Dynamic Analysis of X-ray Angiography for Image-Guided Coronary Interventions

Hua Ma

Colophon

The research described in this thesis was carried out at the Department of Radiology & Nuclear Medicine, Erasmus MC, University Medical Center Rotterdam (the Netherlands). This work was supported by the IMAGIC project under the iMIT program of NWO-domein TTW (grant number 12703).

This work was carried out in the ASCI graduate school.
ASCI dissertation series number 410.

Financial support by the Dutch Heart Foundation for the publication of this thesis is gratefully acknowledged.

Additional financial support for the publication of this thesis was kindly provided by

- the Department of Radiology & Nuclear Medicine, Erasmus MC,
- the ASCI graduate school,
- Medis Medical Imaging Systems bv,
- Pie Medical Imaging BV.

Cover design by Hua Ma.

ISBN: 978-94-6375-752-2

Printed by Ridderprint | www.ridderprint.nl

Dynamic Analysis of X-ray Angiography for Image-Guided Coronary Interventions

Analyse van X-ray Angiografische Beeldreeksen voor Beeldgeleide Interventies aan de Kransslagaders

Thesis

to obtain the degree of Doctor from the
Erasmus University Rotterdam
by command of the
rector magnificus

Prof.dr. R.C.M.E. Engels

and in accordance with the decision of the Doctorate Board.

The public defence shall be held on
Wednesday, 12 February 2020 at 11:30 hrs

by

Hua Ma

born in Wuhan, China

Erasmus University Rotterdam

Doctoral Committee

Promotor:	**Prof.dr. W.J. Niessen**
Other members:	**Prof.dr.ir. A.F.W. van der Steen**
	Dr.ir. J. Dijkstra
	Dr. E.S. Regar
Copromotor:	**Dr.ir. T. van Walsum**

Contents

Chapter One

Introduction

1.1 Coronary Arteries: Anatomy and Functions

The heart is the central organ of the human circulation system. It pumps blood through the blood vessels to provide the human body with oxygen and nutrients, and it carries away metabolic waste. The heart muscles receive their own supply of blood via the coronary arteries. These vessels branch off from the aorta near the point where the aorta and the left ventricle meet [7] (Fig. 1.1a). The coronary arteries wrap around the surface of the heart, with small branches entering into the heart muscle to supply it with blood [2]. Thus, the coronary arteries play a significant role to the heart, and so to speak, to human life.

Coronary arteries have two main branches: the left main and right coronary artery, each of which further divides into smaller branches (Fig. 1.1a).

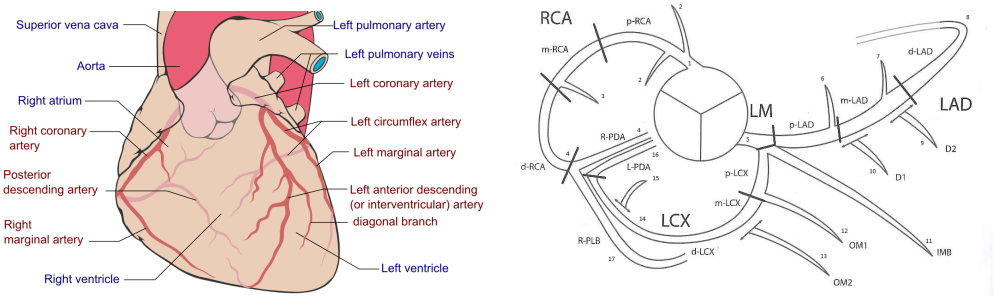
The left main coronary artery (LM or LCA) supplies blood to the left ventricle and left atrium. It divides into two major branches (Fig. 1.1a): the left anterior descending artery (LAD) and the left circumflex artery (LCX). The LAD and its diagonal branches supply blood to the front and the left side of the heart, mainly the anterior ventricular septum and the major part of the anterior portion of the left ventricle. The LCX encircles the heart muscle and provides blood to the left atrium and the posterior-lateral aspect of the left ventricle that are mainly at the outer side and back of the heart [2, 75].

The right coronary artery (RCA) supplies blood to the right ventricle, the right atrium, and the SA (sinoatrial) and AV (atrioventricular) nodes. The RCA divides into smaller branches (Fig. 1.1a), such as the posterior descending artery (PDA) and the acute marginal artery, providing blood to the inferior part of the heart and the lateral portion of the right ventricle, respectively [75]. The RCA also supplies blood to the septum of the heart, together with the LAD.

A brief schematic view summarizing all major and small branches of the coronary arteries is illustrated in Fig. 1.1b, where the relative positions of coronary artery branches and aorta are shown.

1.2 Coronary Artery Disease

Coronary artery disease, also known as *ischemic heart disease* or *atherosclerosis*, is one of the leading causes of death in the world [88]. It involves a reduction of blood supply to the heart muscle due to the build-up of plaque in the coronary arteries. Plaque



(a) (b)

Figure 1.1: The anatomy of coronary arteries: (a) Coronary arteries in the anterior view of the heart [64] (labeled in red text). (b) the branches of the coronary arteries (the figure is adapted from [57]).

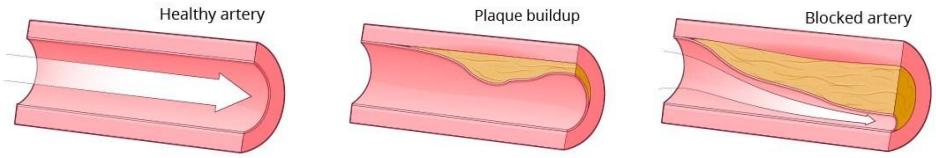


Figure 1.2: The formation of atherosclerosis caused by plaque built up inside blood vessels. (the figure is adapted from [86])

mainly consists of fat, cholesterol or calcium. Accumulation of these substances at the plaque narrows the lumen of coronary arteries, reducing the amount of blood flowing to the heart muscle. A rupture of the plaque may also cause the formation of blood clots in the coronary arteries that can lead to partially or complete obstruction of the vessel lumen. The consequence of a narrowed or blocked vessel lumen is that insufficient oxygen-rich blood can reach the heart muscles, which can cause angina or heart attack and may lead to heart failure or arrhythmias in long term [6]. Fig. 1.2 illustrates the formation of atherosclerosis in the vessel lumen.

1.3 Treatment of Coronary Artery Disease

The treatment of coronary artery disease includes medication and medical procedures. Drugs, such as cholesterol-modifying medications, may alleviate the symptoms of coronary artery disease by decreasing the amount of cholesterol in the blood, one of the major component materials of the plaque [5]. Furthermore, a patient suffering from severe cases may undergo medical procedures: a percutaneous coronary intervention or coronary artery bypass grafting.

