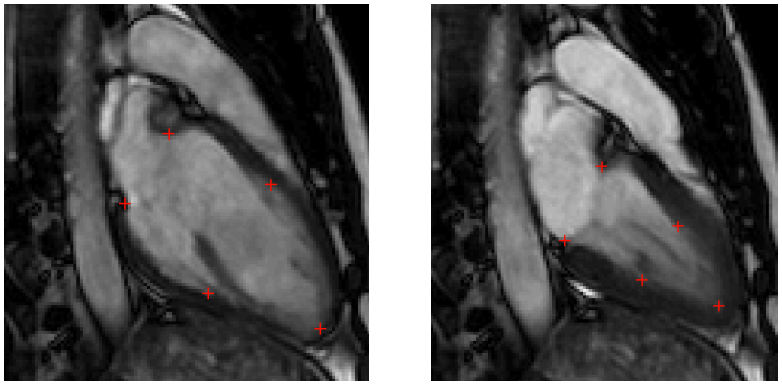


Figure 5.7 – This figures shows five landmarks and their corresponding positions are annotated manually in ED and ES phase.

Table 5.1 – Landmark motion tracking error in LA view

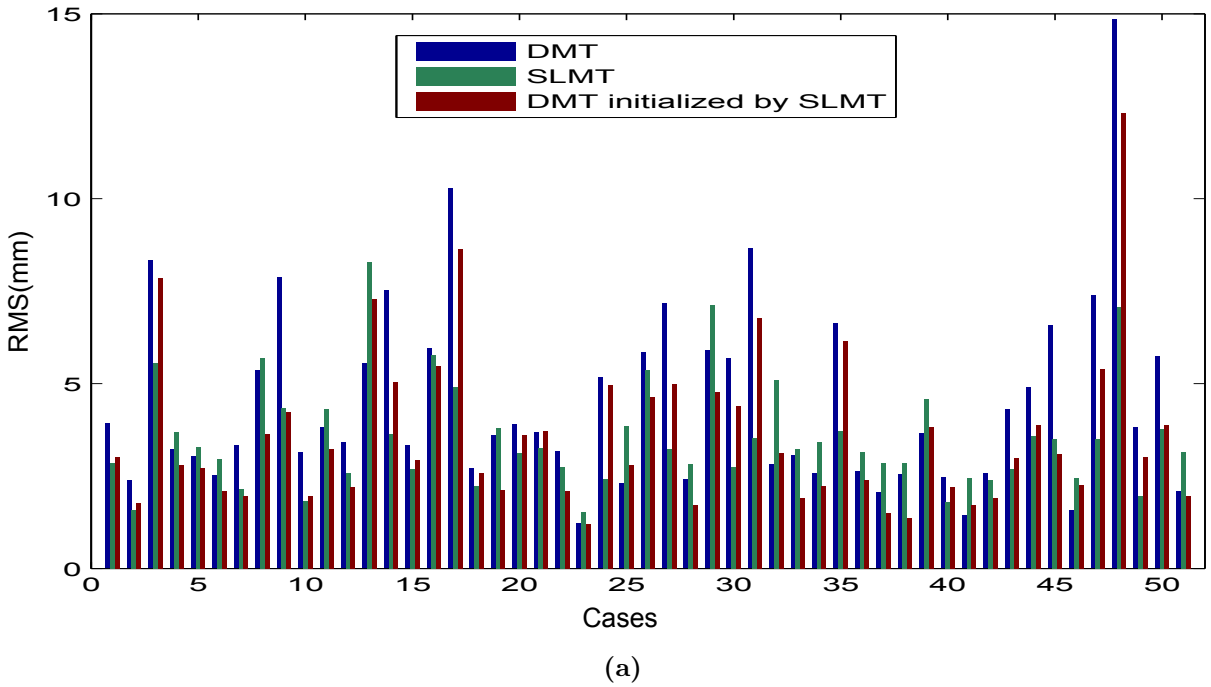
	DMT	SLMT	DMT initialized by SLMT
RMS	$4.48 \pm 2.11mm$	$3.55 \pm 1.92mm$	$3.61 \pm 1.86mm$
Improvement compared to DMT	-	20.8%	19.4%

19.4% respectively. Figure 5.9 shows examples of tracking results of five landmarks using DMT without initialisation, SLMT and DMT initialised by the results from SLMT.

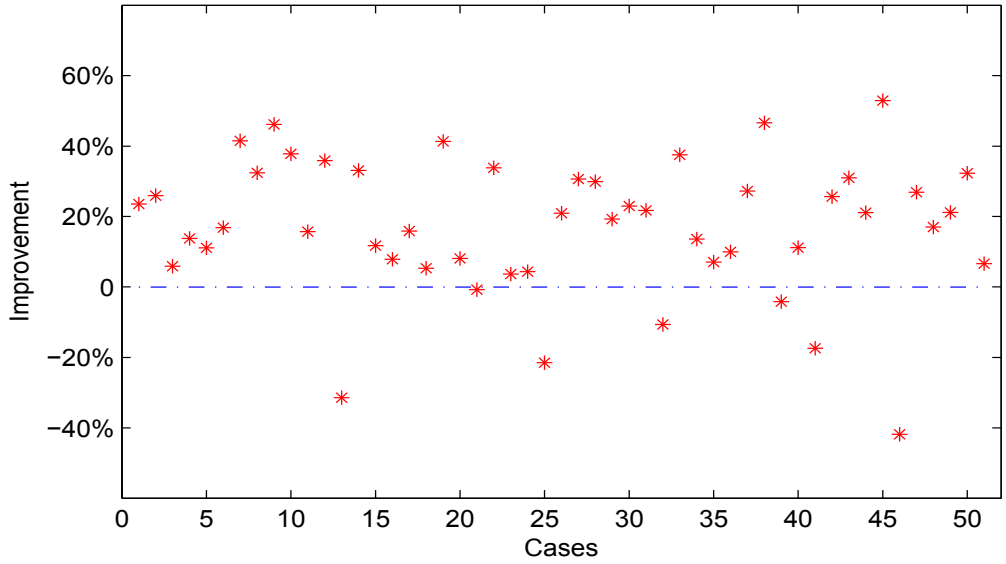
5.3.2 Tracking results for the SA view sequences

We evaluate the extended 4D tracking approach proposed in section 5.2.3 using the SA view sequences. Furthermore, the tracking results are used to estimate the LV contraction dyssynchrony, which are compared with clinical measurement. 47 subjects, including 44 CRT patients and three normal volunteers are used in this evaluation because that the clinical measurements are not available for the other four subjects.

As before in section 5.3.1, five landmarks are manually annotated on the endocardium in the middle slice in the ED phase. In addition their corresponding positions are annotated in ES phase. The positions of these five landmarks in the ES phase are also automatically estimated by the transformation defined by the 4D sparse landmark motion tracking (SLMT) proposed in section 5.2.3. Thereafter, the accuracy of the motion tracking is computed as the distance



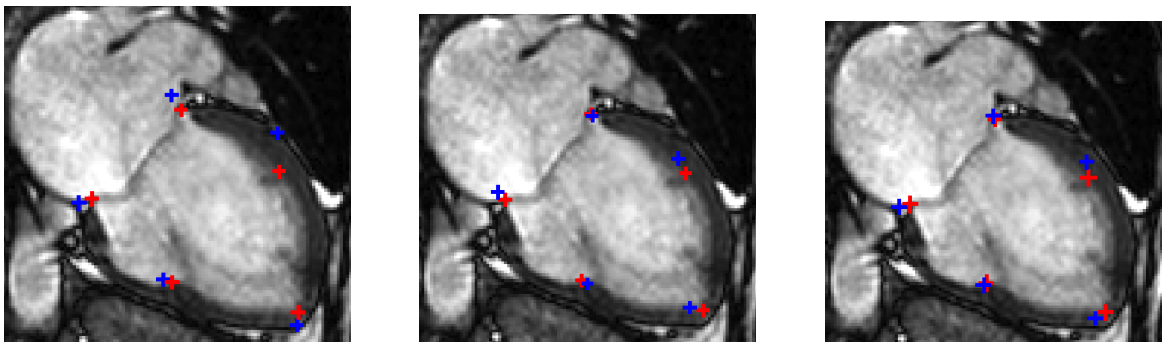
(a)



(b)

Figure 5.8 – This figure shows the landmark tracking errors in LA view for 51 cases when compared with the results of manual tracking. (a) The results using the DMT, the SLMT and DMT initialised by the result from SLMT. (b)The improvement of landmark tracking error in percentage when comparing results from DMT with and without initialisation.

between the tracked position of the landmarks and their corresponding manually annotated position in ES. For comparison we also tracked the landmarks using dense motion tracking (DMT) (using non-rigid registration [Rueckert et al., 1999]) with and without being initialised



(a) (b) (c)

Figure 5.9 – This figure compares the automatic tracking results of five landmarks (in blue) with manually annotated positions (in red) using a)DMT without initialisation, b)SLMT and c)DMT initialised by SLMT.

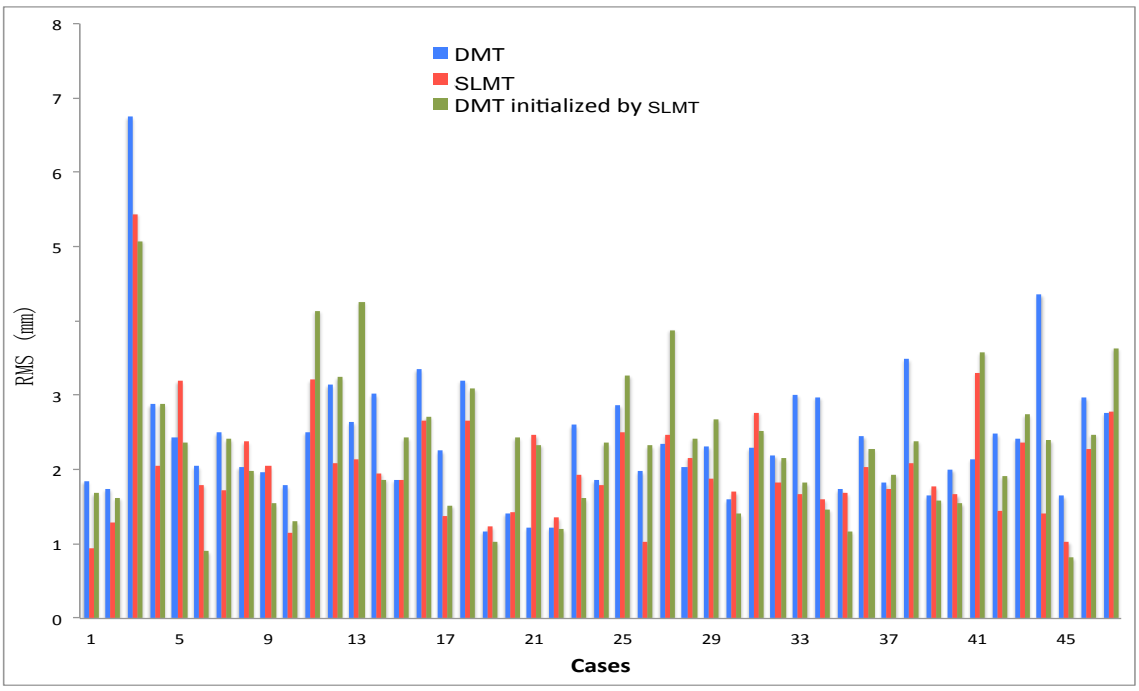
Table 5.2 – Landmark Motion Tracking Error in SA views

	DMT	SLMT	DMT initialized by SLMT
RMS	$2.41 \pm 1.22mm$	$2.03 \pm 1.05mm$	$2.31 \pm 1.22mm$
Improvement compared to DMT	-	15.7%	4.2%

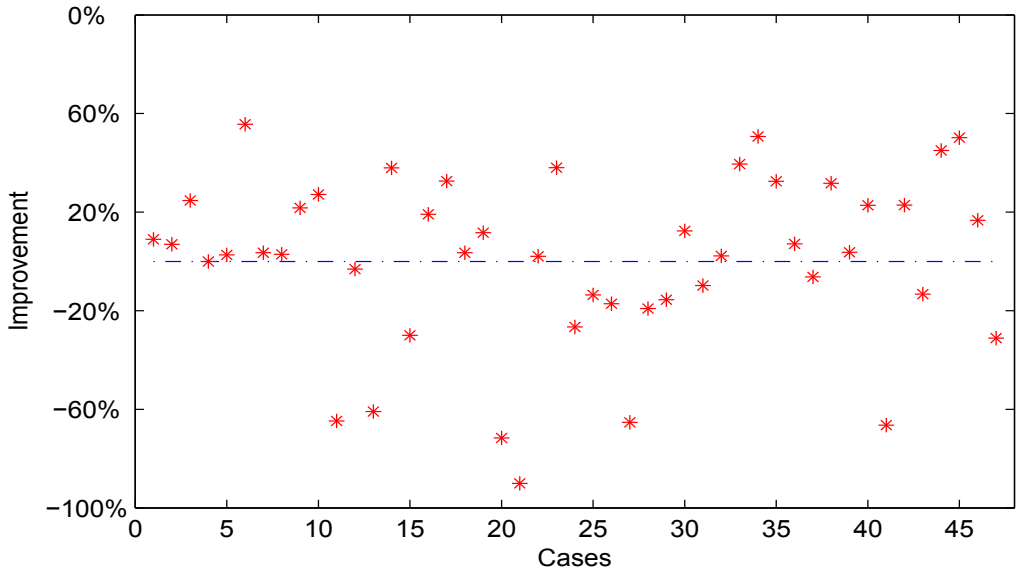
by the results from sparse motion tracking (SLMT). The average tracking errors in terms of root mean square (RMS) using the two approaches are shown in Table 5.2. It can be seen that both SLMT and DMT using SLMT as initialisation outperformed the DMT without initialisation. Figure 5.10 shows the case-by-case tracking errors of the three methods.

5.3.3 Analysis of LV dyssynchrony

As introduced in section 5.2.4, endocardial-epicardial landmark pairs on 16 segments of the LV are used to calculate SDI. For each of the 16 segments of the left ventricular myocardium, there are around two to eight endocardial-epicardial pairs of landmarks in our experiment. Figure 5.11 shows the change of the myocardial wall thickness of each of the 16 segments throughout the whole cardiac cycle from a normal volunteer and two CRT patients respectively. Each of the 16 curves in the figure shows the change in thickness for one of the 16 segments. As shown in the figure, the wall thickening for the normal volunteer is more synchronous across the segments, with almost all curves reaching their maximum thickness at the same timepoint (approximately at 35% in the cycle). By contrast, the curves for the two CRT patient indicate



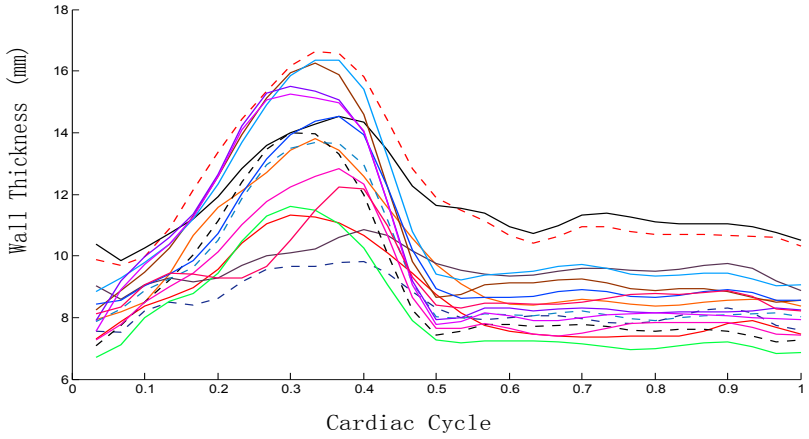
(a)



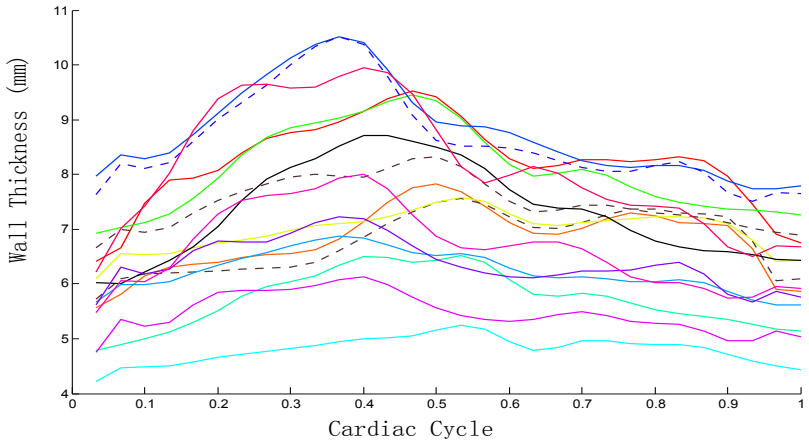
(b)

Figure 5.10 – This figure shows the landmark tracking errors (mm) in SA view (a) using the DMT, the SLMT and DMT initialised by the result from SLMT. (b)The improvement of landmark tracking error in percentage when comparing results from DMT with and without initialisation.

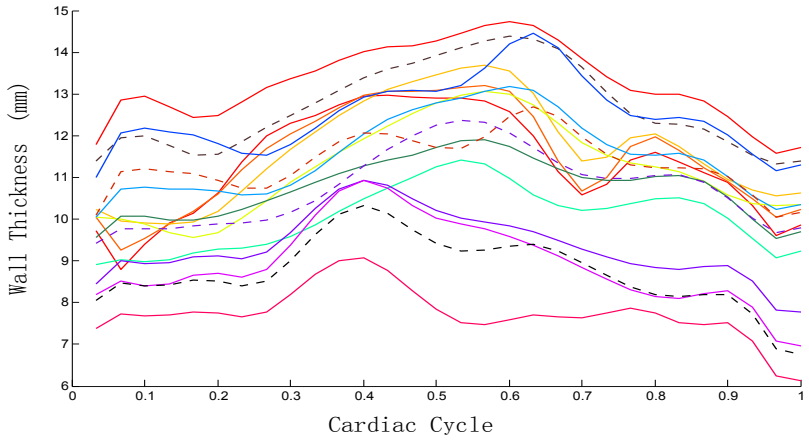
that the wall thickening for the two patients is not synchronous across the segments. In addition, the deformation of myocardium for the normal volunteers is much larger compared to that of the patients. The regional wall thickness SDI for each of the 47 subjects can be computed from the deviation of the wall thickening for each of the segments. For comparison, we use a



(a)



(b)



(c)

Figure 5.11 – This figure shows the wall thickness changing (in *mm*) curves of 16 AHA segments of heart from (a) a normal subject and (b) (c) CRT candidates. Each of the 16 curves in the figure denotes the thickness change of one of the 16 segments. The wall thickening for the normal volunteer is more synchronous across the segments, with almost all curves reach their own maximum thickness at the same timepoint (approximately at 35% in the cycle). By contrast, the curves in the two CRT patient figures moves more diversely indicating that the wall thickening for the two patients are not synchronous across the segments.