Percutaneous coronary intervention (PCI), also known as *Coronary Angioplasty*, is a minimally-invasive procedure to treat narrowing of the coronary arteries. It was

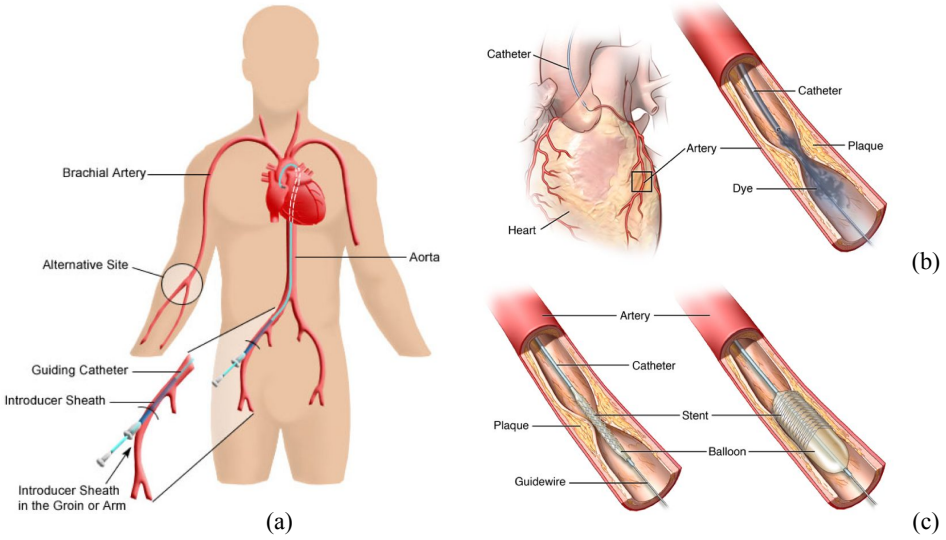


Figure 1.3: The PCI procedure: (a) insertion sites of guiding catheter, (b) injection of contrast agent through the guiding catheter, (c) balloon and stent expansion to widen the vessel lumen. (the figure is adapted from [1, 4])

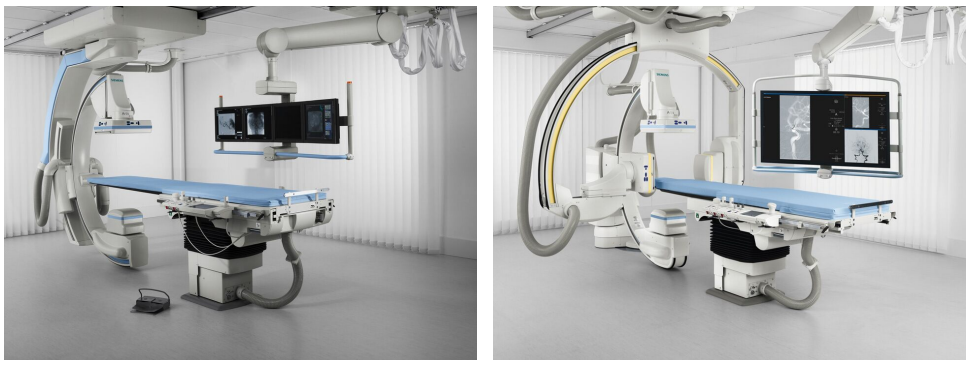
first introduced by Andreas Grünzig in Zurich, Switzerland in 1977 and has become worldwide-adopted since 1980s [73]. During this procedure, the stenosed vessel area is widened by the inflation of a balloon and a stent that are introduced through a long and thin catheter inserted in the blood vessel from a small skin incision.

Coronary artery bypass grafting (CABG) is used for more severe cases that are difficult to treat with PCI, e.g. coronary arteries with multiple stenosed sites [5]. In this procedure, a surgeon creates a graft to bypass the blocked vessel using arteries or veins from other parts of the body. Compared to PCI, CABG is superior for patients with multivessel disease [44], yet it is more invasive, as it requires to open the chest in order to reach the heart.

The target application of this thesis is the PCI procedure. Its minimally-invasive nature puts patients under lower risk, especially for patients in very old and very young age [97].

1.3.1 Percutaneous Coronary Intervention

Fig. 1.3 shows an overview of the PCI procedure. At first, a guiding catheter is inserted in the blood vessel, e.g. via the groin or arm. The catheter is then manoeuvred through the aorta or brachial artery (depending on the insertion site) towards the ostium of the coronary arteries. Through the catheter, X-ray opaque contrast agent can be injected to identify the lesions in the vessels, and a guidewire is manipulated by the interventional cardiologist to move to the stenosed site. Once the guidewire reaches the site of the lesion, a balloon catheter carrying a stent, a collapsed wire mesh tube,



(a) Monoplane **(b)** Biplane

Figure 1.4: Typical X-ray imaging equipment in a catheterization laboratory in a hospital with a Siemens Artis zee ceiling monoplane system (a) and biplane system (b). (images copyright: Siemens Healthineers AG [3])

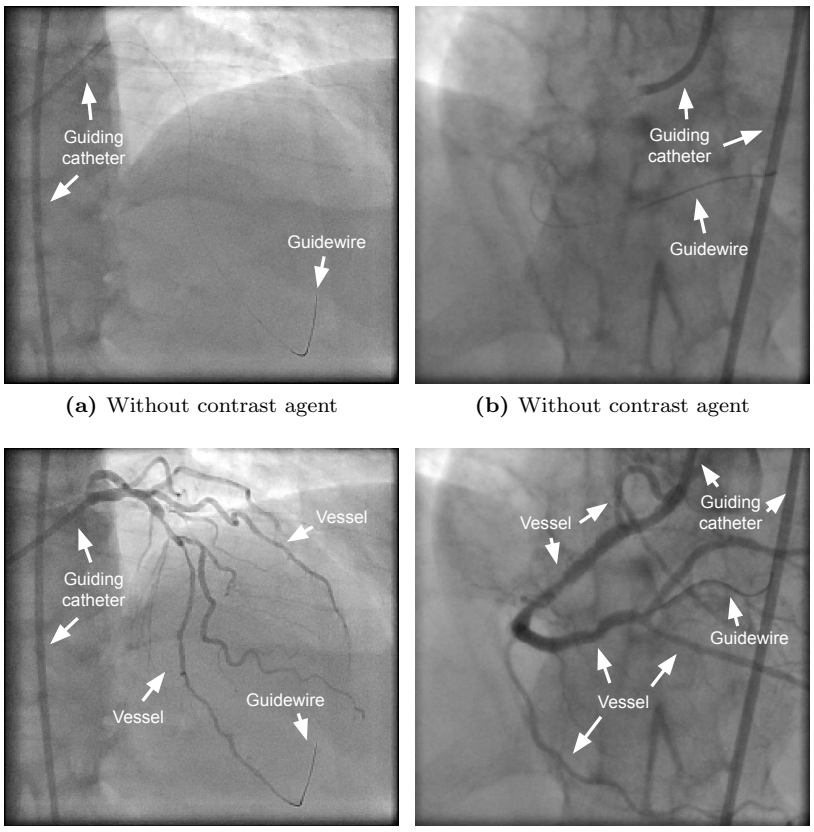
is advanced over the guidewire towards the lesion site. The balloon is then inflated to expand the vessel lumen as well as the stent. Finally, the stent is deployed at the lesion to prevent the vessel from collapsing, and the balloon is retrieved.

1.3.2 X-ray Angiography for Image Guidance

The PCI procedure, as suggested by its name, is performed percutaneously, which means that the interventional cardiologist does not directly see the coronary arteries, vessel lesions and interventional tools in the patient's body during the procedure. In a catheterization laboratory (the operation room where PCI is performed), X-ray angiography (XA) is the imaging technique that is commonly used during PCI for visual guidance of the procedure. The complete set-up is normally an integrated system that typically contains one or two C-arms, a patient table and monitors (Fig. 1.4). The C-arm is a C-shape equipment connecting an X-ray source with an X-ray detector that is mainly used for fluoroscopic imaging, and can be rotated during the procedure to acquire X-ray images from different angles. The acquired images are shown on the monitors together with other relevant patient information (e.g. ECG) to provide real-time visual feedback to the interventional cardiologist during PCI.

Fig. 1.5 illustrates two examples of X-ray angiographic images of left and right coronary arteries (Fig. 1.5c and 1.5d). The vessels are only visible with the use of contrast agent, showing as "tree" structures in the images. When no contrast agent is injected, the vessels are not visible in the X-ray fluoroscopic images, while the guidewire has good visibility (Fig. 1.5a and 1.5b). The real X-ray angiographic images used for PCI are dynamic cine-angiograms with a frame rate ranging between 4 and 30 frames per second (fps). In typical angiograms, moving structures, such as diaphragm and vessels, and transition between contrast and non-contrast phase can be seen.

During PCI, interventional cardiologists use XA images to view the target arter-



(a) Without contrast agent **(b)** Without contrast agent

(c) Left coronary arteries **(d)** Right coronary arteries

Figure 1.5: Examples of X-ray images for PCI. In (a) and (b), contrast agent has not been injected, guiding catheter (the long, dark, thin tube) and guidewire (the thin wire) can be seen in the images. In (c) and (d), contrast agent is used for visualizing the left (c) and the right (d) coronary arteries.

ies, and to find and assess the lesions on the vessels. Since guidewires and stents have better visibility in fluoroscopic images, cardiologists typically manipulate the instruments without continuously injecting contrast agent. The manipulation of instruments therefore mainly relies on the operator's mental map of the vessels and plaques' location from the previous XA images with contrast agent.

1.4 Challenges

1.4.1 Challenges of Image Guidance for PCI

One of the challenges in PCI is that limited visual feedback is provided to the interventional cardiologist when manoeuvring guidewires towards the lesion site. As the navigation of instruments is guided with "vessel-free" fluoroscopic images, the operator needs to mentally reconstruct the vessels from the previous angiographic acquisitions, which relies largely on the skills and experiences of the operator, as it is, in general, a challenging task to accurately imagine the position of moving structures.

To be more certain about the target location, cardiologists sometimes repeatedly inject contrast agent to get a better mental image of the vessels and the lesions, as the opacification of the vessels lasts only for a short period before contrast agent drains from the vessel lumen. However, contrast-induced side effects limit the amount of contrast agent that can be used on patients during PCI. Allergic reaction and nephrotoxicity have been reported as side effects that X-ray contrast agent may have [10, 105]. Especially the latter one, also known as contrast-induced nephropathy (CIN), which may result in chronical renal failure with all its relevant sequelae [105], has been associated to contrast volume [104].

Another challenge in performing PCI comes from the side effect of exposure to X-ray. The ionizing radiation of X-ray has detrimental effect on the exposed human tissues, including tissue reactions and increased risk for stochastic events, such as skin necrosis and radiation-induced cancer [58]. To spare the radiation dose to patients and operators, strategies, such as minimizing the fluoroscopy acquisition time, modulating the fluoroscopy dose per frame, decreasing the radiation detector magnification and reducing the frame rate, should be applied [40, 58]. As some of these strategies may hamper the quality of the acquired X-ray images, operators face a trade-off between a lower X-ray dose and abilities to resolve small vessels or motion details.

The above listed challenges of image guidance for PCI procedure could be addressed with an improved system that can help interventional cardiologists to view relevant information that is needed for PCI, and that meanwhile controls the potential risks the operators and the patients may have from the procedure. This thesis describes the works done during my PhD in order to address the challenges from the perspective of automatic image analysis approaches.

1.4.2 Challenges of Image Analysis for X-ray Angiograms

In addition to the challenges from the image guidance point of view, the physical nature of X-ray angiographic images also poses challenges for developing automatic and robust algorithms for X-ray image analysis. The major challenges are four-fold:

- X-ray images are the results of projection of 3D structures on a 2D plane, therefore, the 3D information is lost during the image formation. The information of anatomical structures overlaps each other in X-ray images, making automatic analysis of those structures challenging.

- Different from X-ray projection radiography which is normally acquired in the anterior-posterior direction, X-ray angiography can be taken from arbitrary viewing directions, depending on the C-arm configuration. This means that X-ray images acquired from different view angles based on the same 3D structures look different. Therefore, the image analysis approaches for XA need to be robust for images acquired with different C-arm angles.

- An X-ray angiogram is a cinematic clip instead of a static image, the structures of interest in the image sequence normally moves instead of staying still, such as diaphragm and coronary arteries. As the motion may be caused by the patient's respiration or heartbeat, the moving structures may present respiratory or cardiac motion patterns. These two types of motions normally need to be taken into consideration for image guidance applications.

- The level of contrast agent in an X-ray angiogram does not stay constant. After the contrast injection, the amount of contrast in the field of view increases rapidly till its maximum, the vessels becomes fully opacified at this phase. After a short period, the contrast agent drains gradually from the target vessels along with the blood flow, the vessels becomes invisible again. These changes between the contrast and the non-contrast phase in an X-ray sequence require additional designs for some image guidance applications.

1.5 This Thesis

1.5.1 Dynamic Analysis of X-ray Angiograms

Despite of the previously mentioned challenges, opportunities may also come from the particular properties of X-ray angiographic image data. Different from many other medical imaging modalities which are static snapshots of anatomical structures, the cinematic nature of X-ray angiograms possesses a time dimension in the data, bringing the possibility of using the temporal information in the image analysis. The temporal information contains the changes of the same tissues at different time points, which may serve as a cue to link different frames instead of treating each image independently. For some problems, the temporal information is the key to overcome the obstacles.

In this thesis, I use the term *dynamic analysis* to call the image analysis approaches that take advantage of the temporal, motion or inter-frame information in X-ray angiograms. Different dynamic analysis approaches will be introduced in this book to address the challenges in various aspects of the image guidance for PCI.

1.5.2 Purpose and Chapter Organization

The purpose of the research presented in this thesis is to develop and evaluate dynamic image analysis techniques that may improve image guidance for percutaneous coronary interventions. To this end, this thesis investigates on the following approaches:

1. **Layer separation, a computational angiography approach** Layer separation is a computational technique to improve vessel visibility in XA images by removing static and moving background structures. In Chapter 2, a layer separation approach using robust PCA was proposed to separate an XA angiographic sequence into three additive layers: a vessel layer, a breathing layer and a quasi-static background layer. In Chapter 3, we developed and evaluated an online layer separation method which dynamically separates streaming XA data into the three layers and have shown its potential on reducing the amount of contrast agent used for PCI.

2. **Layer separation as a component for XA analysis** Layer separation enables independent analysis of the layers it outputs, and can serve as a component in the image processing pipeline for various applications. Two example are presented in this thesis. In Chapter 4, a PCA-based approach was proposed to extract respiratory motion surrogate from the breathing layer that has high correlation with the respiration movement. In Chapter 5, we adopted the vessel layer for automatic detection of contrast inflow in an XA sequence. An approach using a recurrent neural network (RNN) was proposed to classify whether an X-ray image is with contrast or not using features extracted from the enhanced vessel layer. Additionally, we also proposed a second method based on a convolutional neural network (CNN) to perform the frame classification.

3. **Dynamic roadmapping, an augmented fluoroscopy approach** Interventional tools are typically navigated in fluoroscopy mode with non-contrast-enhanced X-ray images, which forces the operator to rely on a mental reconstruction of anatomical structures. In Chapter 6, we developed and evaluated a novel dynamic coronary roadmapping approach to tackle the challenge. The fluoroscopic images are augmented with a dynamic motion-compensated vessel layer to provide real-time visual guidance during PCI, while in the meantime, reducing the use of contrast agent.

In Chapter 7, the thesis concludes with a summary on the methods and result, and a discussion on future research directions.

Chapter Two

Layer Separation for Vessel Enhancement in Interventional X-ray Angiograms Using Morphological Filtering and Robust PCA

Abstract — Automatic vessel extraction from X-ray angiograms (XA) for percutaneous coronary interventions is often hampered by low contrast and presence of background structures, e.g. diaphragm, guiding catheters, stitches. In this paper, we present a novel layer separation technique for vessel enhancement in XA to address this problem. The method uses morphological filtering and Robust PCA to separate interventional XA images into three layers, i.e. a large-scale breathing structure layer, a quasi-static background layer and a layer containing the vessel structures that could potentially improve the quality of vessel extraction from XA. The method is evaluated on several clinical XA sequences. The result shows that the proposed method significantly increases the visibility of vessels in XA and outperforms other background-removal methods.