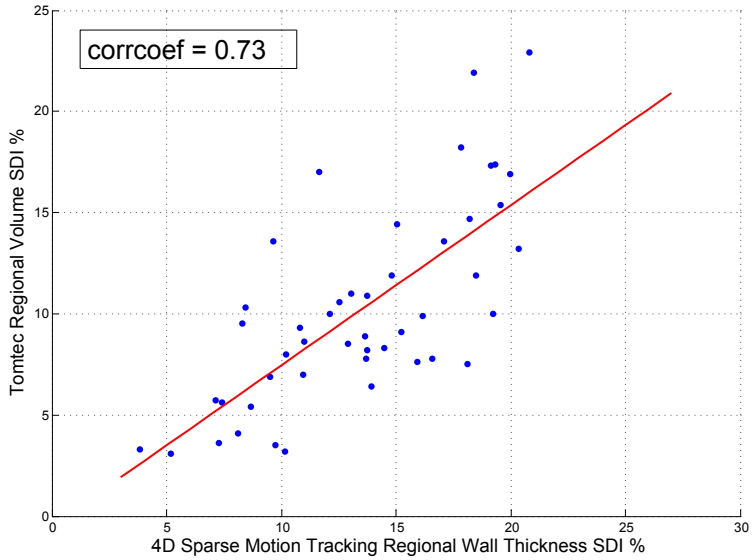


Figure 5.12 – Evaluation of wall-thickness SDI against the Tomtec's regional volume SDI. The Pearson correlation coefficient between them is 0.73.

commercial software tool (TomTec 4D LV analysis tool V2.0 [Kühl et al., 2004]) which relies on manual tracking within tri-plane projections and semi-automated border detection, to obtain 16 segments regional volume SDI for each of all the 47 cases. The volume SDI is calculated based on the peak phase in which the maximum volume is reached. The Pearson correlation coefficient between the proposed wall thickness SDI and the regional volume SDI obtained by the TomTec software, is 0.73. The scatter plot of the two SDI indices, as shown in Figure 5.12, illustrates a strong correlation between them.

5.4 Conclusions

In this chapter we have presented an approach to identify and track a sparse set of landmarks throughout the cardiac cycle. An integrated framework is proposed to combine entropy and SVD-based patch analysis for landmark detection with a motion tracking method. The accuracy of the proposed sparse motion tracking is evaluated by tracking a group of manually annotated landmarks on the endocardium of the left ventricle (LV). We also show the derived regional wall thickness SDI agrees well with the clinical measurements of regional volume SDI.

Comparing the motion tracking results of LV in LA view and SA views listed in Table 5.1

and Table 5.2, the tracking errors in LA view are much larger than that in SA views. We speculate that this is due to fact that the left ventricle demonstrates much larger deformation in longitudinal direction in the LA view than in the SA view. Meanwhile, significant improvements of the tracking accuracy are obtained by the proposed SLMT approach in LA view and less so in the SA views. This may be caused by the larger deformation in the LA views as well. Comparing to DMT method, the proposed SLMT approach benefits from the larger search space of the MRF during the optimization.

We also analyzed of LV dyssynchrony based on the proposed sparse landmark motion tracking results and the AHA LV model [Cerqueira et al., 2002]. Comparing with the analysis generated by a state-of-the-art commercial tool, the results obtained by the proposed method are in good agreement. This shows that it is plausible to use the proposed sparse landmark motion tracking to estimate the dyssynchrony index of the regional deformation.

Chapter 6

Cardiac motion analysis using manifold learning

The work in this chapter is based on the following paper:

- Haiyan Wang, Wenzhe Shi, Wenjia Bai, Antonio M. Simoes Monteiro de Marvao, Timothy J. W. Dawes, Declan P. O'Regan, Philip Edwards, Stuart Cook and Daniel Rueckert. (2015). Prediction of clinical information from cardiac MRI using manifold learning. Functional Imaging and Modeling of the Heart (FIMH), Volume 9126, 2015, pages 91-98.

6.1 Introduction

Cardiac MR images contain rich information that can be used to reveal important information about the anatomy and function of the heart. Given a large number of subjects, the shape, appearance and motion of the heart can vary significantly across subjects due to age, gender, physiological parameters such as blood pressure or disease progression. Finding an appropriate representation of a large dataset of cardiac MRI images in order to extract meaningful features has gained increasing attention and importance in recent years. The most popular approaches include atlasing techniques [Duchateau et al., 2011, Hoogendoorn, 2014, Lombaert et al.,

2012, Fonseca et al., 2011]. However, these atlasing techniques require image segmentation and motion estimation in order to extract the desired information. The process is usually computationally intensive and depends heavily on the accuracy and robustness of the image analysis techniques.

Dimensionality reduction techniques such as manifold learning simplify the data so that it can be efficiently processed and interpreted. Manifold learning has seen increasing use in the analysis of medical images ranging from the segmentation of brain MR images [Wolz et al., 2010], biomarker discovery in knee MR images [Donoghue et al., 2011], the detection of respiratory motion in ultrasound [Wachinger et al., 2010], to study of regional appearance pattern in cardiac and brain MRI images [Bhatia et al., 2014] and the classification of abnormal myocardial motion [Duchateau et al., 2012].

The importance of obesity indices, such as body mass index (BMI) and body fat percentage (Fat%), as risk factors for diabetes and hypertension is well-recognized, but its role for prediction of cardiac function and as a CVD risk factor in healthy individuals is less well established [Abbasi et al., 2002]. Recently, Stelfox [Stelfox et al., 2006] found that BMI is positively correlated with cardiac output (CO) and stroke volume (SV) on healthy volunteers. In [Abbasi et al., 2002] BMI is also found to be a powerful predictor of CHD risk based on the study of 314 health volunteers. AlJaroudi et al. [Aljaroudi et al., 2012] concluded that BMI is independently associated with worsening diastolic dysfunction (DD). Defining the relationship between obesity indices and cardiac images may be able to indicate whether the impact of the body constitution on appearance and motion of the heart can be represented by obesity indices.

In this work, we propose to learn anatomical and functional information directly from cardiac MR images using manifold learning. We hypothesize that the anatomical information can be embedded in the differences between images of the same phase in the cardiac cycle across different subjects and that the functional information can be embedded in the differences between different cardiac phases of the same subject. To exploit both anatomical and functional information from cardiac MR images, we first learn a low dimensional manifold of a large number of cardiac MR images at a single time point to describe inter-subject variation. Then

we consider two alternative approaches to incorporate functional or motion information into the manifold. Both approaches are compared to the manifold learned from a single time point. Finally, we demonstrate that by using different combinations of information, we are able to predict different clinical variables directly from the learnt manifold representation with good accuracy.

The remainder of the chapter is organised as follows: Section 6.2 explains the methods used to derive different forms of manifolds for incorporating motion information. The dataset used in this chapter is also described. Section 6.3 presents the results of using the derived manifold coordinates to discriminate according to gender and between subjects with high blood pressure and those with normal blood pressure, and to regress age and obesity indices (BMI, FFMI and Fat%). Finally, Section 6.4 concludes the work.

6.2 Method and materials

This chapter proposes to learn anatomical and functional information directly from cardiac MR images using manifold learning, and obtain classifiers and regression models for cardiac functions.

6.2.1 Materials

The dataset used in this chapter consists of 209 randomly selected subjects, sampled from the UK 1000 Cardiac Phenomes project [de Marvao et al., 2014]. Subjects were excluded at screening if they had known cardiovascular diseases or were being treated for hypertension, diabetes or hypercholesterolemia. Additionally, female participants were not included if they were pregnant or breastfeeding, but were eligible if they took oral contraceptives. An overview of the subjects is given in Table 6.1. In the table, BMI refers to body mass index, which evaluates body mass relative to the height; FFMI (fat free mass index) expresses the fat free component of body mass relative to the body size, and Fat% is body fat percentage of the body

Table 6.1 – Characteristics of the subjects whose images were used in this chapter. *Data available for 200 subjects

Number of subjects	209
Gender M/F	95/114
Blood-pressure High/Normal	35/174
LV volume (ED)	$145.35 \pm 41.98[80 \sim 256]$
LV volume (ES)	$52.55 \pm 20.50[25 \sim 127]$
Stroke volume	$92.85 \pm 24.93[54 \sim 152]$
LV mass	$108.14 \pm 35.81[60 \sim 204]$
Age	40.79 ± 12.65 $[18 \sim 73]$
BMI	24.93 ± 3.84 $[18.02 \sim 42.30]$
FFMI*	18.14 ± 2.35 $[11.86 \sim 25.89]$
Fat%*	36.38 ± 9.43 $[3.77 \sim 56.88]$

mass. All these three indices are measures that can be used to define obesity [WHO Expert Consultation, 2004] [WHO, 1995] [Schutz et al., 2002], hence we will refer to them collectively as obesity indices. More detailed information about some of the characteristics in the table, including LV volume, stroke volume, LV mass, BMI, FFMI and Fat% are given in section 3.1.

Images were obtained on a 1.5T Philips Achieva system (Best, Netherlands). The maximum gradient strength is $33mT/m$. A 32 element cardiac phased-array coil is used for signal reception. Scout images were obtained and used to plan a single breath-hold 3D cine balanced steady-state free precession (b-SSFP) image in the left ventricular short axis (LVSA) plane from base to apex using the following parameters: pixel size $= 2.0 \times 2.0mm$; slice thickness $= 2mm$; reconstructed voxel size $= 1.2 \times 1.2 \times 2.0$mm; cardiac phases $= 20$; sensitivity encoding (SENSE) factor. Figure 6.1 shows one example of the images obtained from one subject. In addition, Figure 6.2 illustrates the variabilities between subjects.

6.2.2 Image pre-processing

For a population of cardiac MR sequences, a region of interest (ROI) is defined by five manually placed landmarks in the ED phase of the sequence. These five landmarks are carefully placed by clinicians and include the two insertion points of the right ventricle (RV), the apex of the LV and the two longitudinal LV axis points (one in the most basal slice and one in the most apical slice), as shown in Figure 6.3. Then, a ROI is generously chosen around these five landmarks

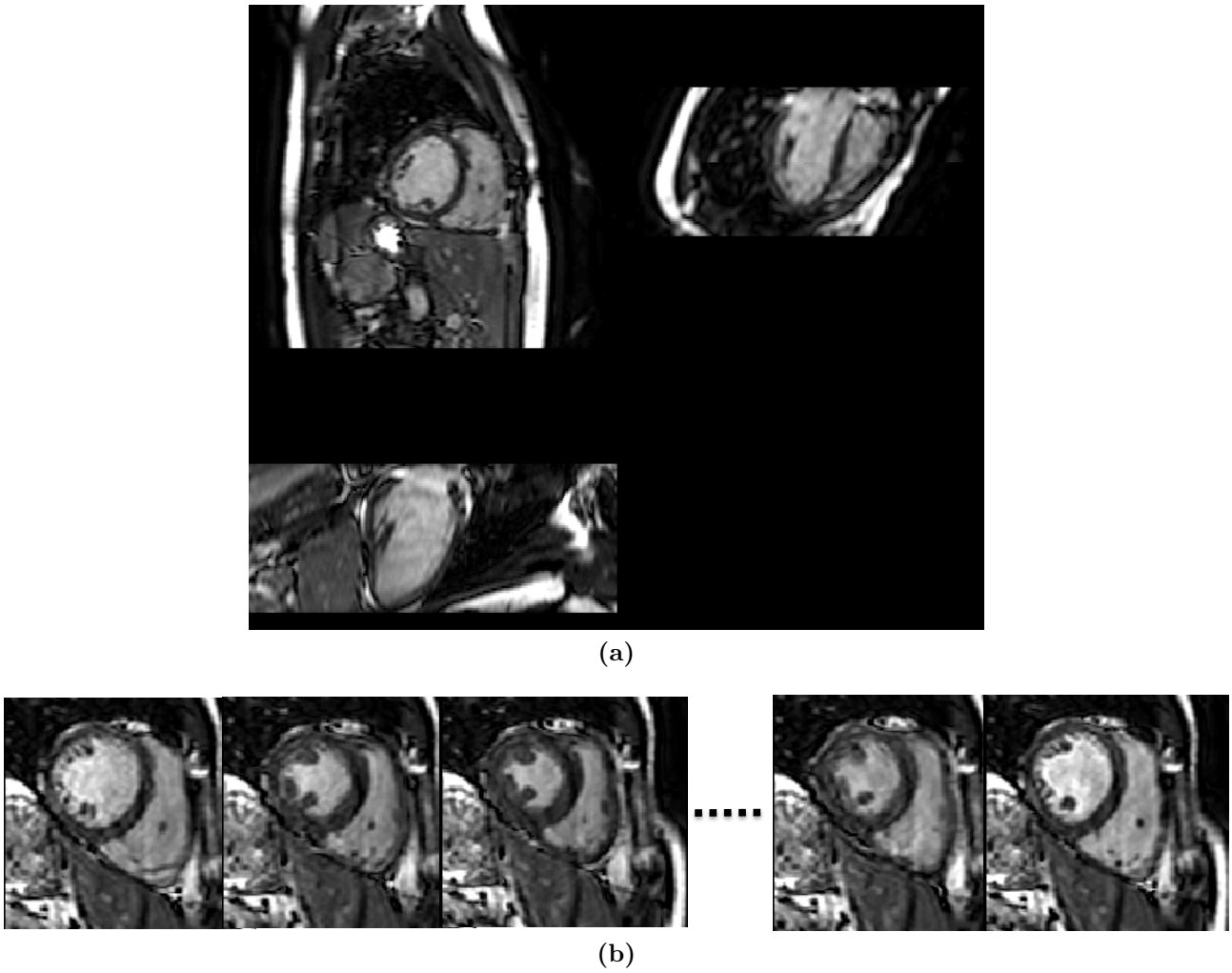


Figure 6.1 – Examples of cardiac MR image from the UK1000 study. (a) The SA view and two LA views of one example images obtained from one subject and (b) images of different time phases of one example subject.

to ensure that the left ventricle is well enclosed and we will not lose any useful information during the spatial normalisation performed afterwards. Each image is then transformed by aligning the landmarks to a randomly chosen reference image using rigid registration. This spatial normalisation is performed to correct for variations in position and orientation across the subjects but to preserve variations due to size differences. The pre-processing is kept as simple as possible in order to avoid the loss of any useful information; as well as to save on computational cost because no explicit image segmentation and motion tracking is required.

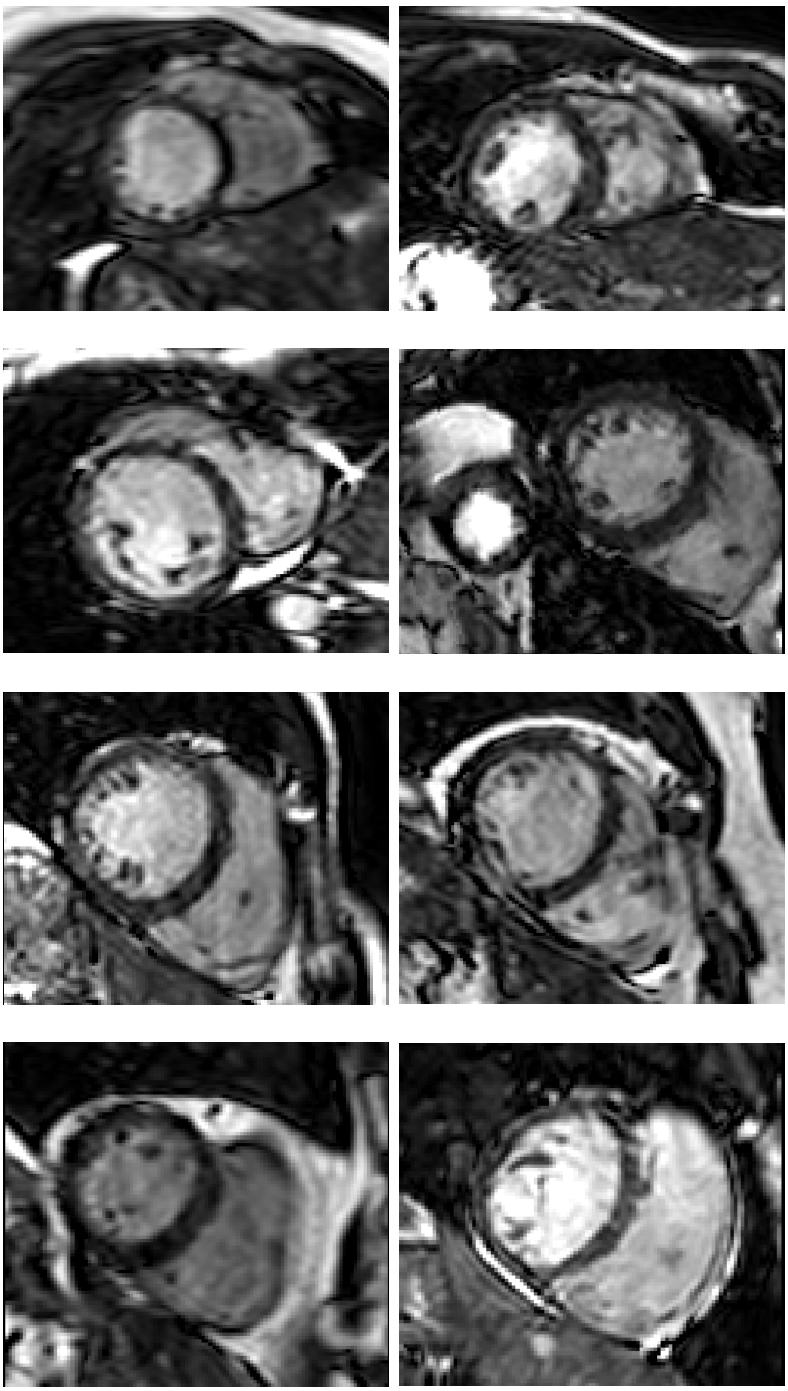
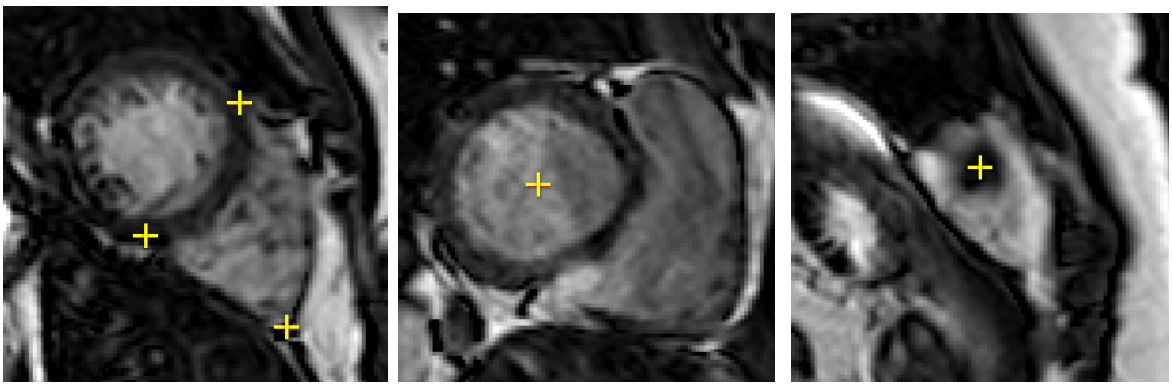


Figure 6.2 – Example images from eight different subjects in the UK1000 dataset.