Based upon: **H. Ma**, G. Dibildox, J. Banerjee, W.J. Niessen, C. Schultz, E. Regar and T. van Walsum: Layer Separation for Vessel Enhancement in Interventional X-ray Angiograms Using Morphological Filtering and Robust PCA. Workshop on Augmented Environments for Computer-Assisted Interventions (AE-CAI 2015), Lecture Notes in Computer Science, vol. 9365, pp. 104-113, 2015.

2.1 Introduction

Percutaneous coronary intervention (PCI) is a minimally invasive procedure for treating patients with advanced coronary artery disease. It is usually performed under guidance of X-ray angiograms (XA) where coronary arteries are opacified with contrast agent. Automatic processing of XA images, e.g. vessel extraction of coronary arteries, may serve as a basis for further processing, such as coronary motion analysis [78] and pre/intra-operative information fusion [14].

Hessian-based vessel enhancement filtering, e.g. Frangi vesselness filter [38], is commonly used for extraction of vessels in medical images. Applying such filters directly on interventional XA, however, often also enhances non-vascular structures, such as catheter segments and vertebral contours, due to their tubular or curvilinear structural appearances.

Related works have reported on methods to remove non-vessel structures or improve the visibility of vessels in XA images. In [15], a method that subtracts the median frame was used for removing static structures in XA, such as vertebral bodies. Schneider et al. [93] proposed a post-processing technique on vesselness images that combines a local probability map with local directional vessel information for artifact reduction and catheter removal. Layer separation methods provide an alternative way of vessel enhancement. In [116], a multi-scale framework was developed to separate XA images into three layers based on different motion patterns such that coronary arteries are better visible in the fast motion layer. This method involves human-interactions to label corresponding control points in XA images for motion field estimation. In another study [118], a Bayesian framework was developed that combines dense motion estimation, uncertainty propagation and statistical fusion to achieve motion layer separation. Both layer separation methods require to compute motion field. Robust principal component analysis (Robust PCA) is a data decomposition technique that has e.g. been used for background modeling from surveillance video in [23]. In [43], Robust PCA was adopted for registration of DCE MR time series.

In this paper, we propose an automatic method to robustly separate foreground (contrast-enhanced vessels, guiding catheter tip) from (quasi) static background, such as vertebral bodies and guiding catheters in the aorta, while ignoring large-scale motion such as diaphragm movement. Our contributions are three folds: 1) the development of a Robust PCA based layer separation method that does not require computation of the motion field; 2) qualitative and quantitative evaluations on four clinical XA sequences; 3) comparison to other related background-removal approaches.

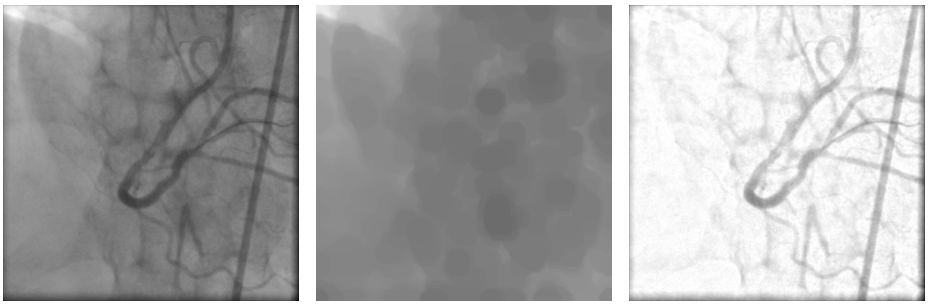
2.2 Method

The method enhances vessels in XA images by separating an image into three layers, i.e. a large-scale breathing layer, a quasi-static background layer and a foreground layer containing the vessels. To this end, our proposed method consists of two steps: first, separation and removal of large-scale breathing structures, such as diaphragm, from the original images, using morphological closing; second, separation of a quasi-

static background from the moving structures using Robust PCA. In the remainder of this section, we describe both steps in more details, followed by the integrated layer separation.

2.2.1 Separation of Breathing Structures

To obtain a separate layer containing large-scale structures, we remove small objects from the original image, including guiding catheters, guide wires, stitches and vertebral bodies. Similar to the approach in [66] (Chapter 4 of this thesis), we apply morphological closing to the image with a circular structuring element of 8.5 *mm* in diameter. Pilot experiments indicated that this size was adequate for a complete removal of vessels and guiding catheters from our images while not causing too much circular artifacts. An example of a resulting image is shown in Fig. 2.1b. Compared to the original image, the guiding catheter and coronary arteries are removed and vertebral contours are blurred, while structures that presents respiratory motion, such as the diaphragm and lung tissue, remain in the image (white area in the upper left part of the image). The resulting image that contains large scale structures which exhibit respiratory motion is called the breathing layer, and will later be subtracted from the original image to obtain the difference image (DI, Fig. 2.1c) of an XA frame for further processing.



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

Figure 2.1: Morphological closing operation on an XA image: (a) the original image, (b) image processed with morphological closing, (c) the difference image (DI) of (a) and (b).

2.2.2 Background Separation Using Robust PCA

Robust PCA decomposes a data matrix into two different sources: a low-rank matrix and a sparse matrix. Suppose that M is an $m \times n$ matrix to be decomposed, which contains n observations of m dimensional data in its columns. Robust PCA is formulated as the following optimization problem [23]:

$$\begin{aligned} \text{minimize} \quad & \|L\|_* + \lambda \|S\|_1 \\ \text{subject to} \quad & L + S = M \end{aligned} \tag{2.1}$$

where L is a low-rank matrix and S is a sparse matrix of the same size as M. $\|L\|_*$ denotes the nuclear norm of L and $\|S\|_1$ is the L_1 norm of S. λ is the tuning parameter of regularization. Source decomposition is achieved by solving this optimization problem. In this work, we use inexact Augmented Lagrange Multiplier (ALM) method [63] to solve the problem. Robust PCA can be applied for separation of the background layer of DI from the vessel layer. The background of an XA sequence is an image series with small changes of pixel intensity containing (quasi) static structures, while the foreground, or the vessel layer, consists of moving objects. Thus, resizing the background image into a column vector and combining all these vectors from a background series together results in a low rank matrix. Likewise, the image series of vessel layer can be modeled as a sparse matrix, as either vessels or guiding catheters take up only a small part of the whole image content. Therefore, the background layer and vessel layer of DIs can be separated by solving the Robust PCA problem.

2.2.3 Image Processing Pipeline of XA Layer Separation

The proposed layer separation algorithm consists of the following steps. All steps are illustrated in Fig. 2.2.

1. Given an XA sequence, apply morphological closing on each frame of the series, as described in Section 2.2.1. For each frame, subtract the morphological-closed image from the original image to obtain the DI.
2. Rearrange the DIs of the XA sequence to construct a matrix whose columns represent the frames. This matrix is considered as the input matrix M in Equation 2.1.
3. Solve the Robust PCA problem to obtain the background layer matrix L and vessel layer matrix S. Resize L and S to get the background layer and vessel layer of the previous size for each frame of the sequence.

2.3 Experiments

Fully anonymized imaging data were used in our experiments. Four XA image series that were acquired with Siemens AXIOM-Artis biplane system were analyzed. The frame rate of all sequences is 15 frames per second. The number of frames per series ranges from 55 to 169. From our data, the image matrix is 512×512 pixels for one of the series and 600×600 for the other three, with resolution 0.216×0.216 and 0.184×0.184 mm^2, respectively.

To quantify the visibility of vessels in an image, the contrast-to-noise ratio (CNR) is used in the experiments. CNR is a measure of image quality based on contrast. Once the background and foreground of an image is defined, the definition of CNR can be formulated as:

$$CNR = \frac{|\mu_F - \mu_B|}{\sigma_B} \tag{2.2}$$

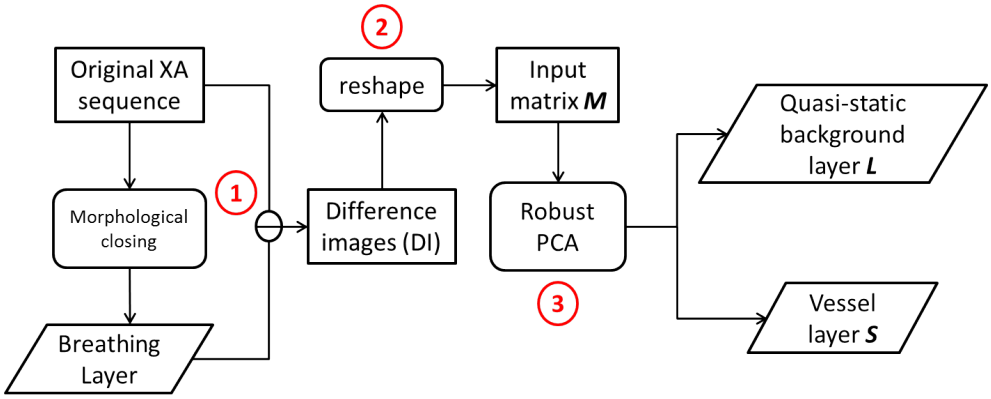


Figure 2.2: The pipeline of the proposed layer separation method.

where μ_F and μ_B are the mean of foreground and background pixel values respectively, and σ_B is the standard deviation of the background pixel values. This definition of CNR measures the contrast between the foreground and background pixel intensities in relation to the standard deviation of the background pixel intensities. Larger CNR values imply a better contrast.

Two different versions of CNR are computed, using two different masks for defining the foreground (vessel) and the background in XA images (Fig. 2.3). In mask 1, as shown in Fig. 2.3 column 1, a 4 *mm*-wide image area around the manually-labeled vessel centerline is defined as the foreground (the dark area inside white region); the background are its 3 *mm*-wide neighborhood area (white region surrounding the vessel). This mask can be used to assess the local contrast around vessels in XA. In mask 2, as shown in Fig. 2.3 column 3, everything outside the foreground is considered background, which thus also evaluates the removal of the diaphragm, guiding catheters, etc.. In our experiments, we randomly select 5 frames once from each sequence for the mask generation and compute the average CNR of the 5 frames.

We compare the performance of our approach to 3 other related methods. In [15], static background is eliminated by subtracting the median of the first 10 frames from each frame in the sequence. This method is referred to as *MedSubtract 1*. Second, we considered an advanced version of median subtraction by firstly removing the breathing layer using morphological closing and then subtracting the median. This is called *MedSubtract 2* in the experiments. Third, a conventional PCA technique is explored. The breathing layer is first removed to generate the difference image and the background layer is later reconstructed with the first principal component using PCA. This is referred to as *Normal PCA*.

For the parameter λ in the formulation of Robust PCA, we use the value suggested in [23]. All experiments were implemented in MATLAB 2013b on an Intel Core i7-4800MQ 2.70 GHz computer with 16 GB RAM running Windows.

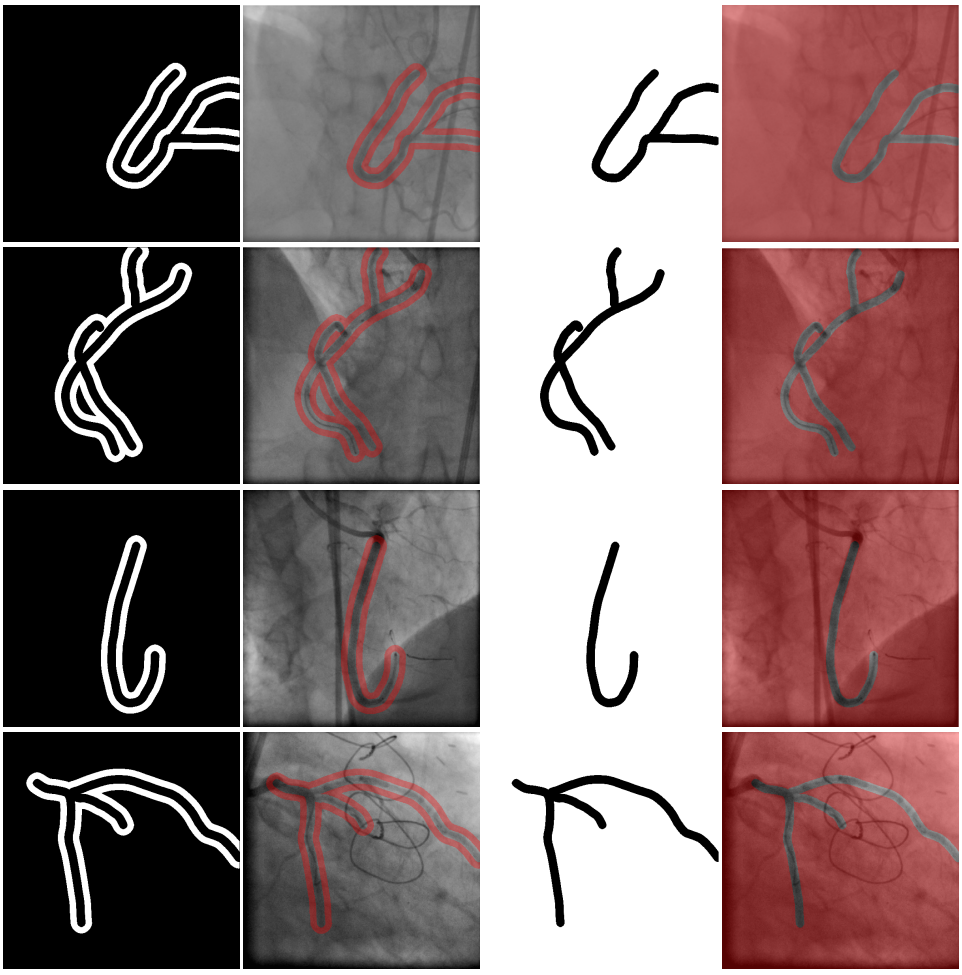
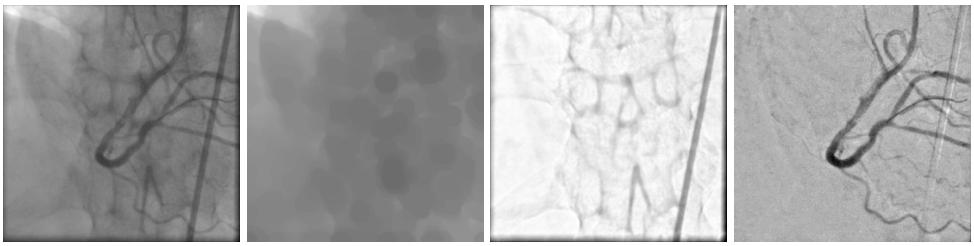


Figure 2.3: Two types of mask images. Background is defined as the white image region, foreground is defined as the dark area within the white part: (Column 1) Mask 1 for one frame in the four XA sequences; (Column 2) Mask 1 overlaid on the corresponding XA frames; (Column 3) Mask 2 for one frame in the four XA sequences; (Column 4) Mask 2 overlaid on the corresponding XA frames.