6.2.3 Manifold learning

Manifold learning technique is applied in this chapter for dimensionality reduction. Manifold learning aims to preserve the underlying local structure to yield a lower dimensional embedding of the data in a high dimensional space. More detailed information about manifold learning can be found in section 2.3.5.



(a) (b)

Figure 6.3 – This figure shows five landmarks in the ED phase which are carefully placed by clinicians: (a) two right ventricle (RV) insert points and apex point in SA view, and (b) the two longitudinal LV axis points (one in the most basal slice and one in the most apical slice).

In general, manifold learning techniques can be divided into local and global methods: Local methods such as local linear embedding (LLE) [Roweis and Saul, 2000] and Laplacian Eigenmaps embedding [Belkin and Niyogi, 2003] aim to preserve local distances between data points. Global methods like Isomap [Tenenbaum et al., 2000] considers the distance between all pairs of data points in its objective function, hence preserves global distances. For our study, we are more interested in preserving the inter-subject and intra-subject variations across the data at the same time. Linear methods such as PCA [Jolliffe, 2002] [Pearson, 1901] and MDS are not considered as a linear transformation is rarely appropriate for dimensionality reduction in the complex dataset. Laplacian Eigenmaps have the advantage of generating manifold coordinates that are not only dependent upon the original pairwise similarity between them, but implicitly emphasize natural clusters in the dataset. This makes the obtained embedding more robust [Belkin and Niyogi, 2003].

In the original high-dimensional space image $I = \{I_1, I_2, ..., I_n\} \in \mathbb{R}^M$ represents a data point of dimension M, where M is the number of voxels in each image. Assuming there exists a lower dimension manifold which can well represent the local geometric relationship between images, the manifold learning seeks to find a low dimensional representation $Y = \{y_1, y_2, ..., y_n\} \in \mathbb{R}^m$ of the input images I where $m \ll M$. Laplacian Eigenmaps ([Belkin and Niyogi, 2003, Von Luxburg, 2007]) build a sparsely connected graph from a pairwise similarity matrix W computed from the data set. The connectivity between each pair of data points in the graph

is defined by a similarity measure S. A schematic overview of manifold learning techniques is given in Figure 6.4.

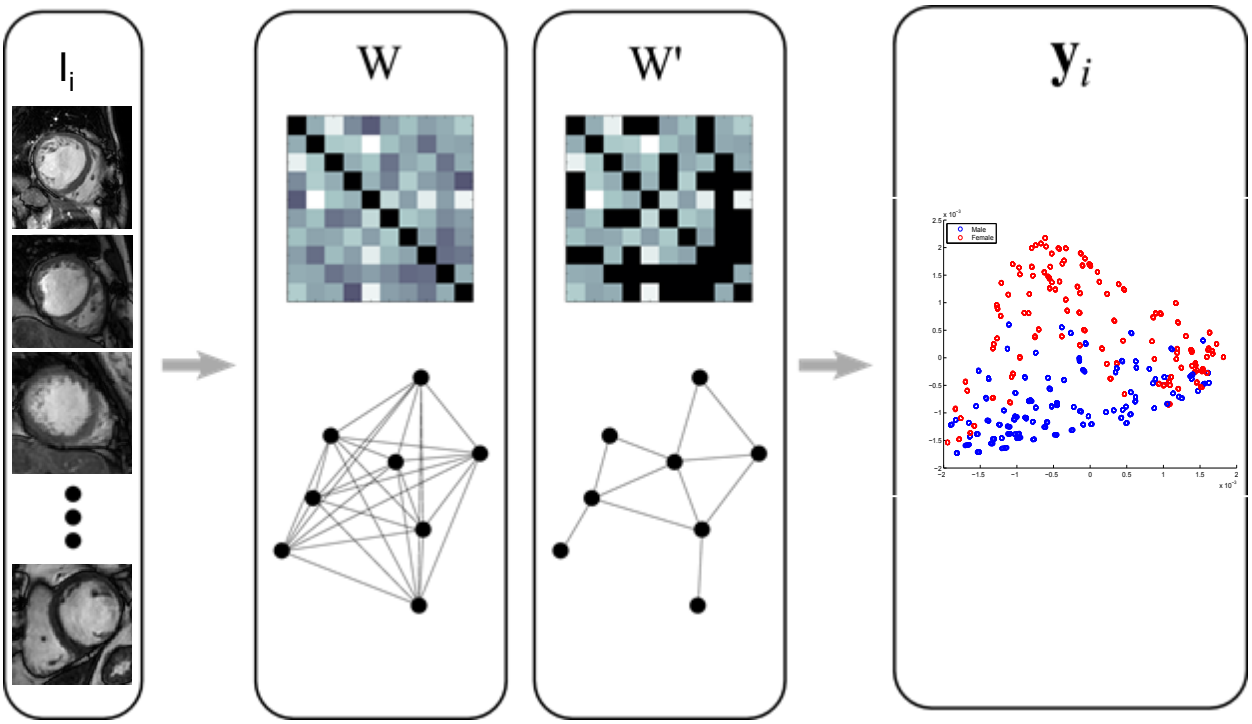


Figure 6.4 – Overview of manifold learning. For input images I_i, either a full graph/matrix W or a sparse one W' is constructed to represent the pairwise similarity or distance. Then a low dimensional manifold y_i, typically using the eigenvalue-eigenvector structure, is derived to represent the original data. Here the first two dimensions of y_i are shown. Adopted from [Aljabar et al., 2012].

For a sparsely connected graph (W' in Figure 6.4), the connectivity weight $w_{i,j}$ is defined as follows:

$$w_{i,j} = \begin{cases} s_{ij} & \text{, if } I_i \in \psi(I_j) \\ 0 & \text{, otherwise} \end{cases} \tag{6.1}$$

Here $\psi(I_j)$ denotes a local neighbourhood of I_j and s_{ij} measures the similarity of images I_i and I_j. For this work, we constructed a sparse k-nearest neighbour graph. We will examine the parameter choice of k in more detail in section 6.3.1.

In manifold learning, it is common to use a similarity measure/metric to define the strength of connection between pairs of data points, i.e. images in our case. The similarity matrix may also be viewed as a graph in which each vertex denotes an image and the weight of each edge corresponds to the similarity or dissimilarity between the image pair it connects. For this

work the similarity metric is defined as the sum of squared difference, SSD, which provides the simplest yet robust form of similarity metric to measure the differences between a pair of images I_1 and I_2, which is given by:

$$SSD = \sum_{\mathbf{x} \in \Omega_{1,2}} (I_1(\mathbf{x}) - I_2(\mathbf{x}))^2$$

SSD was selected in favour of an entropy-based measure because all images are based on the same modality. Alternatively, cross correlation can also be used as similarity metrics here. More information about similarity metrics is given in section 3.2.1.

We investigate three different ways of constructing the manifold graph: a single manifold, a pooled manifold and an aligned manifold.

- To preserve the inter-subject variation and extract anatomical information, first a low dimensional manifold is derived from a single time-point. In this work, we use the ED phase as this single time point but other choices are possible. We refer to this as *single manifold*.
- To preserve the inter- and intra-subject variation and in order to capture functional and motion information in cardiac MR images, we perform the manifold learning on the cardiac MR image sequences. For this purpose, we propose two different approaches:
  - *Pooled manifold*. More than one time frames from the same subject are simultaneously embedded together with their ED phase images.
  - *Aligned manifold*. Previous studies [Chang et al., 2007] revealed that for the pooled manifold important intra-subject variation can be lost in the embedding while the inter-subject variation dominates the manifold. Therefore, we derive an individual manifold for each individual cardiac MR image sequences. These manifold are then aligned into the final combined manifold using Procrustes analysis [Wang and Mahadevan, 2008].

The three ways of constructing the manifold are explained in detail as following:

Single manifold

The single manifold $\boldsymbol{Y^{ED}}$ is obtained using the ED phase of the cardiac sequence for each subject. The similarity matrix $\boldsymbol{W^{ED,ED}}$ is defined in Equation 6.2 as follows:

$$\boldsymbol{W^{ED,ED}} = \begin{pmatrix} 1 & w_{1,2}^{ED,ED} & \dots & w_{1,j}^{ED,ED} & \dots & w_{1,n}^{ED,ED} \\ w_{2,1}^{ED,ED} & 1 & & \dots & & w_{2,n}^{ED,ED} \\ & \dots & & \dots & & \\ w_{i,1}^{ED,ED} & \dots & \dots & w_{i,j}^{ED,ED} & \dots & w_{i,n}^{ED,ED} \\ & \dots & & \dots & & \\ w_{n,1}^{ED,ED} & \dots & \dots & \dots & w_{n,n-1}^{ED,ED} & 1 \end{pmatrix} \tag{6.2}$$

Here $w_{i,j}^{ED,ED}$ denotes the strength of connection between the ED phase of subject s_i and the ED phase of s_j ($i, j \in [1, n]$, n is the number of subjects).

Pooled manifold

In this model, more than one frame from the image sequence per subject are used to construct the similarity matrix for the pooled manifold learning. Similar to Equation 6.2, if we define $\boldsymbol{W_{i,j}}$ as the similarity matrix measuring the connectivity between the n frames in subject s_i and n frames in subject s_j as follows:

$$\boldsymbol{W_{i,j}} = \begin{pmatrix} w_{i,j}^{1,1} & w_{i,j}^{1,2} & \dots & w_{i,j}^{1,v} & \dots & w_{i,j}^{1,f} \\ w_{i,j}^{2,1} & w_{i,j}^{2,2} & & \dots & & w_{i,j}^{2,f} \\ & \dots & & \dots & & \\ w_{i,j}^{u,1} & \dots & \dots & w_{i,j}^{u,v} & \dots & w_{i,j}^{u,f} \\ & \dots & & \dots & & \\ w_{i,j}^{f,1} & \dots & \dots & w_{i,j}^{f,v} & \dots & w_{i,j}^{f,f} \end{pmatrix} \tag{6.3}$$

Here $w_{i,j}^{u,v}$ denotes the weight between frame u in subject s_i and frame v in subject s_j. f is the total number of frames used per subject. Therefore the full similarity matrix for the pooled

manifold can be defined as:

$$\boldsymbol{W_{pooled}} = \begin{pmatrix} W_{1,1} & W_{1,2} & ... & W_{1,j} & ... & W_{1,n} \\ W_{2,1} & W_{2,2} & & ... & & W_{2,n} \\ & ... & & ... & & \\ W_{i,1} & ... & ... & W_{i,j} & ... & W_{i,n} \\ & ... & & ... & & \\ W_{n,1} & ... & ... & W_{n,j} & ... & W_{n,n} \end{pmatrix} \tag{6.4}$$

Aligned manifold

In addition, we constructed the aligned manifold model in order to not only preserve global properties in data, but also maintain local properties of the images in the embedded space. In this work, the manifold alignment is performed using Procrustes analysis, which seeks to find the scaling, rotation and translation required to best match two data sets [Wang and Mahadevan, 2008].

In our proposed approach we first learn a low dimensional embedding for each frame set. For example, if two cardiac phases are used per subject, e.g. ED and ES, then a manifold $\boldsymbol{Y^{ED}}$ for all ED images and a manifold $\boldsymbol{Y^{ED}}$ for all ES images are computed. The similarity matrix $\boldsymbol{W^u}$ for measuring the similarity between frame u in n subjects is defined as following:

$$\boldsymbol{W^u} = \begin{pmatrix} 1 & w_{1,2}^{u,u} & ... & w_{1,j}^{u,u} & ... & w_{1,n}^{u,u} \\ w_{2,1}^{u,u} & 1 & & ... & & w_{2,n}^{u,u} \\ & ... & & ... & & \\ w_{i,1}^{u,u} & ... & ... & w_{i,j}^{u,u} & ... & w_{i,n}^{u,u} \\ & ... & & ... & & \\ w_{n,1}^{u,u} & ... & ... & ... & w_{n,n-1}^{u,u} & 1 \end{pmatrix} \tag{6.5}$$

Here $w_{i,j}^{u,u}$ denotes the strength of connection between frame u of subject s_i and frame u in subject s_j ($i, j \in [1, n]$, n is the total number of subjects). Then we derive a set of manifolds

$\{\boldsymbol{Y^u}, u = 1..f\}$ (f is the total number of frames used) for each cardiac phase using the weight matrix $\boldsymbol{W^u}$. In our notation $\boldsymbol{Y^1}$ is $\boldsymbol{Y^{ED}}$.

Thereafter, each derived manifold $\boldsymbol{Y^u}$ ($u \neq 1$) is mapped to $\boldsymbol{Y^{ED}}$ as $\boldsymbol{Y^{u*}}$ using Procrustes analysis. Procrustes analysis alignment removes the translational, rotational and scaling components so that optimal alignment between the two manifolds can be achieved. The final aligned manifold is formed as:

$$\boldsymbol{Y_{aligned}} = \{\boldsymbol{Y^{ED}}, \boldsymbol{Y^{2*}}, \boldsymbol{Y^{3*}}, ..., \boldsymbol{Y^{f*}}\} \tag{6.6}$$

In order to investigate that how many frames are required to capture functional information, we derive the aligned manifold from images with different temporal resolutions. The single manifold can be seen as the manifold with a temporal resolution of one image during a cardiac cycle. We then increase the temporal resolution to two frames (the ED and ES frames) and so on. In this work, we gradually increase the temporal resolution from one frame to twenty frames per cardiac cycle and build the aligned manifold for each different temporal resolution respectively.

6.2.4 Classification

The representation obtained through manifold learning allows the embedding of inter- and intra-subject variation of the cardiac anatomy and function into a lower dimensional space. This makes it possible to use the embedding coordinates $\mathbf{y_i}$ as the feature vectors to analyze the population in terms of cardiac morphology and function. We investigate the usefulness of the extracted manifold coordinates by evaluating two classification scenarios: a) male vs. female and b) subject with high blood pressure vs. subject with normal blood pressure. Figure 6.5 illustrates the four groups of low, ideal, pre-high and high that [HighBloodPressUK, 2015] the blood pressure is generally classified into. In our dataset, subjects are classified into two categories according to their blood pressure: high blood pressure refers to the high blood pressure group in Figure 6.5 and normal blood pressure refers to the rest three groups.

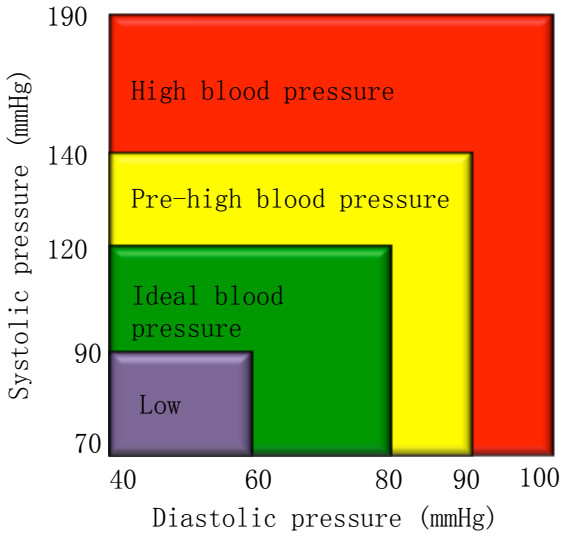


Figure 6.5 – The blood pressure chart, adapted from [HighBloodPressUK, 2015].

Support vector machines (SVMs) [Vapnik, 1995] [Vapnik, 1998] can be used for this binary classification problem. SVMs arrange the original input space into a high-dimensional feature space according to the kernel function and separate classes in the feature space by minimizing the classification error on the training set and maximizing the margin between classes. Typically, the SVM kernel functions $\phi(X)$ are linear, polynomial, hyperbolic tangent and radial basis functions. Through the application of the kernel function, the hyperplane can be constructed in a transformed feature space. The Gaussian radial basis function kernel is used in this work to map the data into a higher dimensional space. More information about SVMs can be found in section 2.3.1.

6.2.5 Regression

The manifold embeddings can also be used to predict continuous clinical variables. To achieve such continuous prediction, a regression model is constructed to investigate the relationship between the manifold coordinates and continuous clinical variables. The clinical variable that we are interested in are age and obesity indices. Aging is the most important risk factor in developing heart diseases [Finegold et al., 2012] and also causes structural change of the heart [Jani and Rajkumar, 2006] [Lakatta, 1990] [Sandstede et al., 2000]. In addition, obesity

is associated with a number of heart diseases. It would therefore be valuable to define the relationship between age/obesity and features learnt from the cardiac images.

In this work, a SVM regression (SVR) model is used to regress age, BMI, FFMI and Fat% from derived manifold coordinates. In SVM regression, the inputs are first mapped into a high-dimensional feature space using a nonlinear mapping, and thereafter a linear model is constructed in this feature space. As in the classification task in this work, radial basis function kernel is chosen as the mapping function. Detailed information about SVM regression can be found in section 2.3.4.

As described in section 2.3.4, the root mean squared error (RMSE) and the coefficient of determination (R^2) are used as statistics for regression analysis. The correlation coefficient (CC) is also calculated to investigate the strength of the correlation between the predicted and real value of the variables. In particular, the median error is also used as accuracy measure. This value refers to the median absolute difference between the predicted and real value. For instance, a median error of 5.84 years indicates that predicted age differs by less than 5.84 years in 50% of subjects. The median error, which is used in [Horvath, 2013] to evaluate the age predictor proposed in this paper, eliminates the influence of the standard deviation and thus is well suited for datasets with large standard deviation.

6.3 Experiments and results

6.3.1 Impact of the parameter setting

Selecting value for neighbourhood size and feature dimension

The neighbourhood size (k) and feature dimension (d) are two important parameters in the Laplacian eigenmap embedding. The neighbourhood size (k) is selected for the sparsification of a full similarity matrix W of size $N \times N$, which is typically performed by retaining the k-nearest neighbours for each data point in W and setting all other values to zero. The selection of the

optimal feature dimension d is not trivial. It is advisable that $d \ll N$ for manifold learning and SVMs. The correct choice of these two parameters is important to obtain meaningful representations of the data. In this work, these parameters are selected empirically using an exhaustive 2D grid search.

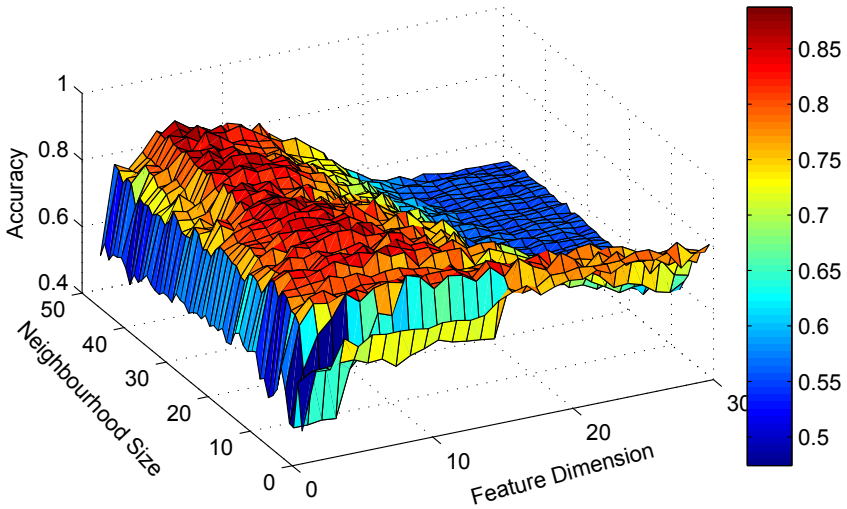
Figure 6.6 shows classification accuracies computed based on a single manifold within the parameter grid for classifying subjects according to gender and blood pressure respectively. We show the average classification accuracies obtained after 100 runs of seven-fold cross-validation. As can be seen for the gender classification task, the accuracy initially increases sharply as k and d increase and reaches its peak at $k = 45, d = 7$. The classification accuracy decays significantly for large d. By contrast, for the blood pressure classification task, the accuracy increases steadily and remains stable for $k > 10$ and $d > 15$.

For the above two cases, increasing the dimension of the features initially improves the classification accuracy. However, including further additional dimensions after a certain level has less of an influence on the classification result.

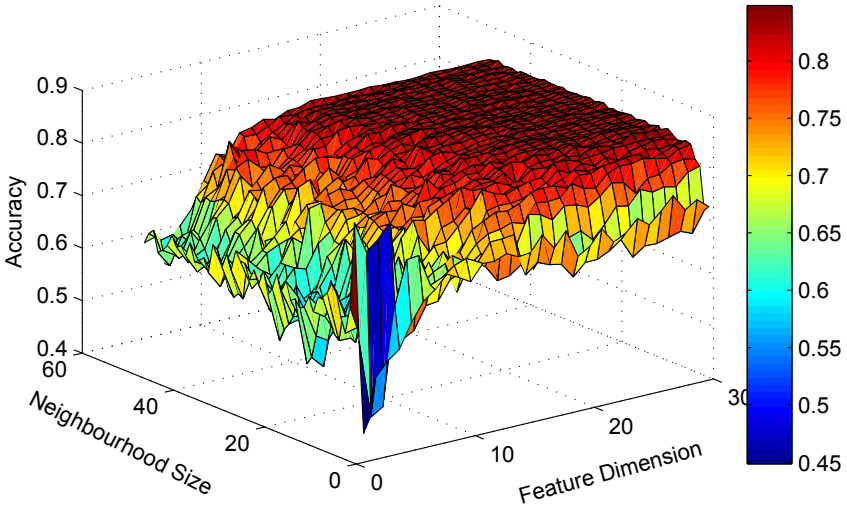
Selecting temporal resolution for aligned-manifold

In order to achieve the best classification rate and investigate the impact of the neighbourhood size, feature dimension and temporal resolution for the pooled and aligned manifold, the optimal parameter settings are investigated using grid search as well. The results for the classification according to blood pressure using the aligned manifold is reported in Figure 6.7. 100 runs of seven-fold cross-validation are performed for each of the settings: neighbourhood size $k \in [2..50]$, feature dimension $d \in [1..20]$ and number of frames $f \in [1..20]$. The search range chosen for the neighbourhood size and feature dimension are based on the results obtained in the previous section. Clear improvements in the classification accuracy can be observed in both figures when increasing any of the parameters. Thereafter the accuracy stabilizes with the best performance obtained at $k = 29, d = 4, f = 16$.

Figure 6.8 plots the change of R^2 when regressing age and three forms of obesity indices on



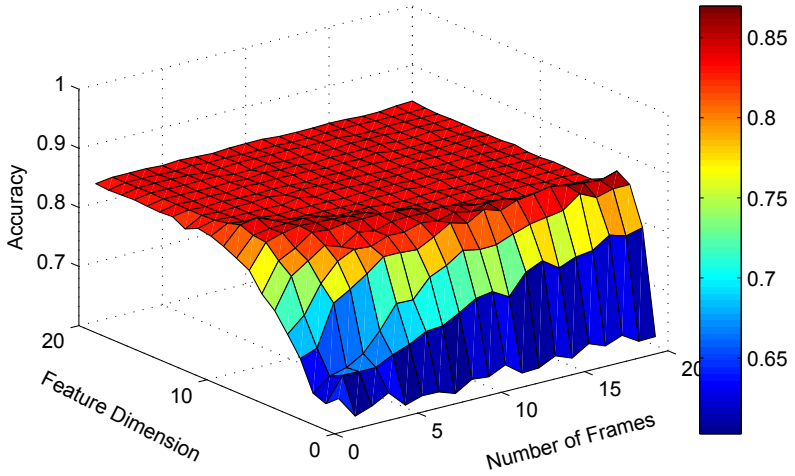
(a)



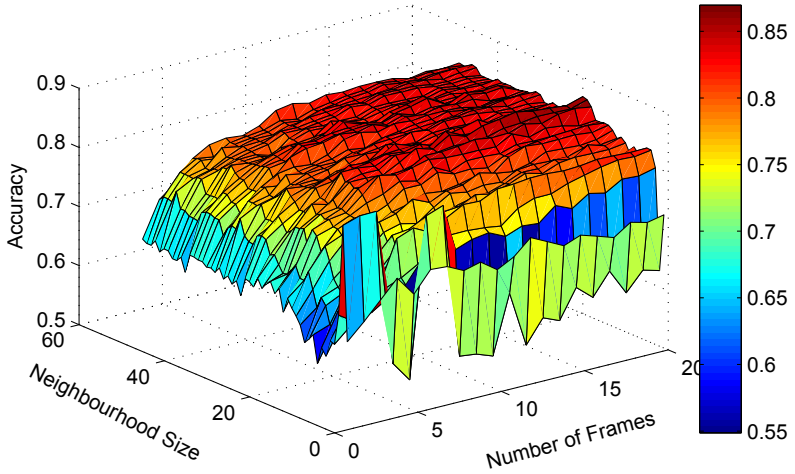
(b)

Figure 6.6 – Average classification accuracies computed based on the single manifold with the varying neighbourhood size and feature dimension for discriminating between (a) gender and (b) subject with high vs. normal blood pressure.

aligned manifold coordinates with varying temporal resolution. The temporal resolution shows moderate influence on regression performance. The improvements of the regression R^2 are rather minor along with the increase of the temporal resolution for all the four cases; and the R^2 reach their peaks all before the available full resolution used.



(a)



(b)

Figure 6.7 – The impact of (a) the temporal resolution f and feature dimension d and (b) neighbourhood size k and the temporal resolution f for classifying subjects with high and normal blood pressure using the aligned manifold.

6.3.2 Results

Visualising the low dimensional embedding

Visualisation of single manifold Figure 6.9 visualizes the first two manifold coordinates for both the fully connected graph and sparsely connected graph learned from ED images. Comparing the two figures, the sparse graph illustrates a clear separation of subjects according to gender. Furthermore, the same single manifold plotted in Figure 6.10 is color-coded by

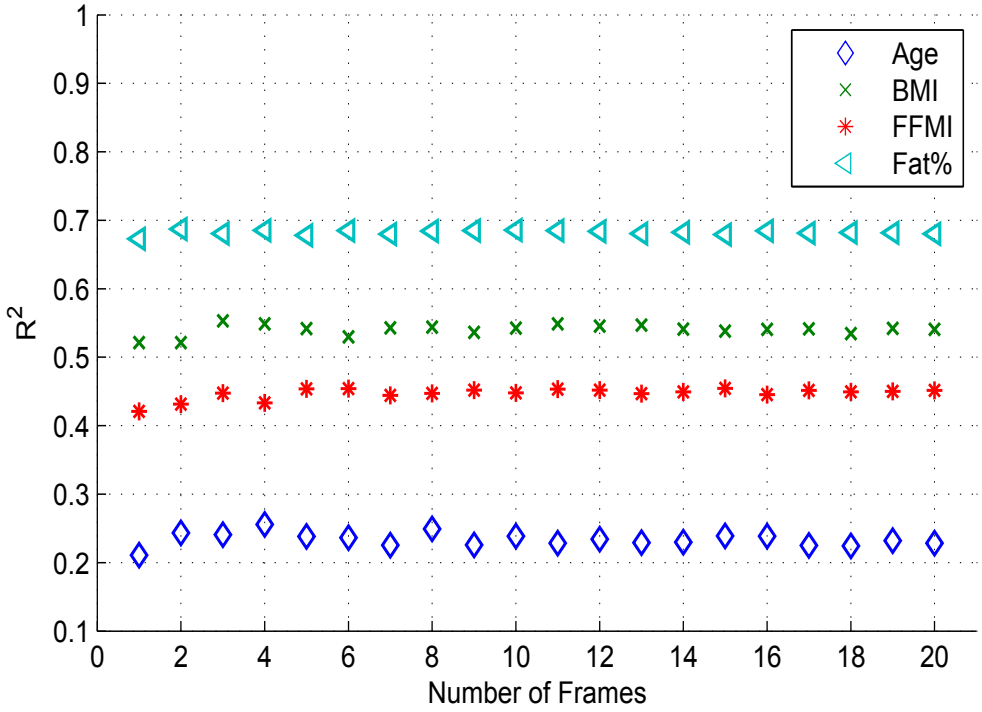
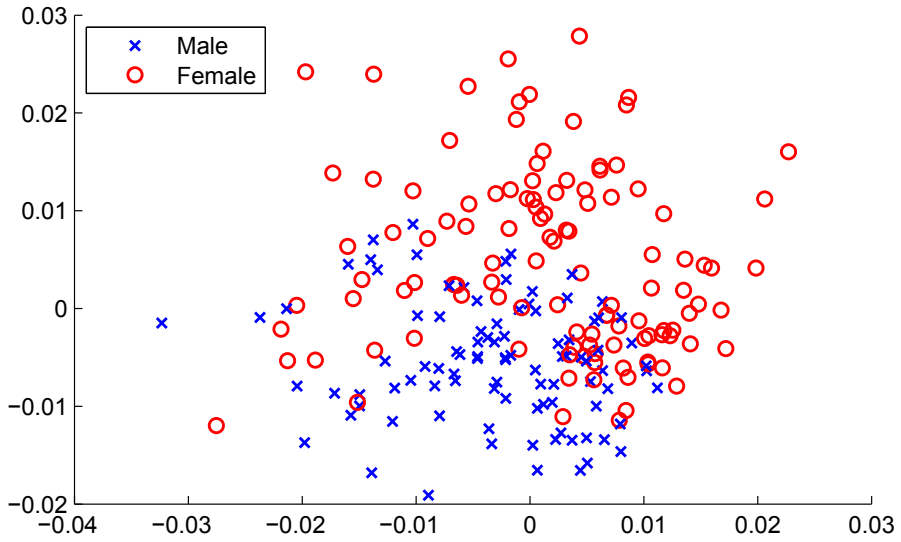


Figure 6.8 – The impact of the temporal resolution on regressing manifold coordinates versus age, BMI, FFMI and fat% respectively.

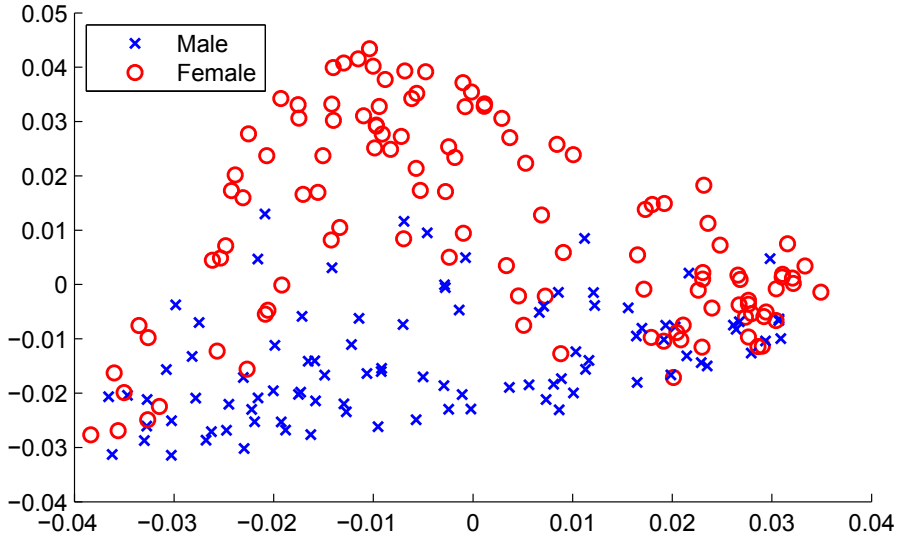
age, BMI, FFMI and Fat% respectively. The visualisation is able to show that subjects with similar values cluster together and this value associated distribution is more obvious in Figure 6.10(b)-(c).

Visualisation of pooled manifold Figure 6.11 illustrates two manifolds learned from two frames (ED and ES images) and from all frames respectively. Images are labeled by gender. The ED-ES manifold shows an encouraging pattern: the connection between some ED-ES pairs varies from pair to pair despite most of the ED-ES pairs from the same subject staying overlapped. However, the manifold constructed from all frames shows only the exact same plot as ED-ES manifold, without any more variations in the intra-subject connections. This is due to the fact that the inter-subject variation are much larger than that of the intra-subject; and the ED-ES variation is already the largest among those intra-subject variations. Consequently we propose the aligned manifold, which can illustrate the inter-subject variation and maximise the presentation of the intra-subject information at the same time.

Visualisation for aligned manifold The alignment of the manifolds is performed by



(a)



(b)

Figure 6.9 – 2D visualizations of single manifolds labeled by gender, (a) a full graph, (b) a sparse graph ($k = 24$).

Procrustes analysis. Figure 6.12 visualizes the aligned manifold labeled according to gender, while Figure 6.12(a) shows the manifold including all subject and Figure 6.12(b) only shows 50 subjects for illustration purpose. It can be observed that the frames from the same subject cluster together. The overlaid images show similar anatomical appearance with its neighbours (as annotated in yellow). Furthermore, the aligned manifolds successfully demonstrate the varying pattern of the intra-subject connections.

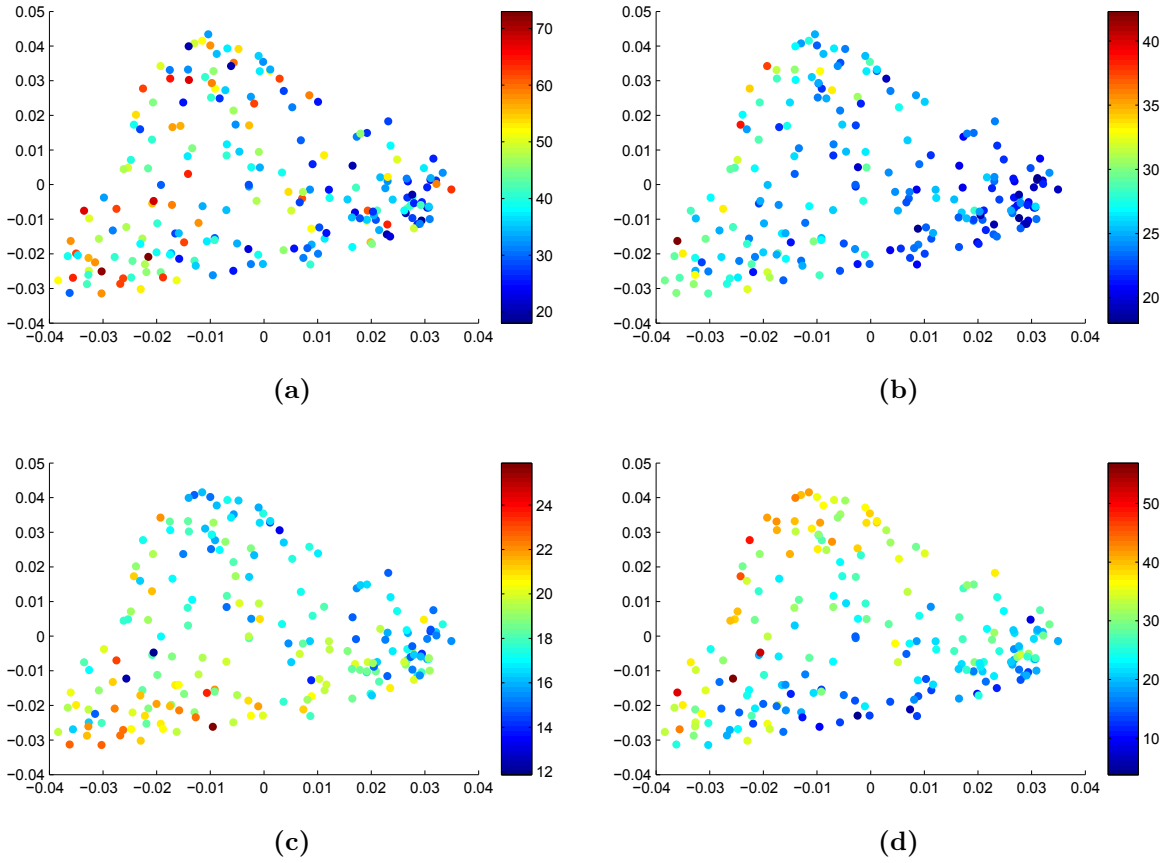
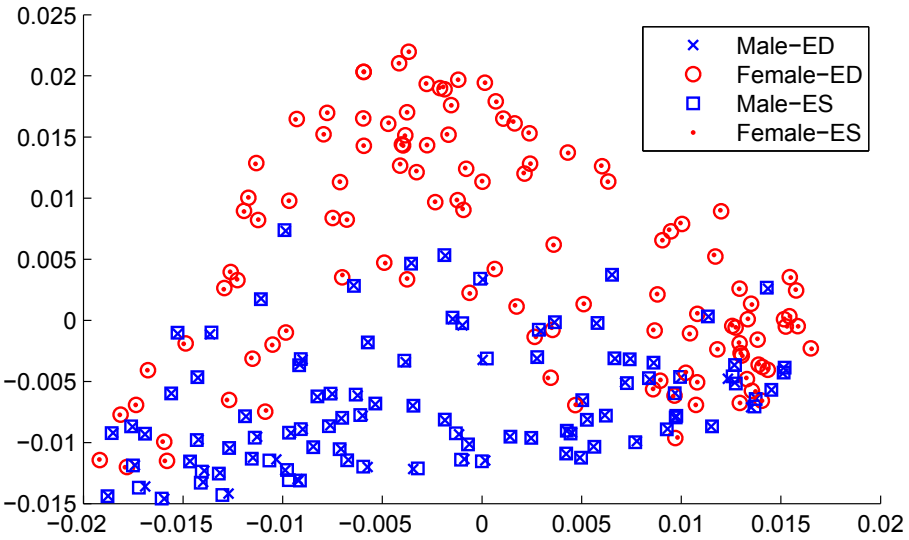


Figure 6.10 – 2D visualizations of the first two coordinates of single manifolds for subjects colour-labeled by (a) age, (b) BMI, (c)FFMI and (d)Fat%.