2.4 Results

Fig. 2.4 shows an example result of layer separation on one XA sequence. Note that in the original image (Fig. 2.4a), the presence of the diaphragm, the vertebral structures and the long guiding catheter segment makes extracting the vessels challenging. In the vessel layer image (Fig. 2.4d), those structures are removed, and the contrast between vessels and their neighborhood pixels is larger than in the original image.

Fig. 2.5 presents the comparison of our proposed method (Row 5) to three other



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

Figure 2.4: An example of layer separation: (a) the original image, (b) breathing layer, (c) quasi-static background layer, and (d) vessel layer.

background-removal methods (Row 2-4) applied on four XA sequences. For each of the sequences, we selected a representative frame. It can be observed that all the four methods increase the visibility of vessels in XA with better contrast. However, the result of *MedSubtract 1* method (Row 2) still presents artifacts in the foreground due to the motion of diaphragm, whereas our method successfully removes the diaphragm using morphological closing. Compared to *MedSubtract 2* (Row 3) and *Normal PCA* methods (Row 4), the method based on Robust PCA performs better on removing quasi-static structures, such as the guiding catheter segment in aorta (column 1-3) and stitches (column 4).

The CNR values of XA sequences and vessel layers are illustrated in Fig. 2.6. Compared to the original XA, as shown in both Fig. 2.6a and Fig. 2.6b, all methods improve the CNR values. For CNR 1, when only local contrast around vessels is measured, *Robust PCA* method performs better than the other approaches for patient 1 and 2, but has slightly lower CNR than *Normal PCA* for patient 3 and 4. In the case that the removal of diaphragm and guiding catheter is considered, as what CNR 2 indicates, *Robust PCA* is superior in all four patients.

2.5 Discussion and Conclusion

We have developed an automatic method for layer separation of interventional XA images, to enhance vessel visualization. The method separate XA images into a breathing layer, a quasi-static background layer and a vessel layer using morphological filtering and applying Robust PCA. The separation is evaluated on four XA

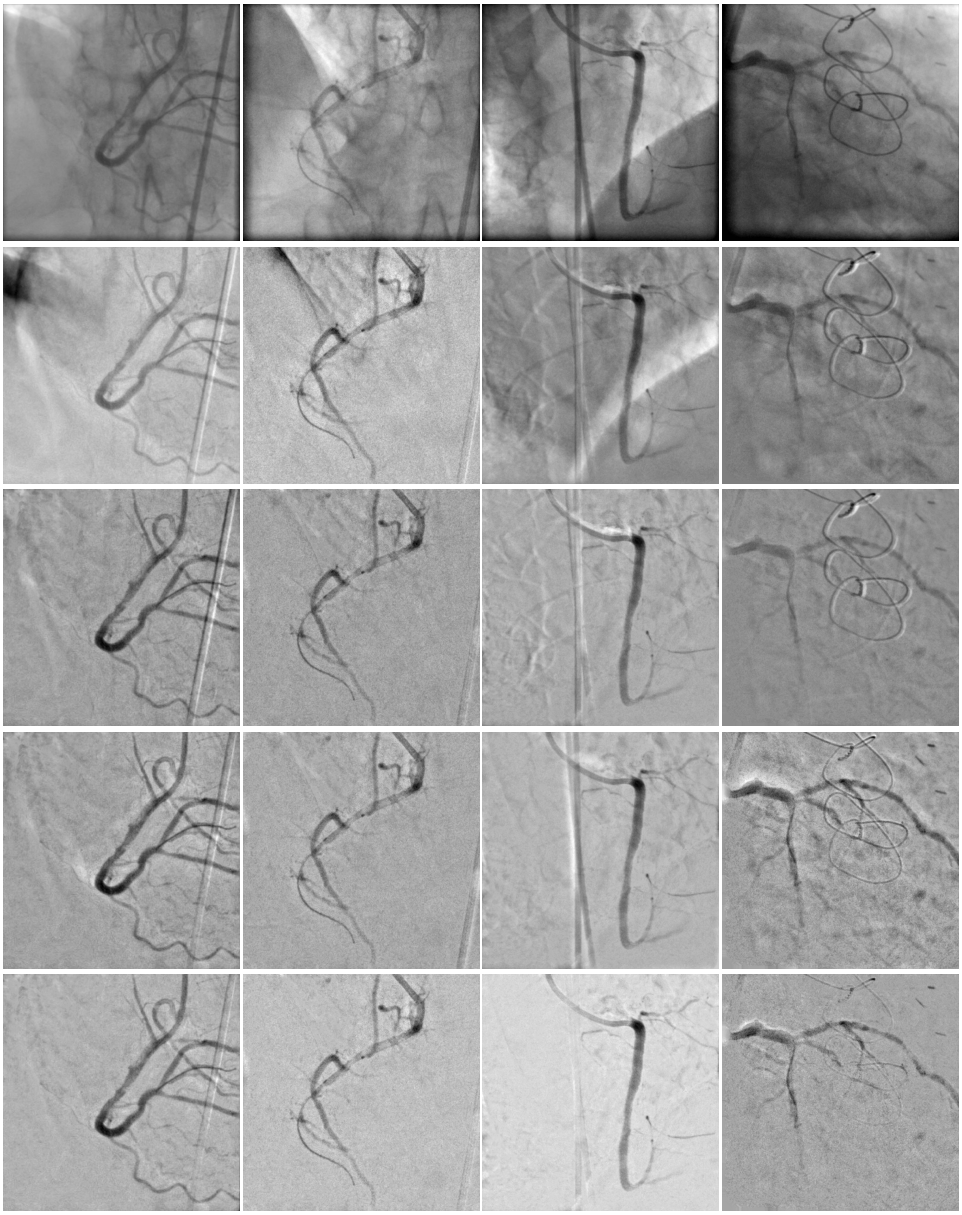
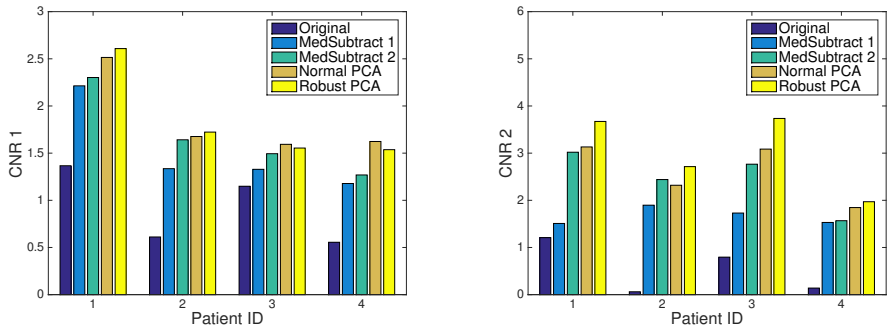


Figure 2.5: Example frames of foreground images obtained by different background-removal techniques applied on four XA sequences: (Column 1-4) The four different XA sequences, (Row 1) The original image, (Row 2) *MedSubtract 1*, (Row 3) *MedSubtract 2*, (Row 4) *Normal PCA*, (Row 5) our method using Robust PCA.



(a) CNR 1 **(b)** CNR 2

Figure 2.6: The average CNR over 5 randomly-chosen frames using two types of masks for the four XA sequences.

sequences, demonstrating better separation of the coronary arteries and reduced inclusion of breathing or quasi-static structures compared to other approaches.