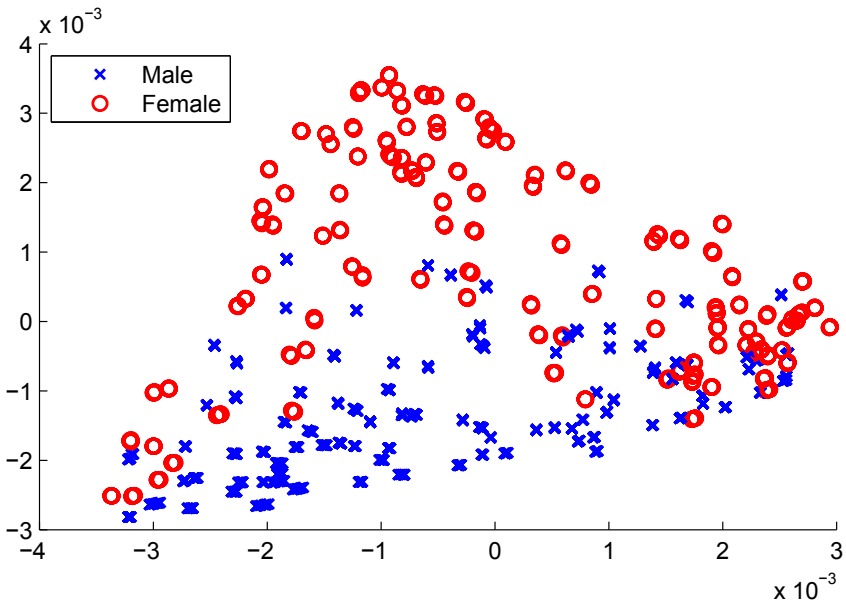
Classification Results

The single manifold, pooled manifold and aligned manifold proposed in this work are used to select features for classification between the gender and between subjects with high and normal blood-pressure. The single manifold is learned from the ED images, while the pooled and aligned manifolds are learned from different numbers of temporal frames. A SVM with a Gaussian radial basis function kernel is used for classification.

Table 6.2 reports the classification accuracy results based on the three types of manifolds for both classification tasks. For comparison, the classification accuracy obtained using clinical information such as LV volume, stroke volume and LV mass are also shown. It can be seen that the classification accuracy using manifold learning coordinates as features outperforms the use of clinical information. Among the three types of manifolds, the aligned manifold has a distinct advantage over the other two manifold models for classification according to blood pressure.



(a)



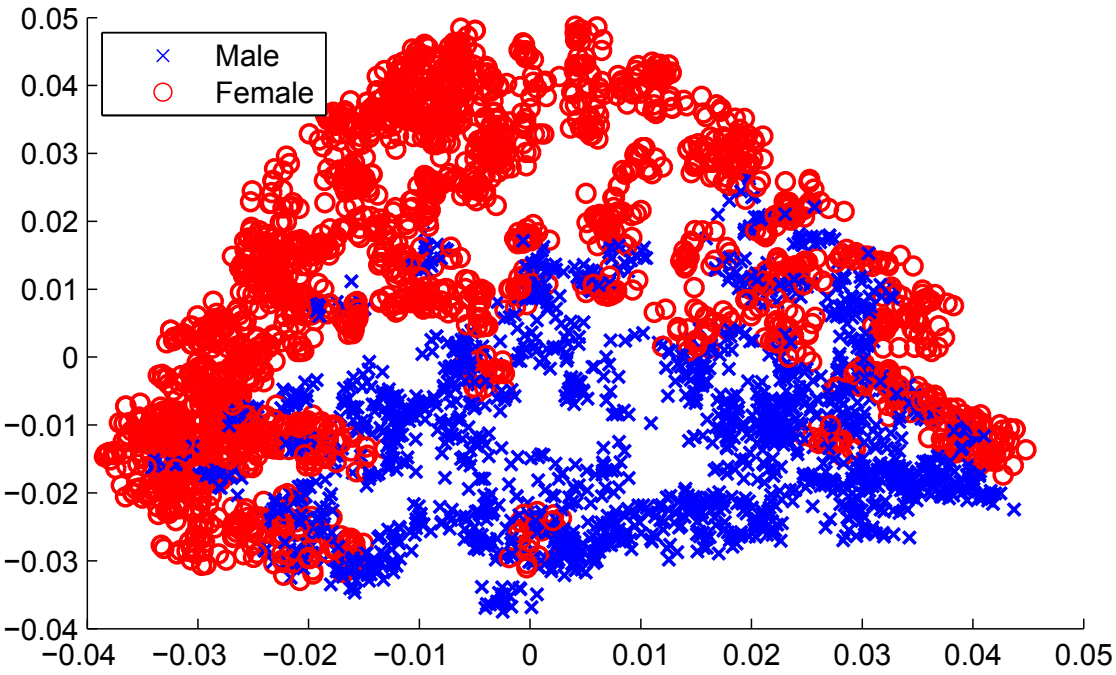
(b)

Figure 6.11 – 2D visualizations of pooled manifolds for subjects labeled by gender, learned from images of (a) ED,ES phases, and (b) all frames. The connection between some ED-ES pairs varies from pair to pair despite most of the ED-ES pairs from the same subject staying overlapped.

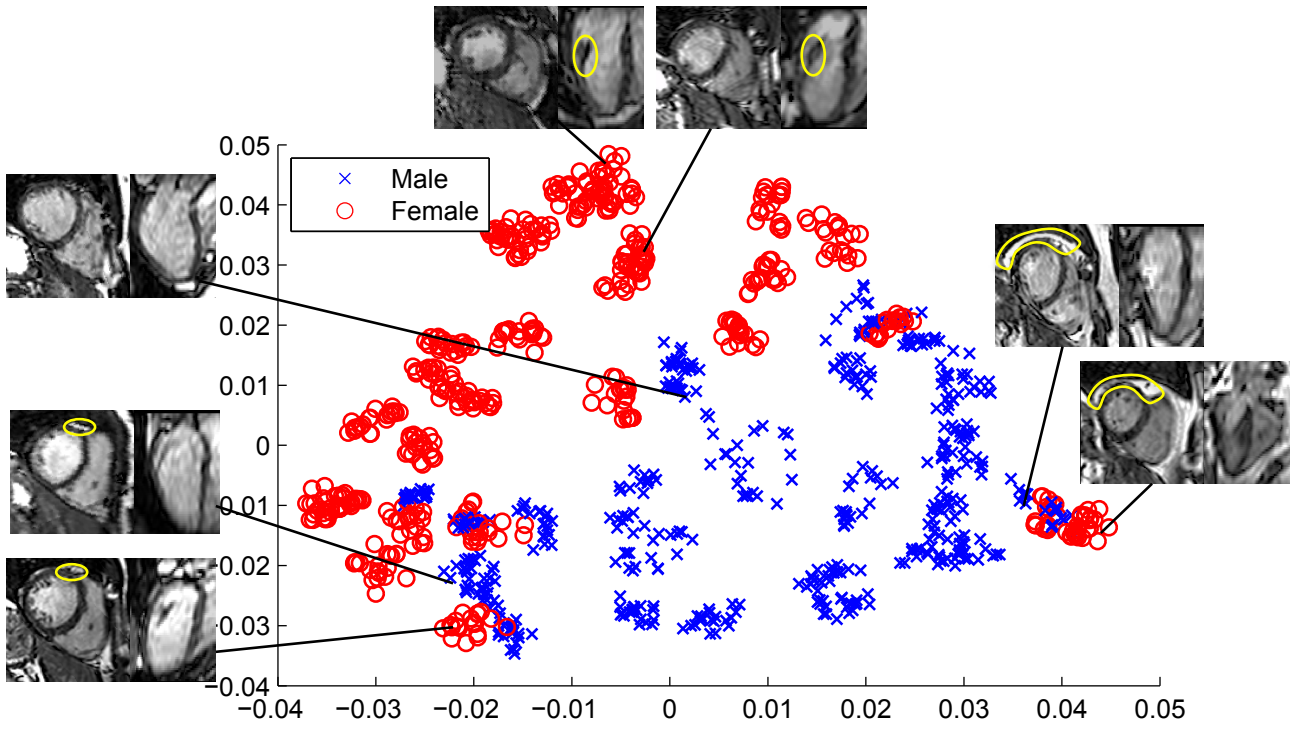
For gender discrimination, all the three manifold models reach the same classification accuracy.

Regression Results

SVM regression are performed for regressing age, BMI, FFMI and Fat% using single manifold coordinates and aligned manifold coordinates. The single manifold is learned from the ED



(a)



(b)

Figure 6.12 – Visualization of the aligned manifold labeled by genders: (a) shows the manifold including all subject, and (b) shows only 50 subjects for illustration purpose. The overlaid images show similar anatomical appearance with its neighbours (as annotated in yellow).

Table 6.2 – Comparison of classification accuracy based on the different types of manifolds. For comparison, we have also investigated how well clinical measurements can be used to separate subjects.

	Gender	Blood Pressure
Single manifold	0.89 ± 0.02	0.83 ± 0.01
Pooled manifold	0.89 ± 0.01	0.84 ± 0.01
Aligned manifold	0.89 ± 0.02	0.87 ± 0.01
LV volume (ED)	0.77 ± 0.01	0.57 ± 0.01
LV volume (ES)	0.75 ± 0.01	0.54 ± 0.02
Stroke volume	0.72 ± 0.02	0.58 ± 0.02
LV mass	0.85 ± 0.01	0.63 ± 0.01

images, while the aligned manifolds are derived with varying temporal resolutions. SVM ϵ regression with Gaussian Radial Basis function kernel is applied using ten-fold cross-validation. Average regression metrics for cross-validation of 100 runs are obtained. The fat mass/fat free mass value is missing or invalid for nine subjects in the dataset, hence 200 subjects are used for regression analysis of FFMI and Fat% whereas the total 209 subjects are used for age and BMI.

The results from the SVM regression demonstrate that there is a high correlation between the manifold coordinates with obesity indices. Among them, BMI and Fat% perform particularly well that median error of BMI achieved as small as 0.38(1.6% of median BMI) and median error of Fat% achieved 0.55(3.0% of median $Fat\%$), as shown in Figure 6.13 and Table 6.3. The median error of 5.84 for age shows a reasonable good prediction despite of the large RMSE. In order to validate the significance of the results, we also performed permutation test. We randomly shuffle the clinical variables, i.e. the clinical variables are no longer paired with their own feature vector (manifold coordinates) but a random one, then repeat the SVM regression process. These permutation results are plotted in Figure 6.13 for comparison.

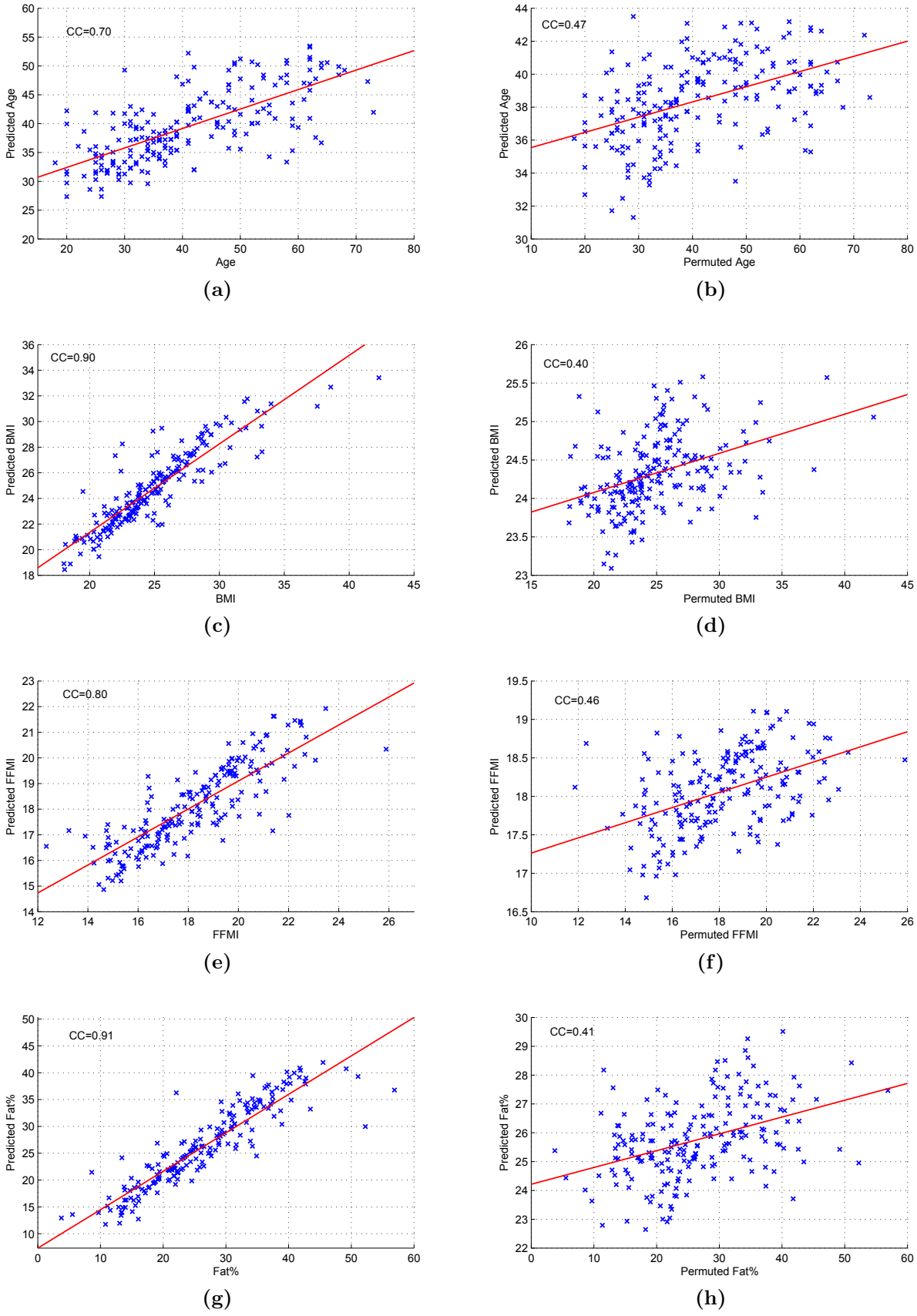


Figure 6.13 – Prediction results using SVM regression for (a)age, (b)permuted age, (c)BMI, (d)permuted BMI, (e)FFMI, (f)permuted FFMI, (g)Fat% and (h)permuted Fat%.

Table 6.3 – Results of SVM regression, using the coordinates from the manifold as predictor where neighbourhood size, feature dimension and temporal resolution are chosen for each task individually. p-values < 0.001 for the R^2 values quoted. Results are reported for best temporal resolution of aligned manifold and single manifold. Median error refers to the median absolute difference between the predicted and real value. The percentages of Median error/Median are also computed.

	RMSE		Median Error		R^2		R^2
	aligned	single	aligned	single	aligned	single	permutation test
Age	11.22 ± 0.12	11.53 ± 0.12	5.84(15.4%)	4.20(11.1%)	0.26 ± 0.02	0.21 ± 0.02	0.07
BMI	2.58 ± 0.04	2.68 ± 0.04	0.38(1.6%)	0.70(2.9%)	0.55 ± 0.02	0.52 ± 0.02	0.06
FFMI	1.77 ± 0.03	1.82 ± 0.02	0.55(3.0%)	0.72(4.0%)	0.45 ± 0.02	0.42 ± 0.02	0.11
Fat%	5.40 ± 0.09	5.53 ± 0.08	1.22(4.8%)	1.17(4.6%)	0.68 ± 0.01	0.67 ± 0.02	0.08

For the relatively poor correlation between the predicted age and the actual age, we speculate about whether the obesity status has an impact on it. Therefore we investigated the top 10% outliers with the actual age considerably above or below the age predicted by regression. These outliers are grouped based on the difference between the predicted and the actual age as follows: above set (averagely 18.80 above the actual age) and below set(averagely 21.52 below the actual age). For comparison, two reference groups are assembled with the same actual age ranges respectively. Table 6.4 reports the statistics for the two outlier sets and reference groups. The average BMI of the subjects in both above and below sets are significantly different from that of the reference groups. Fat% also differs distinctly in the above set and FFMI differs distinctly in the below set. The average BMI of the subjects in the above set is 26.35, which is larger than the WHO cut-off point for overweight of 25 [WHO Expert Consultation, 2004]. By contrast, the subjects in the below group have very healthy obesity indices, especially at their age (mean age is 63.06).

Table 6.4 – Statistics for outliers from age regression and reference groups, which have the same age ranges as outliers groups. Outliers are divided into above and below sets based on the difference between the predicted and the actual age. p-value is computed using t-test. Significant differences with p-value < 0.05 are labeled in **bold**.

		Real Age	Predicted Age	BMI	FFMI	Fat%	BP	Size
Above set	Outliers	23.60	**42.40**	**26.35**	18.29	**30.56**	116/75	5
	Ref.	26.33	**33.35**	**22.99**	17.66	**23.02**	115/75	39
	p-val	0.052	< 0.001	0.005	0.567	0.044	/	/
Below set	Outliers	62.93	**41.50**	**23.07**	**16.77**	26.67	131/82	16
	Ref.	60.21	**48.82**	**27.82**	**18.84**	31.17	123/81	19
	p-val	0.065	< 0.001	0.003	0.019	0.201	/	/

6.4 Conclusion

We successfully used manifold learning as a feature selection approach for a SVM-based classifier to analyse cardiac MR image sequences for 209 subjects. The SVM-based classifier directly operates on the manifold coordinates of the MR images without the need for any explicit image analysis such as registration or segmentation. We also demonstrated that, by considering

both inter- and intra-subject variation in the manifold learning, we are able to extract both anatomical and functional information. This can be used to construct powerful and reliable classifiers that are more predictive than global indices like LV volume and mass. The manifold enables the investigation of how much temporal information is necessary to help improve the classification performance.

Our experiments show that the neighbourhood size, feature dimension and amount of motion information have significant impact on the prediction results. For gender prediction additional temporal information and features beyond a certain point do not add useful information. This can be explained by the fact that gender is highly correlated to the volume of the heart but not to other properties of the myocardium. By contrast, blood pressure is closely related to both anatomy and function of the heart, and although including further additional manifold dimensions beyond a certain level has less of an influence on the classification result, temporal information is essential for accurate prediction. It is also interesting to see that the best classification performance is reached before using the full available temporal resolution.

The regression experiments demonstrate that there exists a strong correlation between the manifold embeddings and obesity indices. Among these indices, BMI and Fat% are significantly related to manifold embeddings: for BMI $R^2 = 0.55, r = 0.90 (p - value < 0.001)$; and $R^2 = 0.68, r = 0.91 (p - value < 0.001)$ for Fat%. The strong correlation discovered between obesity indices and manifold coordinates suggests that the body constitution has a strong impact on the appearance of the heart, and obesity indices are good indicator of the appearance and motion status of the heart, which might be used to provide an early warning for diabetes and CVD risks for healthy individuals.