Fig. 2.5 shows that the proposed method is able to improve the visibility of vessels and performs better on representative frames of the four XA sequences. Fig. 2.6a shows that the *Robust PCA* method is advantageous over the two median subtraction methods on improving the local contrast, and has similar performance with *Normal PCA*. Fig. 2.6b, which displays the global CNR measure, shows that *Robust PCA* is superior on all four patients which indicates that the superiority of *Robust PCA* to other approaches is more on removing respiratory and quasi-static structures from XA to improve the contrast of vessels in the whole image. This advantage could potentially reduce the generation of spurious vessels when applying vessel extraction methods on XA.

Compared to original images, the *Robust PCA* method improves image quality in the vessel layer by removing breathing structures and background objects. Compared to the absolute-static background resulted from the median-subtraction-based methods, *Robust PCA* models a quasi-static background with small changes, which is more adaptive to the change of image content caused by coronary motion. *Normal PCA* also models a flexible background , which could be the reason why it has similar performance with *Robust PCA*. Compared to *Normal PCA*, *Robust PCA* produces less residuals of guiding catheter in the vessel layer after the removal of the background layer. The regularization parameter of *Robust PCA* enables better flexibility of balancing between moving objects and background in layer separation. Compared to other related techniques e.g. in [116] [118], the main difference of the proposed method is that it does not rely on motion field, therefore, no motion field is required to extract before doing layer separation.

Several factors might have impact on CNR values. The masks defines the background and foreground, therefore the mask-related factors could directly influence the CNR values, e.g. the width of the foreground or background, whether or not including small vessels or the guiding catheter distal segment in the foreground. In addition,

the number of the selected frames for mask generation from each XA sequence might also be an important factor. More in-depth analysis of these factors is part of the future work.

In conclusion, we proposed a novel layer separation method based on morphological operation and Robust PCA. We also demonstrated that the method improves the visibility of coronary arteries in XA and has advantages over several other related approaches. In the future, we will assess this technique in prospective settings and study its application in approaches that improve image guidance in XA guided cardiac interventions.

Acknowledgement This work was supported by Technology Foundation STW, IMAGIC project under the iMIT program (grant number 12703).

Chapter Three

Automatic Online Layer Separation for Vessel Enhancement in X-ray Angiograms for Percutaneous Coronary Interventions

Abstract — Percutaneous coronary intervention is a minimally invasive procedure that is usually performed under image guidance using X-ray angiograms in which coronary arteries are opacified with contrast agent. In X-ray images, 3D objects are projected on a 2D plane, generating semi-transparent layers that overlap each other. The overlapping of structures makes robust automatic information processing of the X-ray images, such as vessel extraction which is highly relevant to support smart image guidance, challenging. In this paper, we propose an automatic online layer separation approach that robustly separates interventional X-ray angiograms into three layers: a breathing layer, a quasi-static layer and a vessel layer that contains information of coronary arteries and medical instruments. The method uses morphological closing and an online robust PCA algorithm to separate the three layers. The proposed layer separation method ran fast and was demonstrated to significantly improve the vessel visibility in clinical X-ray images and showed better performance than other related online or prospective approaches. The potential of the proposed approach was demonstrated by enhancing contrast of vessels in X-ray images with low vessel contrast, which would facilitate the use of reduced amount of contrast agent to prevent contrast-induced side effects.

Based upon: **H. Ma**, A. Hoogendoorn, E. Regar, W.J. Niessen and T. van Walsum: Automatic Online Layer Separation for Vessel Enhancement in X-ray Angiograms for Percutaneous Coronary Interventions. *Medical Image Analysis*, vol. 39, pp. 145-161, 2017.

3.1 Introduction

3.1.1 Motivation

Percutaneous coronary intervention (PCI) is a minimally invasive procedure for patients with advanced coronary artery disease. In this procedure, a stent pre-mounted on a delivery catheter is advanced over a guide-wire and through a guiding catheter at the site of narrowing in a patient's coronary arteries. Once the lesion site is reached, the delivery balloon is inflated and the stent is deployed against the coronary wall, assuring optimal patency of the artery. As there is no direct eyesight on the target area, these procedures are commonly performed under image guidance using X-ray angiography (XA), where coronary arteries are visualized with X-ray contrast agent. During the intervention, clinicians use XA images to navigate catheters and guidewires inside the patients.

As XA images contain useful information on anatomy and instrument position, many works have been published on extracting relevant information to improve the image guidance for cardiac interventions. For example, Panayiotou et al. [78] have developed a retrospective motion gating technique of interventional X-ray images through vessel extraction. Also using the information of vessels, pre/intra-operative information fusion between CT angiography and XA have been reported [14, 89]. Apart from vessels in XA, there is interest to track structures such as the lungs, catheters and guidewires. Shechter et al. [96] have used the position of diaphragm as an indicator of respiratory phase and constructed a patient specific coronary motion model based on that. In [13], the position of guiding catheter tip has been related to the combination of respiratory and cardiac motion.

Since X-ray images are projections of 3D structures on a 2D plane, the image content can be interpreted as a composition of several opaque or semi-transparent structures, which have different appearances and motion patterns. The overlapping nature of the structures makes automatic analysis of XA challenging. Separating the structures from each other enables visualizing and analyzing different structures independently, which would, therefore, potentially facilitate the information processing of XA. For example, vessel extraction using Hessian-based filtering method in XA is often hampered by non-vascular structures, such as guiding catheters, diaphragm border and vertebral body edges, because of their tubular or curvi-linear appearance in XA. Separating non-vascular structures would improve the visibility of vessels and promote automatic vessel extraction that would ultimately facilitate the image guidance during interventions.

In the context of this work, we interpret the process of separating those structures in XA images as a separation of a set of additive 2D layer images which add up to the original image, and each of them has different structures. The purpose of this work is to develop and evaluate a fast method that can run prospectively for the effective and efficient separation of the structures on different layers for XA sequences. Following the terminology from earlier works (in Section 3.1.2), we adopt the term "layer separation" to refer to the separation of structures and putting them in different layers.

3.1.2 Related Works

Existing methods for layer separation for X-ray fluoroscopic sequences can be categorized into two approaches: motion-based and motion-free .

Motion-based layer separation methods treat each frame of an X-ray fluoroscopic sequence as the outcome of the motion of each layer. Hence, the key part of obtaining the layers in these methods is estimating the motion of every layer. Various assumptions on the type of motion have been proposed. For instance, Close et al. [26] have estimated translation, rotation and scaling for each layer in a region of interest. The layers are computed by transforming each frame with the estimated motion and averaging the transformed frames. This method computed a total of four layers for a sequence. Zhu et al. [118] have proposed a two-layer separation scheme. They have developed a Bayesian framework that combines dense motion estimation, uncertainty propagation and statistical fusion to achieve layer separation. In a three-layer separation approach proposed in [116], a multi-scale framework has been developed based on different motion patterns for the static background, lung and vessels. In this work, a dense motion field of each layer has been constructed using thin plate splines. Fischer et al. [37] have further extended this method by introducing a regularization term for layers with a Bayesian model to aid layer separation. In particular, they have proposed to use a robust data term and edge-preserving regularization. In [12], a joint layer segmentation and parametric motion estimation scheme has been proposed for transparent image sequences. Similarly, Preston et al. [85] jointly estimated layers and their corresponding smooth deformation to model the non-smooth motion observed in a fluoroscopic sequence. A total variation based regularization was used to encourage sparsity of gradients within and across the layer images.

Unlike motion-based methods, motion-free approaches do not require estimating the motion of layers. Instead, they directly model the background layer or/and foreground (vessel) layer of an image sequence under certain hypotheses. One of the simplest ways of modeling the background of XA is computing the median of several frames in a sequence, and obtaining the foreground by subtracting the median image from the original frames [15]. This method worked well for the background that is entirely static, but generates artefacts when there are moving objects in the background, e.g. diaphragm in XA. A more advanced method has been proposed in [103] in which they assumed that the vessel and the backgrounds generate independent signals that are mixed in a sequence, so that the vessel-background separation becomes a blind source separation problem that is commonly solved by independent component analysis (ICA) [52].