In contrast to the visible influence of the temporal resolution on identifying subjects with high blood pressure, the improvements of the regression R^2 are rather flat along with the increase of the temporal resolution for all the obesity index studies. All the R^2s reach their peaks when only a few frames of the image sequence are used. This observation may suggest that the difference between the body constitution of healthy individuals can be observed in the appearance of heart, but is not yet observable in the cardiac motion. Additionally, the

investigation of outliers from the age-specific regression model indicates that higher BMI or Fat% were associated with the subjects whose predicted age is considerably above the actual age; while the lower obesity indices are associated with the subjects whose predicted age is considerably below the actual age.

The proposed approach is computationally efficient as no registration or segmentation is required. The presented work so far has focused on healthy volunteers. Future work will investigate the prognostic capabilities of the proposed manifold learning method in diseases such as dilated cardiomyopathy and pulmonary arterial hypertension. The correlation between obesity indices, age and cardiac MR images should be further explored in a larger dataset.

Chapter 7

Conclusion and Future Work

7.1 Conclusion

The ultimate objective of cardiac image analysis is to provide useful and efficient tools for the diagnosis and treatment of patients with cardiovascular diseases. The work in this thesis has been focused on the use of cardiac motion information to help achieve this goal.

In this work we have presented a machine learning algorithm to identify and track a set of key landmarks throughout the cardiac cycle in order to perform cardiac motion tracking using both tagged and untagged image sequences from short-axis and long-axis views simultaneously. By combining complementary information using a spatially adaptive weighting and valve plane constraint, we have successfully built an accurate and robust cardiac motion analysis framework.

In order to improve the speed and accuracy of the motion tracking, we have presented an approach to identify and track a sparse set of anatomical distinctive landmarks in the presence of relatively large deformations. This tracking does not require dense motion tracking. An integrated framework is proposed to combine entropy and SVD-based sparse landmark detection with a MRF-based motion tracking method. The accuracy of the proposed sparse motion tracking is evaluated by tracking a group of manually annotated landmarks on the endocardial border of the left ventricle (LV). The improvement of the tracking accuracy compared to

the dense motion tracking method shows that the proposed sparse motion tracking approach benefits from the larger search space of the MRF during the optimization.

We have developed a method to derive regional wall thickness SDI to provide accurate quantification of left ventricular motion. The regional cardiac functional analysis has been performed based on the proposed sparse landmark motion tracking results and the AHA LV model [Cerqueira et al., 2002]. Comparing with the analysis generated by a state-of-the-art commercial tool, the results obtained by the proposed method are in good agreement. This shows that it is plausible to use the proposed sparse landmark motion tracking to estimate the dyssynchrony index of the regional deformation.

We successfully used manifold learning as a feature selection approach for a SVM-based classifier to analyse 209 cardiac MR image sequences. The SVM-based classifier directly operates on the manifold coordinates of the MR images without the need for any explicit image analysis such as registration or segmentation. We also demonstrated that, by considering both inter- and intra-subject variation in the manifold learning, we are able to extract both anatomical and functional information. This can be used to construct powerful and reliable classifiers that are more predictive than global indices like LV volume and mass. The manifold allows for investigating how much temporal information is needed to improve the classification performance. The proposed manifold learning approach is computationally efficient as no non-rigid registration or segmentation is required.

Our experiments demonstrate that there exists a strong correlation between the manifold embeddings and obesity indices. Among these indices, BMI and Fat% are significantly related to manifold embeddings: for BMI $R^2 = 0.55, CC = 0.90 (p - value < 0.001)$; and $R^2 = 0.68, CC = 0.91 (p - value < 0.001)$ for Fat%. The strong correlation discovered between obesity indices and manifold coordinates suggests that the body constitution has a strong impact on the appearance of the heart, and obesity indices are good indicator of the appearance and motion status of the heart, which might be used to provide early warning for diabetes and CVD risks for healthy individuals.

In contrast to the visible influence of the temporal resolution on identifying subjects with

high blood pressure, the improvements of the regression R^2 are rather minor along with the increase of the temporal resolution for all the obesity index studies. All the R^2s reach their peaks when only a few frames of the image sequence are used. This observation may suggest that the difference between the body constitution of healthy individuals can be observed in the appearance of heart, but is not yet observable in the cardiac motion. Additionally, the investigation of outliers from the age-specific regression model indicates that higher BMI or Fat% are associated with the subjects whose predicted age is considerably above the actual age; while the lower BMI or Fat% are associated with the subjects whose predicted age is considerably below the actual age.

7.2 Future work

The work presented in chapter 4 and chapter 5 focuses on sparse motion tracking and the work in chapter 6 focuses on manifold learning based classification and regression. It would be beneficial to combine sparse motion tracking with aligned manifold learning for classification in order to utilise anatomically important voxels while keeping the computational cost low. The set of the sparse landmarks in each frame can be used to represent the corresponding anatomy. Because the presented framework selects and tracks a set of anatomically distinctive and important landmarks accurately, the landmarks are able to efficiently represent the image in manifold learning in a way that not only well captures the myocaridal motion but also discards redundant information. Therefore it is likely to improve the classification and regression performance.

The work on manifold learning so far has focused on healthy volunteers. We have successfully discriminated subjects according to gender and between subjects with high blood pressure and normal blood pressure. It would be valuable to extend this work to more clinical phenotypes. Future work will investigate the prognostic capabilities of the proposed manifold learning method in diseases such as dilated cardiomyopathy and pulmonary arterial hypertension. Mutations in more than 30 genes have been found to cause familial dilated cardiomyopathy, however, it usually takes many years for symptoms to appear. The UK1000 project, which

provides the dataset we used in chapter 6, has already detected mutations in protein-coding gene TTN in subjects, which account for approximately 20 percent of cases of familial dilated cardiomyopathy. As more data becomes available, a study to apply manifold learning to discover the correlation between MRI image and genetic mutation would be useful.

Obesity is associated with a number of heart diseases. Our work on regressing obesity indices in the manifold embeddings performs remarkable well. As pointed out in [Abbasi et al., 2002], the importance of obesity as a risk factor for diabetes and hypertension is well-recognized, but its role as a CVD risk factor in healthy individuals is less well established. In particular, to our knowledge no previous research has studied the relationship between the obesity indices and the cardiac images. A further study of the correlation between obesity indices and cardiac images in a larger dataset would be interesting. Additionally, a large body of research indicates that cardiac function gets changed with age. With out preliminary inspection of the relationship between age-specific predictor with regarding to cardiac images, it would also worth to further investigate along this direction. Associating age, gender and obesity indices with cardiac images analysis as metadata is also planned for future work.

Appendix A

Publications

- Haiyan Wang, Wenzhe Shi, Wenjia Bai, Antonio M. Simoes Monteiro de Marvao, Timothy J. W. Dawes, Declan P. O'Regan, Philip Edwards, Stuart Cook and Daniel Rueckert. (2015). Prediction of clinical information from cardiac MRI using manifold learning. Functional Imaging and Modeling of the Heart (FIMH), Volume 9126, 2015, 91-98.

- Haiyan Wang, Wenzhe Shi, Xiahai Zhuang, Xianliang Wu, Kai-Pin Tung, Sebastien Ourselin, Philip Edwards, and Daniel Rueckert. (2013). Landmark detection and coupled patch registration for cardiac motion tracking. Proc. SPIE 8669, Medical Imaging 2013: Image Processing, 86690J, J1-J6 .

- Haiyan Wang, Wenzhe Shi, Wenjia Bai, Philip Edwards, and Daniel Rueckert. (2013). 4D Sparse Landmark Cardiac Motion Tracking and Regional Function Analysis. Proc. MIUA 2013, 33-38.

- Haiyan Wang, Wenzhe Shi, Xiahai Zhuang, Simon Duckett, Kai Pin Tung, Philip Edwards, Reza Razavi, Sebastien Ourselin, and Daniel Rueckert. (2011). Automatic cardiac motion tracking using both untagged and 3D tagged MR images. Statistical Atlases and Computational Models of the Heart: Imaging and Modelling Challenges (STACOM), 45-56.

- Wenzhe Shi, Martin Jantsch, Paul Aljabar, Luis Pizarro, Wenjia Bai, Haiyan Wang,

Declan ORegan, Xiahai Zhuang, Daniel Rueckert (2013)Temporal sparse free-form deformations. Medical image analysis 17 (7), 779-789.

- Wenjia Bai, Wenzhe Shi, Declan P O'Regan, Tong Tong, Haiyan Wang, Shahnaz Jamil-Copley, Nicholas S Peters, Daniel Rueckert. (2013). A probabilistic patch-based label fusion model for multi-atlas segmentation with registration refinement: application to cardiac MR images. Medical Imaging, IEEE Transactions on 32 (7), 1302-1315

- Caroline Petitjean, Maria A. Zuluaga, Wenjia Bai, Jean-Nicolas Dacher, Damien Grosgeorge, Jrme Caudron, Su Ruan, Ismail Ben Ayed, M. Jorge Cardoso, Hsiang-Chou Chen, Daniel Jimenez-Carretero, Maria J. Ledesma-Carbayo, Christos Davatzikos, Jimit Doshi, Guray Erus, Oskar M.O. Maier, Cyrus M.S. Nambakhsh, Yangming Ou, Sbastien Ourselin, Chun-Wei Peng, Nicholas S. Peters, Terry M. Peters, Martin Rajchl, Daniel Rueckert, Andres Santos, Wenzhe Shi, Ching-Wei Wang, Haiyan Wang, Jing Yuan. (2014) Right ventricle segmentation from cardiac MRI: A collation study. Medical image analysis 19(1), 187-202.

- Xiahai Zhuang, Wenzhe Shi, Haiyan Wang, Daniel Rueckert, Sebastien Ourselin. (2013). Computation on shape manifold for atlas generation: application to whole heart segmentation of cardiac MRI. SPIE Medical Imaging, 866941-866941-7

- Kai-Pin Tung, Wen-Jia Bei, Wen-Zhe Shi, Hai-Yan Wang, Tong Tong, Ranil De Silva, Ephilip Edwards, Daniel Rueckert. (2013). Multi-atlas based neointima segmentation in intravascular coronary OCT. Biomedical Imaging (ISBI), 2013 IEEE 10th International Symposium on, 1280-1283

- Wenzhe Shi, Xiahai Zhuang, Luis Pizarro, Wenjia Bai, Haiyan Wang, Kai-Pin Tung, Philip Edwards, Daniel Rueckert. (2012). Registration using sparse free-form deformations. Medical Image Computing and Computer Assisted Intervention (MICCAI), 659-666.

- Wenzhe Shi, Xiahai Zhuang, Haiyan Wang, Simon Duckett, Duy V.N. Luong, Catalina Tobon-Gomez, KaiPin Tung, Philip Edwards, Kawal Rhode, Reza Razavi, Sebastien

Ourselin, Daniel Rueckert. (2012). A comprehensive cardiac motion estimation framework using both untagged and 3D tagged MR images based on non-rigid registration. IEEE Transactions on Medical Imaging, 31(6):1263-1275.

- Wenjia Bai, Wenzhe Shi, Haiyan Wang, Nicholas S Peters, Daniel Rueckert. (2012) Multi-atlas based segmentation with local label fusion for right ventricle MR images. Image 6, 9.

- Kai-Pin Tung, WenZhe Shi, Luis Pizarro Quiroz, HaiYan Wang, Ricardo Guerrero Moreno, Ranil De Silva, Philip Edwards, Daniel Rueckert. (2012). Automatic detection of coronary stent struts in intravascular OCT imaging. SPIE Medical Imaging, Computer Aided Diagnosis, Proc.8315, 2012.

- Wenzhe Shi, Xiahai Zhuang, Robin Wolz, Duckett Simon, KaiPin Tung, Haiyan Wang, Sebastien Ourselin, Philip Edwards, Reza Razavi, and Daniel Rueckert. (2011). A multi-image graph cut approach for cardiac image segmentation and uncertainty estimation. Statistical Atlases and Computational Models of the Heart: Imaging and Modelling Challenges (STACOM), 178-187.

- Wenzhe Shi, Xiahai Zhuang, Haiyan Wang, Duckett Simon, Declan Oregan, Sebastien Ourselin, Eddie Edwards, Daniel Rueckert. (2011). Automatic segmentation from cardiac cine MRI with different pathologies using registration and multiple component EM estimation. Functional Imaging and Modeling of the Heart (FIMH), 163-170.

- Xiahai Zhuang, Wenzhe Shi, S Duckett, Haiyan Wang, Reza Razavi, Daniel Rueckert, Sebastien Ourselin. (2011). A Framework Combining Multi-Sequence MRI for Fully Automated Quantitative Analysis of Cardiac Global And Regional Functions. Functional Imaging and Modeling of the Heart (FIMH), 367-374.

Bibliography

[WHO, 1995] (1995). Physical status: the use and interpretation of anthropometry. report of a who expert committee. *World Health Organ Tech Rep Ser*, 854:1–452.

[Ech, 2015] (2015). Echocardiography. http://www.echocardiographer.org/index.html. [Online; accessed 6-Feb-2015].

[Abbasi et al., 2002] Abbasi, F., Brown, B. W., Lamendola, C., McLaughlin, T., and Reaven, G. M. (2002). Relationship between obesity, insulin resistance, and coronary heart disease risk. *Journal of the American College of Cardiology*, 40(5):937–943.

[Abe, 2005] Abe, S. (2005). *Support Vector Machines for Pattern Classification (Advances in Pattern Recognition)*. Springer, 1 edition.

[Alahi et al., 2012] Alahi, A., Ortiz, R., and Vandergheynst, P. (2012). Freak: Fast retina keypoint. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 510–517. IEEE.

[Aljabar et al., 2012] Aljabar, P., Wolz, R., and Rueckert, D. (2012). Manifold learning for medical image registration, segmentation, and classification. *Machine learning in computer-aided diagnosis: medical imaging intelligence and analysis. IGI Global.*

[Aljaroudi et al., 2012] Aljaroudi, W., Halley, C., Houghtaling, P., Agarwal, S., Menon, V., Rodriguez, L., Grimm, R. A., Thomas, J. D., and Jaber, W. A. (2012). Impact of body mass index on diastolic function in patients with normal left ventricular ejection fraction. *Nutr Diabetes*, 2:e39.

[Amit and Geman, 1997] Amit, Y. and Geman, D. (1997). Shape quantization and recognition with randomized trees. *Neural computation*, 9(7):1545–1588.

[Andrews and Patterson III, 1976] Andrews, H. and Patterson III, C. (1976). Singular value decomposition (svd) image coding. *IEEE Transactions on Communications*, 24(4):425–432.

[Axel and Dougherty, 1989] Axel, L. and Dougherty, L. (1989). Mr imaging of motion with spatial modulation of magnetization. *Radiology*, 171(3):841–845.

[Axel et al., 2005] Axel, L., Montillo, A., and Kim, D. (2005). Tagged magnetic resonance imaging of the heart: a survey. *Medical image analysis*, 9(4):376–393.

[Azhari et al., 1990] Azhari, H., Sideman, S., Weiss, J. L., Shapiro, E. P., Weisfeldt, M. L., Graves, W. L., Rogers, W. J., and Beyar, R. (1990). Three-dimensional mapping of acute ischemic regions using mri: wall thickening versus motion analysis. *American Journal of Physiology-Heart and Circulatory Physiology*, 259(5):H1492–H1503.

[Bauer et al., 2011] Bauer, S., Nolte, L.-P., and Reyes, M. (2011). Fully automatic segmentation of brain tumor images using support vector machine classification in combination with hierarchical conditional random field regularization. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2011*, pages 354–361. Springer.

[Baumberg, 2000] Baumberg, A. (2000). Reliable feature matching across widely separated views. In *Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on*, volume 1, pages 774–781. IEEE.

[Bax et al., 2004] Bax, J. J., Bleeker, G., Marwick, T. H., Molhoek, S. G., Boersma, E., Steendijk, P., van der Wall, E. E., and Schalij, M. J. (2004). Left ventricular dyssynchrony predicts response and prognosis after cardiac resynchronization therapy. *J Am Coll Cardiol*, 44(9):1834–40.

[Bay et al., 2008] Bay, H., Ess, A., Tuytelaars, T., and Van Gool, L. (2008). Speeded-up robust features (surf). *Computer vision and image understanding*, 110(3):346–359.

[Bay et al., 2006] Bay, H., Tuytelaars, T., and Van Gool, L. (2006). Surf: Speeded up robust features. In *Computer Vision–ECCV 2006*, pages 404–417. Springer.

[Belkin and Niyogi, 2003] Belkin, M. and Niyogi, P. (2003). Laplacian eigenmaps for dimensionality reduction and data representation. *Neural computation*, 15(6):1373–1396.

[Bhatia et al., 2013] Bhatia, K., Rao, A., Price, A., Wolz, R., Hajnal, J., and Rueckert, D. (2013). Hierarchical manifold learning for regional image analysis.

[Bhatia et al., 2014] Bhatia, K., Rao, A., Price, A., Wolz, R., Hajnal, J., and Rueckert, D. (2014). Hierarchical manifold learning for regional image analysis.

[Bistoquet et al., 2008] Bistoquet, A., Oshinski, J., and Skrinjar, O. (2008). Myocardial deformation recovery from cine mri using a nearly incompressible biventricular model. *MEDICAL IMAGE ANALYSIS*, pages 69–85.

[Blake et al., 2011] Blake, A., Kohli, P., and Rother, C. (2011). *Markov random fields for vision and image processing*. MIT Press.

[Blumenthal et al., 1983] Blumenthal, D., Becker, L., Bulkley, B., Hutchins, G., Weisfeldt, M., and Weiss, J. (1983). Impaired function of salvaged myocardium: two-dimensional echocardiographic quantification of regional wall thickening in the open-chest dog. *Circulation*, 67(1):225–233.

[Bray et al., 1994] Bray, J. J., Cragg, P. A., Macknight, A. D., et al. (1994). *Lecture notes on human physiology*. London: Blackwell Scientific Publications, 1994.

[Breiman, 2001] Breiman, L. (2001). Random forests. *Machine learning*, 45(1):5–32.

[Brenner and Hall, 2007] Brenner, D. J. and Hall, E. J. (2007). Computed tomographyan increasing source of radiation exposure. *New England Journal of Medicine*, 357(22):2277–2284.

[Bruzzi et al., 2006] Bruzzi, J. F., Rémy-Jardin, M., Delhaye, D., Teisseire, A., Khalil, C., and Rémy, J. (2006). When, why, and how to examine the heart during thoracic ct: Part 1, basic principles. *AJR Am J Roentgenol*, 186(2):324–32.

[Buades et al., 2005] Buades, A., Coll, B., and Morel, J. (2005). A non-local algorithm for image denoising. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005. CVPR 2005.*, volume 2, pages 60–65. IEEE.

[Calonder et al., 2012] Calonder, M., Lepetit, V., Ozuysal, M., Trzcinski, T., Strecha, C., and Fua, P. (2012). Brief: Computing a local binary descriptor very fast. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 34(7):1281–1298.

[Calonder et al., 2010] Calonder, M., Lepetit, V., Strecha, C., and Fua, P. (2010). Brief: binary robust independent elementary features. In *Computer Vision–ECCV 2010*, pages 778–792. Springer.

[Camara et al., 2009] Camara, O., Oubel, E., Piella, G., Balocco, S., De Craene, M., and Frangi, A. (2009). Multi-sequence Registration of Cine, Tagged and Delay-Enhancement MRI with Shift Correction and Steerable Pyramid-Based Detagging. *Functional Imaging and Modeling of the Heart*, pages 330–338.

[Cerqueira et al., 2002] Cerqueira, M., Weissman, N., Dilsizian, V., Jacobs, A., Kaul, S., Laskey, W., Pennell, D., Rumberger, J., Ryan, T., and Verani, M. (2002). Standardized myocardial segmentation and nomenclature for tomographic imaging of the heart. *Journal of Cardiovascular Magnetic Resonance*, pages 203–210.

[Chander et al., 2008] Chander, A., Brenner, M., Lautamäki, R., Voicu, C., Merrill, J., and Bengel, F. M. (2008). Comparison of measures of left ventricular function from electrocardiographically gated 82rb pet with contrast-enhanced ct ventriculography: a hybrid pet/ct analysis. *Journal of Nuclear Medicine*, 49(10):1643–1650.

[Chandrashekara et al., 2004a] Chandrashekara, R., Mohiaddin, R., and Rueckert, D. (2004a). Analysis of 3-D myocardial motion in tagged MR images using nonrigid image registration. *IEEE Tansactions on Medical Imaging*, pages 1245–1250.

[Chandrashekara et al., 2004b] Chandrashekara, R., Mohiaddin, R., and Rueckert, D. (2004b). Cardiac motion tracking in tagged mr images using a 4d b-spline motion model and nonrigid

image registration. *2nd IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, 1-2:468–471.

[Chang et al., 2007] Chang, W.-Y., Chen, C.-S., and Hung, Y.-P. (2007). Analyzing facial expression by fusing manifolds. In *Computer Vision–ACCV 2007*, pages 621–630. Springer.

[Chapelle and Vapnik, 1999] Chapelle, O. and Vapnik, V. (1999). Model selection for support vector machines. In *NIPS*, pages 230–236.

[Chen et al., 2010] Chen, J., Shan, S., He, C., Zhao, G., Pietikainen, M., Chen, X., and Gao, W. (2010). Wld: A robust local image descriptor. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32(9):1705–1720.

[Cigarroa et al., 1993] Cigarroa, C., Brickner, M., Alvarez, L., Wait, M., Grayburn, P., et al. (1993). Dobutamine stress echocardiography identifies hibernating myocardium and predicts recovery of left ventricular function after coronary revascularization. *Circulation*, 88(2):430–436.

[Collignon et al., 1995] Collignon, A., Maes, F., Delaere, D., Vandermeulen, D., Suetens, P., and Marchal, G. (1995). Automated multi-modality image registration based on information theory. In *Information processing in medical imaging*, volume 3, pages 264–274.

[Cootes et al., 2010] Cootes, T., Twining, C., Petrovic, V., Babalola, K., and Taylor, C. (2010). Computing accurate correspondences across groups of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(11):1994–2005.

[Coupé et al., 2011] Coupé, P., Manjón, J. V., Fonov, V., Pruessner, J., Robles, M., and Collins, D. L. (2011). Patch-based segmentation using expert priors: application to hippocampus and ventricle segmentation. *Neuroimage*, 54(2):940–54.

[Coupe et al., 2008] Coupe, P., Yger, P., Prima, S., Hellier, P., Kervrann, C., and Barillot, C. (2008). An optimized blockwise nonlocal means denoising filter for 3-d magnetic resonance images. *IEEE Transactions On Medical Imaging*, 27(4):425–441.

[Cox and Cox, 2000] Cox, T. F. and Cox, M. A. (2000). *Multidimensional scaling*. CRC Press.

[Criminisi et al., 2013] Criminisi, A., Robertson, D., Pauly, O., Glocker, B., Konukoglu, E., Shotton, J., Mateus, D., Möller, A. M., Nekolla, S., and Navab, N. (2013). Anatomy detection and localization in 3d medical images. In *Decision Forests for Computer Vision and Medical Image Analysis*, pages 193–209. Springer.

[Criminisi and Shotton, 2013] Criminisi, A. and Shotton, J. (2013). *Decision forests for computer vision and medical image analysis*. Springer Science & Business Media.

[Crow, 1984] Crow, F. (1984). Summed-area tables for texture mapping. In *Proceedings of SIGGRAPH*, volume 18, pages 207–12.

[Crum et al., 2014] Crum, W. R., Hartkens, T., and Hill, D. (2014). Non-rigid image registration: theory and practice.

[Dalal and Triggs, 2005] Dalal, N. and Triggs, B. (2005). Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 886–893. IEEE.

[De Craene et al., 2012] De Craene, M., Piella, G., Camara, O., Duchateau, N., Silva, E., Doltra, A., D'hooge, J., Brugada, J., Sitges, M., and Frangi, A. F. (2012). Temporal diffeomorphic free-form deformation: application to motion and strain estimation from 3d echocardiography. *Medical Image Analysis*, 16(2):427–450.

[de Marvao et al., 2014] de Marvao, A., Dawes, T. J., Shi, W., Minas, C., Keenan, N. G., Diamond, T., Durighel, G., Montana, G., Rueckert, D., Cook, S. A., et al. (2014). Population-based studies of myocardial hypertrophy: high resolution cardiovascular magnetic resonance atlases improve statistical power. *Journal of Cardiovascular Magnetic Resonance*, 16(1):16.

[Dijkstra, 1959] Dijkstra, E. W. (1959). A note on two problems in connexion with graphs. *Numerische mathematik*, 1(1):269–271.

[Donnan et al., 2008] Donnan, G. A., Fisher, M., Macleod, M., and Davis, S. M. (2008). Stroke. *Lancet*, 371(9624):1612–23.

[Donoghue et al., 2011] Donoghue, C., Rao, A., Bull, A., and Rueckert, D. (2011). Manifold learning for automatically predicting articular cartilage morphology in the knee with data from the osteoarthritis initiative (oai). In *SPIE Medical Imaging*, pages 79620E–79620E. International Society for Optics and Photonics.

[Duchateau et al., 2012] Duchateau, N., Craene, M. D., Piella, G., and Frangi, A. F. (2012). Constrained manifold learning for the characterization of pathological deviations from normality. *Medical Image Analysis*.

[Duchateau et al., 2011] Duchateau, N., De Craene, M., Piella, G., Silva, E., Doltra, A., Sitges, M., Bijnens, B. H., and Frangi, A. F. (2011). A spatiotemporal statistical atlas of motion for the quantification of abnormal myocardial tissue velocities. *Medical image analysis*, 15(3):316–328.

[Eknoyan, 2008] Eknoyan, G. (2008). Adolphe quetelet (1796-1874)–the average man and indices of obesity. *Nephrol Dial Transplant*, 23(1):47–51.

[Finegold et al., 2012] Finegold, J. A., Asaria, P., and Francis, D. P. (2012). Mortality from ischaemic heart disease by country, region, and age: Statistics from world health organisation and united nations. *International journal of cardiology*.

[Fisher, 1922] Fisher, R. A. (1922). The goodness of fit of regression formulae, and the distribution of regression coefficients. *Journal of the Royal Statistical Society*, pages 597–612.

[Floyd, 1962] Floyd, R. W. (1962). Algorithm 97: shortest path. *Communications of the ACM*, 5(6):345.

[Fonseca et al., 2011] Fonseca, C. G., Backhaus, M., Bluemke, D. A., Britten, R. D., Do Chung, J., Cowan, B. R., Dinov, I. D., Finn, J. P., Hunter, P. J., Kadish, A. H., et al. (2011). The cardiac atlas projectan imaging database for computational modeling and statistical atlases of the heart. *Bioinformatics*, 27(16):2288–2295.

[Foppa et al., 2005] Foppa, M., Duncan, B. B., and Rohde, L. E. (2005). Echocardiography-based left ventricular mass estimation. how should we define hypertrophy? *Cardiovascular ultrasound*, 3(1):17.

[Frangi et al., 2001] Frangi, A. F., Niessen, W. J., and Viergever, M. A. (2001). Three-dimensional modeling for functional analysis of cardiac images, a review. *Medical Imaging, IEEE Transactions on*, 20(1):2–5.

[Freeman and Adelson, 1991] Freeman, W. T. and Adelson, E. H. (1991). The design and use of steerable filters. *IEEE Transactions on Pattern analysis and machine intelligence*, 13(9):891–906.

[Freund and Schapire, 1997] Freund, Y. and Schapire, R. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, pages 119–139.

[Gatenby, 2006] Gatenby, P. (2006). Examination medicine–a guide to physician training. *Journal of Paediatrics and Child Health*, 42(12):828–829.

[Go et al., 2013] Go, A. S., Mozaffarian, D., Roger, V. L., Benjamin, E. J., Berry, J. D., Borden, W. B., Bravata, D. M., Dai, S., Ford, E. S., Fox, C. S., et al. (2013). Heart disease and stroke statistics2013 update a report from the american heart association. *Circulation*, 127(1):e6–e245.

[Grau and Noble, 2005] Grau, V. and Noble, J. A. (2005). Adaptive multiscale ultrasound compounding using phase information. *Med Image Comput Comput Assist Interv*, 8(Pt 1):589–96.

[Harris and Stephens, 1988] Harris, C. and Stephens, M. (1988). A combined corner and edge detector. In *Alvey vision conference*, volume 15, page 50. Manchester, UK.

[Hasegawa et al., 2006] Hasegawa, H., Takano, H., Shindo, S., Takeda, S., Funabashi, N., Nakagawa, K., Toyozaki, T., Kuwabara, Y., and Komuro, I. (2006). Images in cardiovascular medicine. transition from left ventricular hypertrophy to massive fibrosis in the cardiac variant of fabry disease. *Circulation*, 113(16):e720–1.

[Heikkilä et al., 2009] Heikkilä, M., Pietikäinen, M., and Schmid, C. (2009). Description of interest regions with local binary patterns. *Pattern recognition*, 42(3):425–436.

[Heinrich et al., 2011] Heinrich, M. P., Jenkinson, M., Bhushan, M., Matin, T., Gleeson, F. V., Brady, J. M., and Schnabel, J. A. (2011). Non-local shape descriptor: a new similarity metric for deformable multi-modal registration. *Med Image Comput Comput Assist Interv*, 14(Pt 2):541–8.

[Heisele et al., 2003] Heisele, B., Serre, T., Prentice, S., and Poggio, T. (2003). Hierarchical classification and feature reduction for fast face detection with support vector machines. *Pattern Recognition*, pages 2007–2017.

[HighBloodPressUK, 2015] HighBloodPressUK (2015). High blood pressure uk association. http://www.bloodpressureuk.org/Home. [Online; accessed 31-January-2015].

[Ho, 1999] Ho, T. K. (1999). Classification technique using random decision forests. US Patent 5,930,392.

[Hoogendoorn, 2014] Hoogendoorn, C. (2014). A statistical dynamic cardiac atlas for the virtual physiological human: construction and application.

[Horvath, 2013] Horvath, S. (2013). Dna methylation age of human tissues and cell types. *Genome biology*, 14(10):R115.

[Jani and Rajkumar, 2006] Jani, B. and Rajkumar, C. (2006). Ageing and vascular ageing. *Postgraduate medical journal*, 82(968):357–362.

[Jolliffe, 2002] Jolliffe, I. (2002). *Principal component analysis*. Wiley Online Library.

[Joshi and Miller, 2000] Joshi, S. and Miller, M. (2000). Landmark matching via large deformation diffeomorphisms. *IEEE Transactions on Image Processing*, 9(8):1357–1370.

[Kenney et al., 2005] Kenney, C., Zuliani, M., and Manjunath, B. (2005). An axiomatic approach to corner detection. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 191–197. IEEE.

[Kindermann et al., 1980] Kindermann, R., Snell, J. L., et al. (1980). *Markov random fields and their applications*, volume 1. American Mathematical Society Providence, RI.

[Komodakis and Tziritas, 2007] Komodakis, N. and Tziritas, G. (2007). Approximate labeling via graph cuts based on linear programming. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 29(8):1436–1453.

[Komodakis et al., 2008] Komodakis, N., Tziritas, G., and Paragios, N. (2008). Performance vs computational efficiency for optimizing single and dynamic mrfs: Setting the state of the art with primal-dual strategies. *Computer Vision and Image Understanding*, 112(1):14–29.

[Kruskal, 1964] Kruskal, J. B. (1964). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 29(1):1–27.

[Kühl et al., 2004] Kühl, H. P., Schreckenberg, M., Rulands, D., Katoh, M., Schäfer, W., Schummers, G., Bücker, A., Hanrath, P., and Franke, A. (2004). High-resolution transthoracic real-time three-dimensional echocardiography: quantitation of cardiac volumes and function using semi-automatic border detection and comparison with cardiac magnetic resonance imaging. *J Am Coll Cardiol*, 43(11):2083–90.

[Kuranov et al., 2002] Kuranov, A., Rainer Lienhart, R., and Pisarevsky, V. (2002). An empirical analysis of boosting algorithms for rapid objects with an extended set of haar-like features. Technical Report MRL-TR-July02-01, Intel.

[Lakatta, 1990] Lakatta, E. (1990). Changes in cardiovascular function with aging. *European heart journal*, 11(suppl C):22–29.

[Lauer et al., 1992] Lauer, M. S., Anderson, K. M., and Levy, D. (1992). Separate and joint influences of obesity and mild hypertension on left ventricular mass and geometry: the framingham heart study. *J Am Coll Cardiol*, 19(1):130–4.

[Lauterbur et al., 1973] Lauterbur, P. C. et al. (1973). Image formation by induced local interactions: examples employing nuclear magnetic resonance. *Nature*, 242(5394):190–191.

[Lebedev et al., 2014] Lebedev, A., Westman, E., Van Westen, G., Kramberger, M., Lundervold, A., Aarsland, D., Soininen, H., Kłoszewska, I., Mecocci, P., Tsolaki, M., et al. (2014). Random forest ensembles for detection and prediction of alzheimer's disease with a good between-cohort robustness. *NeuroImage: Clinical*, 6:115–125.

[Lee et al., 1996] Lee, S., Wolberg, G., Chwa, K.-Y., and Shin, S. Y. (1996). Image metamorphosis with scattered feature constraints. *Visualization and Computer Graphics, IEEE Transactions on*, 2(4):337–354.

[Leutenegger et al., 2011] Leutenegger, S., Chli, M., and Siegwart, R. Y. (2011). Brisk: Binary robust invariant scalable keypoints. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 2548–2555. IEEE.

[Li and Allinson, 2008] Li, J. and Allinson, N. M. (2008). A comprehensive review of current local features for computer vision. *Neurocomputing*, 71(10):1771–1787.

[Lin and Duncan, 2004] Lin, N. and Duncan, J. S. (2004). Generalized robust point matching using an extended free-form deformation model: application to cardiac images. In *Biomedical Imaging: Nano to Macro, 2004. IEEE International Symposium on*, pages 320–323. IEEE.

[Lin et al., 2003] Lin, N., Papademetris, X., Sinusas, A. J., and Duncan, J. S. (2003). Analysis of left ventricular motion using a general robust point matching algorithm. In *Medical Image Computing and Computer-Assisted Intervention-MICCAI 2003*, pages 556–563. Springer.

[Lombaert et al., 2012] Lombaert, H., Peyrat, J., Croisille, P., Rapacchi, S., Fanton, L., Cheriet, F., Clarysse, P., Magnin, I., Delingette, H., and Ayache, N. (2012). Human atlas of the cardiac fiber architecture: study on a healthy population. *Medical Imaging, IEEE Transactions on*, 31(7):1436–1447.

[Lowe, 1999] Lowe, D. G. (1999). Object recognition from local scale-invariant features. In *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*, volume 2, pages 1150–1157. Ieee.

[Lowe, 2004] Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110.

[Lu et al., 2010] Lu, X., Georgescu, B., Jolly, M., Guehring, J., Young, A., Cowan, B., Littmann, A., and Comaniciu, D. (2010). Cardiac anchoring in MRI through context modeling. In Jiang, T., Navab, N., Pluim, J. P. W., and Viergever, M. A., editors, *MICCAI (1)*, volume 6361 of *Lecture Notes in Computer Science*, pages 383–390. Springer.

[Lu and Jolly, 2013] Lu, X. and Jolly, M.-P. (2013). Discriminative context modeling using auxiliary markers for lv landmark detection from a single mr image. In *Statistical Atlases and Computational Models of the Heart. Imaging and Modelling Challenges*, pages 105–114. Springer.

[Makela et al., 2002] Makela, T., Clarysse, P., Sipila, O., Pauna, N., Pham, Q. C., Katila, T., and Magnin, I. E. (2002). A review of cardiac image registration methods. *Medical Imaging, IEEE Transactions on*, 21(9):1011–1021.

[Manglik et al., 2004] Manglik, T., Axel, L., Pai, W., Kim, D., Dugal, P., Montillo, A., and Qian, Z. (2004). Use of bandpass Gabor filters for enhancing blood-myocardium contrast and filling-in tags in tagged MR images. page 1793.

[Marcus et al., 1999] Marcus, J. T., Götte, M. J., DeWaal, L. K., Stam, M. R., Van der Geest, R. J., Heethaar, R. M., and Van Rossum, A. C. (1999). The influence of through-plane motion on left ventricular volumes measured by magnetic resonance imaging: implications for image acquisition and analysis. *Journal of Cardiovascular Magnetic Resonance*, 1(1):1–6.

[McDonnell, 1981] McDonnell, M. (1981). Box-filtering techniques. *Computer Graphics and Image Processing*, 17(1):65–70.

[McInerney and Terzopoulos, 1996] McInerney, T. and Terzopoulos, D. (1996). Deformable models in medical image analysis: a survey. *Medical image analysis*, 1(2):91–108.

[McMurray and Pfeffer, 2005] McMurray, J. J. V. and Pfeffer, M. A. (2005). Heart failure. *Lancet*, 365(9474):1877–89.

[Metz et al., 2011] Metz, C., Klein, S., Schaap, M., van Walsum, T., and Niessen, W. J. (2011). Nonrigid registration of dynamic medical imaging data using nd+ t b-splines and a groupwise optimization approach. *Medical image analysis*, 15(2):238–249.

[Mika et al., 1998] Mika, S., Schölkopf, B., Smola, A. J., Müller, K.-R., Scholz, M., and Rätsch, G. (1998). Kernel pca and de-noising in feature spaces. In *NIPS*, volume 4, page 7. Citeseer.

[Miksik and Mikolajczyk, 2012] Miksik, O. and Mikolajczyk, K. (2012). Evaluation of local detectors and descriptors for fast feature matching. In *Pattern Recognition (ICPR), 2012 21st International Conference on*, pages 2681–2684. IEEE.

[Mirsky and Parmley, 1973] Mirsky, I. and Parmley, W. W. (1973). Assessment of passive elastic stiffness for isolated heart muscle and the intact heart. *Circulation research*, 33(2):233–243.

[Mitchell, 1997] Mitchell, T. (1997). *Machine Learning*. The McGraw-Hill Companies, Inc.

[MMWR Weekly: January 2, 2009] MMWR Weekly: January 2, . . . (2009). Quickstats: Mean percentage body fat,* by age group and sex — national health and nutrition examination survey, united states, 1999–2004.

[Moravec, 1977] Moravec, H. P. (1977). Towards automatic visual obstacle avoidance. In *International Conference on Artificial Intelligence (5th: 1977: Massachusetts Institute of Technology)*, page 584.

[Morra et al., 2008] Morra, J. H., Tu, Z., Apostolova, L. G., Green, A. E., Toga, A. W., and Thompson, P. M. (2008). Automatic subcortical segmentation using a contextual model. *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2008, Pt I, Proceedings*, 11:194–201.

[national Library of Medicine, 2011] national Library of Medicine, U. (2011). Myocardial perfusion imaging at the us national library of medicine medical subject headings (mesh).

[Nesser et al., 2007] Nesser, H. J., Sugeng, L., Corsi, C., Weinert, L., Niel, J., Ebner, C., Steringer-Mascherbauer, R., Schmidt, F., Schummers, G., Lang, R. M., et al. (2007). Volumetric analysis of regional left ventricular function with real-time three-dimensional echocardiography: validation by magnetic resonance and clinical utility testing. *Heart*, 93(5):572–578.

[NIH, 2014] NIH (2014). http://www.nhlbi.nih.gov, accessed 2- january-2014.

[O'Brien et al., 1995] O'Brien, E., Beevers, D. G., and Marshall, H. J. (1995). *ABC of hypertension*. BMJ Books.

[Ojala et al., 1994] Ojala, T., Pietikainen, M., and Harwood, D. (1994). Performance evaluation of texture measures with classification based on kullback discrimination of distributions. In *Pattern Recognition, 1994. Vol. 1-Conference A: Computer Vision & Image Processing., Proceedings of the 12th IAPR International Conference on*, number 1, pages 582–585.

[Ojala et al., 2002] Ojala, T., Pietikainen, M., and Maenpaa, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(7):971–987.

[Organization, 2013] Organization, W. H. (2013). Fact sheet no 317: cardiovascular diseases (cvds). Technical report.

[Organization et al., 2012] Organization, W. H. et al. (2012). Fact sheet no. 317: cardiovascular diseases (cvds).

[Papageorgiou and Poggio, 1999] Papageorgiou, C. and Poggio, T. (1999). Trainable pedestrian detection. In *ICIP (4)*, pages 35–39.

[Parzen, 1962] Parzen, E. (1962). On estimation of a probability density function and mode. *The annals of mathematical statistics*, pages 1065–1076.

[Pearson, 1901] Pearson, K. (1901). Liii. on lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11):559–572.

[Perperidis et al., 2005] Perperidis, D., Mohiaddin, R. H., and Rueckert, D. (2005). Spatio-temporal free-form registration of cardiac mr image sequences. *Medical image analysis*, 9(5):441–456.

[Peyrat et al., 2010] Peyrat, J.-M., Delingette, H., Sermesant, M., Xu, C., and Ayache, N. (2010). Registration of 4d cardiac ct sequences under trajectory constraints with multichannel diffeomorphic demons. *Medical Imaging, IEEE Transactions on*, 29(7):1351–1368.

[Powell et al., 2008] Powell, S., Magnotta, V. A., Johnson, H., Jammalamadaka, V. K., Andreasen, N. C., and Pierson, R. (2008). Registration and machine learning based automated segmentation of subcortical and cerebellar brain structures. *Neuroimage*, 39(1):238.

[Qian et al., 2007] Qian, T., Huang, R., Metaxas, D., and Axel, L. (2007). A novel tag removal technique for tagged cardiac MRI and its applications. In *Biomedical Imaging: From Nano to Macro, 2007. ISBI 2007. 4th IEEE International Symposium on*, pages 364–367. IEEE.

[Qian et al., 2006] Qian, Z., Metaxas, D. N., and Axel, L. (2006). Boosting and nonparametric based tracking of tagged mri cardiac boundaries. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2006*, pages 636–644. Springer.

[Rahmim and Zaidi, 2008] Rahmim, A. and Zaidi, H. (2008). Pet versus spect: strengths, limitations and challenges. *Nucl Med Commun*, 29(3):193–207.

[Randen and Husoy, 1999] Randen, T. and Husoy, J. H. (1999). Filtering for texture classification: A comparative study. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 21(4):291–310.

[Rangarajan et al., 1997] Rangarajan, A., Chui, H., Mjolsness, E., Pappu, S., Davachi, L., Goldman-Rakic, P., and Duncan, J. (1997). A robust point-matching algorithm for autoradiograph alignment. *Medical Image Analysis*, 1(4):379–398.

[Remme et al., 2005] Remme, E. W., Augenstein, K. F., Young, A. A., and Hunter, P. J. (2005). Parameter distribution models for estimation of population based left ventricular deformation using sparse fiducial markers. *Medical Imaging, IEEE Transactions on*, 24(3):381–388.

[Romero-Corral et al., 2008] Romero-Corral, A., Somers, V. K., Sierra-Johnson, J., Thomas, R. J., Collazo-Clavell, M., Korinek, J., Allison, T. G., Batsis, J., Sert-Kuniyoshi, F., and

Lopez-Jimenez, F. (2008). Accuracy of body mass index in diagnosing obesity in the adult general population. *International journal of obesity*, 32(6):959–966.

[Rosten and Drummond, 2005] Rosten, E. and Drummond, T. (2005). Fusing points and lines for high performance tracking. In *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, volume 2, pages 1508–1515. IEEE.

[Rosten et al., 2010] Rosten, E., Porter, R., and Drummond, T. (2010). Faster and better: A machine learning approach to corner detection. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 32:105–119.

[Rousseau et al., 2011] Rousseau, F., Habas, P. A., and Studholme, C. (2011). A supervised patch-based approach for human brain labeling. *IEEE Trans Med Imaging*, 30(10):1852–62.

[Roweis and Saul, 2000] Roweis, S. T. and Saul, L. K. (2000). Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500):2323–2326.

[Rublee et al., 2011] Rublee, E., Rabaud, V., Konolige, K., and Bradski, G. (2011). Orb: an efficient alternative to sift or surf. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 2564–2571. IEEE.

[Rueckert et al., 1999] Rueckert, D., Sonoda, L., Hayes, C., and et al. (1999). Nonrigid registration using free-form deformations: Application to breast MR images. *IEEE Tansactions on Medical Imaging*, pages 712–721.

[Rutz et al., 2008] Rutz, A. K., Ryf, S., Plein, S., Boesiger, P., and Kozerke, S. (2008). Accelerated whole-heart 3d cspamm for myocardial motion quantification. *Magn Reson Med*, 59(4):755–63.

[Ryf et al., 2002] Ryf, S., Spiegel, M. A., Gerber, M., and Boesiger, P. (2002). Myocardial tagging with 3d-cspamm. *Journal of Magnetic Resonance Imaging*, 16(3):320–325.

[S Wack and D Badgaiyan, 2011] S Wack, D. and D Badgaiyan, R. (2011). Complex singular value decomposition based noise reduction of dynamic pet images. *Current Medical Imaging Reviews*, 7(2):113–117.

[Samuel, 1959] Samuel, A. L. (1959). Some studies in machine learning using the game of checkers. *IBM Journal of research and development*, 3(3):210–229.

[Sandstede et al., 2000] Sandstede, J., Lipke, C., Beer, M., Hofmann, S., Pabst, T., Kenn, W., Neubauer, S., and Hahn, D. (2000). Age-and gender-specific differences in left and right ventricular cardiac function and mass determined by cine magnetic resonance imaging. *European radiology*, 10(3):438–442.

[Savio et al., 2009] Savio, A., García-Sebastián, M., Graña, M., and Villanúa, J. (2009). Results of an adaboost approach on alzheimer's disease detection on mri. In Mira, J. M., Ferrández, J. M., Álvarez, J. R., de la Paz, F., and Toledo, F. J., editors, *IWINAC (2)*, volume 5602 of *Lecture Notes in Computer Science*, pages 114–123. Springer.

[Schiller et al., 1988] Schiller, N. B., Shah, P., Crawford, M., DeMaria, A., Devereux, R., Feigenbaum, H., Gutgesell, H., Reichek, N., Sahn, D., and Schnittger, I. (1988). Recommendations for quantitation of the left ventricle by two-dimensional echocardiography. american society of echocardiography committee on standards, subcommittee on quantitation of two-dimensional echocardiograms. *Journal of the American Society of Echocardiography: official publication of the American Society of Echocardiography*, 2(5):358–367.

[Schlosser et al., 2005] Schlosser, T., Pagonidis, K., Herborn, C. U., Hunold, P., Waltering, K.-U., Lauenstein, T. C., and Barkhausen, J. (2005). Assessment of left ventricular parameters using 16-mdct and new software for endocardial and epicardial border delineation. *American Journal of Roentgenology*, 184(3):765–773.

[Schutz et al., 2002] Schutz, Y., Kyle, U., and Pichard, C. (2002). Fat-free mass index and fat mass index percentiles in caucasians aged 18-98 y. *International journal of obesity and related metabolic disorders: journal of the International Association for the Study of Obesity*, 26(7):953–960.

[Scientific, 2014] Scientific, B. (2014). [http://www.bostonscientific.com; accessed 2-January-2014].

[Shannon, 1949] Shannon, C. E. (1949). Communication in the presence of noise. *Proceedings of the IRE*, 37(1):10–21.

[Shen and Davatzikos, 2002] Shen, D. and Davatzikos, C. (2002). Hammer: hierarchical attribute matching mechanism for elastic registration. *IEEE Tansactions on Medical Imaging*, 21(11):1421–1439.

[Shi et al., 2011] Shi, W., Zhuang, X., et al. (2011). Automatic segmentation of different pathologies from cardiac cine mri using registration and multiple component em estimation. In *Functional Imaging and Modeling of the Heart*, volume 6666 of *Lecture Notes in Computer Science*, pages 163–170.

[Shi et al., 2012] Shi, W., Zhuang, X., Wang, H., Duckett, S., Luong, D. V., Tobon-Gomez, C., Tung, K., Edwards, P. J., Rhode, K. S., Razavi, R. S., et al. (2012). A comprehensive cardiac motion estimation framework using both untagged and 3-d tagged mr images based on nonrigid registration. *Medical Imaging, IEEE Transactions on*, 31(6):1263–1275.

[Singer and Kite, 2008] Singer, D. R. and Kite, A. (2008). Management of hypertension in peripheral arterial disease: does the choice of drugs matter? *European Journal of Vascular and Endovascular Surgery*, 35(6):701–708.

[Slifstein, 2014] Slifstein, M. (2014). Positron emission tomography (pet) imaging. In *Encyclopedia of Psychopharmacology*, pages 1–11. Springer.

[Smith and Brady, 1997] Smith, S. M. and Brady, J. M. (1997). Susana new approach to low level image processing. *International journal of computer vision*, 23(1):45–78.

[Smola and Schölkopf, 2004] Smola, A. J. and Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and computing*, 14(3):199–222.

[Stelfox et al., 2006] Stelfox, H. T., Ahmed, S. B., Ribeiro, R. A., Gettings, E. M., Pomerantsev, E., and Schmidt, U. (2006). Hemodynamic monitoring in obese patients: The impact of body mass index on cardiac output and stroke volume*. *Critical care medicine*, 34(4):1243–1246.

[Studholme et al., 1995] Studholme, C., Hill, D. L., and Hawkes, D. J. (1995). Multiresolution voxel similarity measures for mr-pet registration. In *Information processing in medical imaging*, volume 252. Dordrecht, The Netherlands: Kluwer.

[Studholme et al., 1997] Studholme, C., Hill, D. L., and Hawkes, D. J. (1997). Automated three-dimensional registration of magnetic resonance and positron emission tomography brain images by multiresolution optimization of voxel similarity measures. *Medical physics*, 24:25.

[Studholme et al., 1999] Studholme, C., Hill, D. L., and Hawkes, D. J. (1999). An overlap invariant entropy measure of 3d medical image alignment. *Pattern recognition*, 32(1):71–86.

[Sugihara et al., 2005] Sugihara, H., Yonekura, Y., Matsumoto, T., and Sasaki, Y. (2005). Relationship between asynchronous myocardial contraction and left ventricular systolic and diastolic function–assessment using the ecg-gated polar map with 99mtc-methoxy-isobutyl isonitrile. *Circ J*, 69(2):183–7.

[Sundar et al., 2009] Sundar, H., Litt, H., and Shen, D. (2009). Estimating myocardial motion by 4d image warping. *Pattern recognition*, 42(11):2514–2526.

[Tabachnick et al., 2001] Tabachnick, B. G., Fidell, L. S., et al. (2001). Using multivariate statistics.

[Taimouri and Hua, 2013] Taimouri, V. and Hua, J. (2013). Visualization of shape motions in shape space. *Visualization and Computer Graphics, IEEE Transactions on*, 19(12):2644–2652.

[Tenenbaum et al., 2000] Tenenbaum, J. B., De Silva, V., and Langford, J. C. (2000). A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323.

[Thibodeau et al., 2004] Thibodeau, G. A., Patton, K. T., and Wills (2004). *Structure & function of the body*. Mosby.

[Tomasi and Shi, 1994] Tomasi, C. and Shi, J. (1994). Good features to track. *CVPR94*, pages 593–600.

[Triggs, 2004] Triggs, B. (2004). Detecting keypoints with stable position, orientation, and scale under illumination changes. In *Computer Vision-ECCV 2004*, pages 100–113. Springer.

[Turkbey et al., 2010] Turkbey, E. B., McClelland, R. L., Kronmal, R. A., Burke, G. L., Bild, D. E., Tracy, R. P., Arai, A. E., Lima, J. A. C., and Bluemke, D. A. (2010). The impact of obesity on the left ventricle: the multi-ethnic study of atherosclerosis (mesa). *JACC Cardiovasc Imaging*, 3(3):266–74.

[Tuytelaars and Mikolajczyk, 2008] Tuytelaars, T. and Mikolajczyk, K. (2008). Local invariant feature detectors: a survey. *Foundations and Trends® in Computer Graphics and Vision*, 3(3):177–280.

[Uecker et al., 2010] Uecker, M., Zhang, S., Voit, D., Karaus, A., Merboldt, K.-D., and Frahm, J. (2010). Real-time mri at a resolution of 20 ms. *NMR in Biomedicine*, 23(8):986–994.

[van der Maaten et al., 2009] van der Maaten, L. J., Postma, E. O., and van den Herik, H. J. (2009). Dimensionality reduction: A comparative review. *Journal of Machine Learning Research*, 10(1-41):66–71.

[Vapnik, 1995] Vapnik, V. (1995). *The Nature of Statistical Learning Theory*. Springer, 1st ed. 1995. corr. 2nd printing edition.

[Vapnik, 1998] Vapnik, V. (1998). *Statistical learning theory*. Wiley, 1 edition.

[Vercauteren et al., 2009] Vercauteren, T., Pennec, X., Perchant, A., and Ayache, N. (2009). Diffeomorphic demons: Efficient non-parametric image registration. *NeuroImage*, 45(1):S61–S72.

[Veress et al., 2005] Veress, A. I., Gullberg, G. T., and Weiss, J. A. (2005). Measurement of strain in the left ventricle during diastole with cine-mri and deformable image registration. *Journal of biomechanical engineering*, 127(7):1195–1207.

[Veress et al., 2013] Veress, A. I., Klein, G., and Gullberg, G. T. (2013). A comparison of hyperelastic warping of pet images with tagged mri for the analysis of cardiac deformation. *Journal of Biomedical Imaging*, 2013:9.

[Viola and Jones, 2002] Viola, P. and Jones, M. (2002). Robust real-time object detection. *International Journal of Computer Vision*, 57(2):137–154.

[Viola and Wells III, 1997] Viola, P. and Wells III, W. M. (1997). Alignment by maximization of mutual information. *International journal of computer vision*, 24(2):137–154.

[Von Luxburg, 2007] Von Luxburg, U. (2007). A tutorial on spectral clustering. *Statistics and computing*, 17(4):395–416.

[Wachinger et al., 2010] Wachinger, C., Yigitsoy, M., and Navab, N. (2010). Manifold learning for image-based breathing gating with application to 4d ultrasound. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2010*, pages 26–33. Springer.

[Wang and Mahadevan, 2008] Wang, C. and Mahadevan, S. (2008). Manifold alignment using procrustes analysis. In *Proceedings of the 25th international conference on Machine learning*, pages 1120–1127. ACM.

[Wang and Amini, 2012] Wang, H. and Amini, A. A. (2012). Cardiac motion and deformation recovery from mri: a review. *Medical Imaging, IEEE Transactions on*, 31(2):487–503.

[Wang et al., 2015] Wang, X., Chen, T., Zhang, S., Schaerer, J., Qian, Z., Huh, S., Metaxas, D., and Axel, L. (2015). Meshless deformable models for 3d cardiac motion and strain analysis from tagged mri. *Magnetic resonance imaging*, 33(1):146–160.

[Whisnant, 1996] Whisnant, J. P. (1996). Effectiveness versus efficacy of treatment of hypertension for stroke prevention. *Neurology*, 46(2):301–307.

[WHO Expert Consultation, 2004] WHO Expert Consultation (2004). Appropriate body-mass index for asian populations and its implications for policy and intervention strategies. *Lancet*, 363(9403):157–63.

[Wittman, 2005] Wittman, T. (2005). Manifold learning techniques: So which is the best? Geometric Data Analysis, University of Minnesota.

[Wolz et al., 2010] Wolz, R., Aljabar, P., Hajnal, J., Hammers, A., Rueckert, D., and Initi, A. D. N. (2010). Leap: Learning embeddings for atlas propagation. *Neuroimage*, pages 1316–1325.

[Yan et al., 2007] Yan, P., Sinusas, A., and Duncan, J. S. (2007). Boundary element method-based regularization for recovering of lv deformation. *Medical image analysis*, 11(6):540–554.

[Yang et al., 2011] Yang, L., Georgescu, B., Zheng, Y., Wang, Y., Meer, P., and Comaniciu, D. (2011). Prediction based collaborative trackers (pct): A robust and accurate approach toward 3d medical object tracking. *Medical Imaging, IEEE Transactions on*, 30(11):1921–1932.

[Yigitsoy et al., 2011] Yigitsoy, M., Wachinger, C., and Navab, N. (2011). Temporal groupwise registration for motion modeling. In *Information Processing in Medical Imaging*, pages 648–659. Springer.

[Yu et al., 2014] Yu, Y., Zhang, S., Li, K., Metaxas, D., and Axel, L. (2014). Deformable models with sparsity constraints for cardiac motion analysis. *Medical image analysis*, 18(6):927–937.

[Zheng et al., 1999] Zheng, Z., Wang, H., and Khwang Teoh, E. (1999). Analysis of gray level corner detection. *Pattern Recognition Letters*, 20(2):149–162.

[Zhu et al., 2010] Zhu, Y., Papademetris, X., Sinusas, A., and Duncan, J. (2010). Segmentation of the left ventricle from cardiac mr images using a subject-specific dynamical model. *IEEE Trans Med Imaging*, 29(3):669–87.