Apart from ICA, robust principal component analysis (RPCA) is also a common approach for source decomposition. One of the most popular RPCA methods, principal component pursuit (PCP) [23], splits a data matrix into a low-rank component and a sparse component. It has been used for background modeling or foreground detection for surveillance videos [20]. In the field of medical image analysis, it found applications in reconstruction [76] and motion correction [43] in dynamic MRI. On the topic of layer separation for X-ray images, Ma et al. [65] (Chapter 2 of this thesis) have used morphological closing to remove breathing structures from the images and adopted RPCA to separate a quasi-static layer and a vessel layer from XA. This

method could only be used in a retrospective setting, since it requires all frames of a sequence. Volpi et al. [109] have developed a method that worked in a prospective setting. The method used vesselness filtering [38] and RPCA to separate a foreground that contains interventional devices. They have implemented the foreground separation by solving RPCA with a mini-batch of data: for each new coming mini-batch, the average of the low-rank component was estimated and used as the background for the next mini-batch. The limitation of this method is that the foreground separation of a mini-batch is delayed by the processing of the previous complete block of data.

Online robust PCA (OR-PCA) is an online extension of the original RPCA method, proposed in [35]. OR-PCA overcomes the limitation of RPCA-based methods by reformulating the nuclear norm in the RPCA formulation as an explicit low-rank factorization, so that it does not require to "see" the complete dataset or a mini-batch of data, but can process each single data sample one at a time. This setting enables online processing of streaming data. In [98], a closed-form solution for the subspace basis update in OR-PCA has been proposed and shown to achieve better performance in image alignment tasks. OR-PCA has been used in computer vision tasks, such as background subtraction [54] and foreground detection [53], but its application in the field of medical imaging has not been investigated yet.

3.1.3 Overview and Contributions

In this work we extended the method in [65] (Chapter 2 of this thesis) that only worked in a retrospective or "off-line" setting. To this end, we developed and evaluated an automatic motion-free online layer separation method for X-ray angiograms. The method robustly separates the layer that contains vessels and catheter tip from a (quasi) static background, while ignoring large-scale motion such as diaphragm movement. Our contributions are:

- We integrated OR-PCA in the layer separation scheme, enabling online layer separation for XA, which is a key ingredient for its potential application in a clinical workflow.

- Inspired by the work in [70], we proposed and analyzed three ways to downweight past information that is able to improve the layer separation performance using the original OR-PCA algorithm.

- We compared the proposed method with other related background-removal approaches and evaluated the results visually and quantitatively on real patient XA data.

- We investigated the potential of improving the contrast of vessels in a low-contrast scenario using the proposed method with synthetic low-contrast XA sequences and real sequences acquired in a pig experiment in which various contrast levels were used.

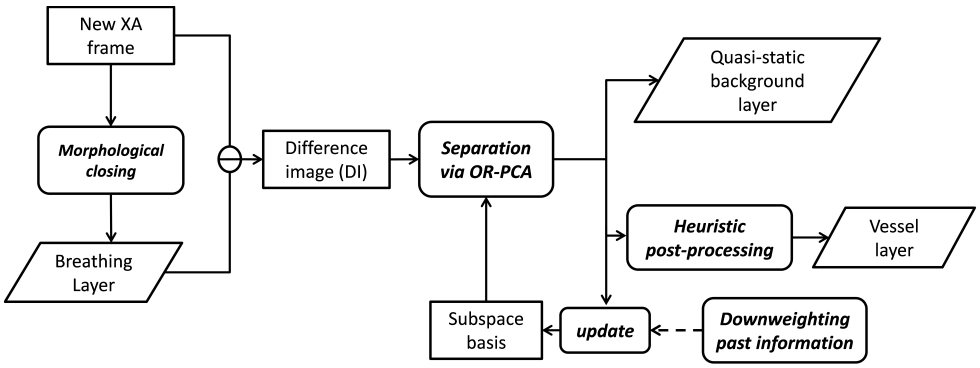


Figure 3.1: The overview of online layer separation for an XA frame.

3.2 Method

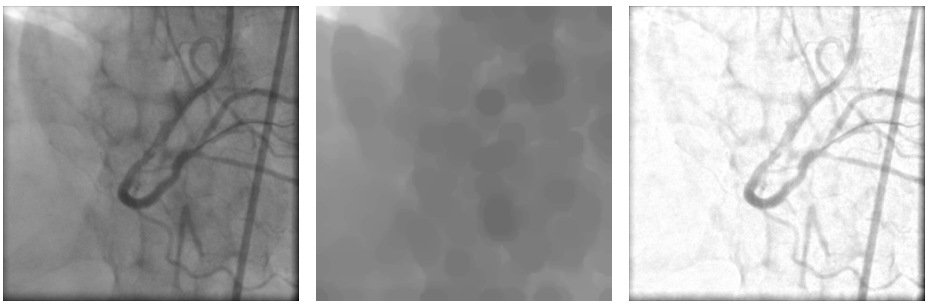
3.2.1 Overview

The proposed method treats the intensity of an XA frame as the sum of three layers, i.e., a "breathing" layer, a quasi-static layer and a layer that contains vessels. The method consists of two main steps: first, large-scale breathing structures, e.g. diaphragm, are separated and removed from the original XA frame, and second, smaller moving structures, e.g. vessels and guiding catheters, are separated from a quasi-static background using online robust PCA (OR-PCA). Fig. 3.1 provides an overview of the complete method, details are described in the remainder of Section 2.

3.2.2 Separation of Breathing Structures

To prevent artefacts remaining in the vessel layer due to respiratory motion, the layer that contains large-scale breathing structures, such as diaphragm, is removed from the original XA images in the first step.

The layer of breathing structures was obtained by removing "small" objects from the original X-ray angiographic frame. Depending on the field of view, those objects could include vessels, guiding catheters, guide wires, stitches and vertebral bodies. Following the approach in [65] (Chapter 2 of this thesis), as a preprocessing step, we applied a morphological closing operation to the XA image with a circular structuring element of 8.5 *mm* in diameter, in order to remove any tubular and curvilinear structures smaller than that size. An example of a resulting image is shown in Figure 3.2b, where the guiding catheter and vessels are removed and vertebral contours are blurred, while structures that are susceptible to breathing motion remain in the image (diaphragm and lung tissue are shown as the white area in the upper left part of the image). The resulting image is referred to as the "breathing layer" in this paper and was next subtracted from the original image to obtain the difference image (DI, Figure 3.2c) of the XA frame for further processing.



(a) (b) (c)

Figure 3.2: Morphological closing operation applied on an XA frame: (a) original frame, (b) image processed with morphological closing, (c) difference image (DI) (a-b).

3.2.3 Separation of Vessel Layer via OR-PCA

In this section, we briefly review the formulation of the online robust PCA method proposed in [35] and different subspace basis update strategies for solving the OR-PCA problem [35, 98]. Then we propose three different ways of coping with previous frames to improve on these methods.

3.2.3.1 Notation

Bold letters are used to denote vectors. With the difference image (DI) of an XA frame represented with a $k \times k$ matrix, we concatenated all pixels in this matrix to form a single column vector $\mathbf{z} \in \mathbb{R}^p$, where $p = k^2$ is the dimension of the observed sample. Likewise, we use $\mathbf{x} \in \mathbb{R}^p$ to denote the quasi-static background of the XA frame and $\mathbf{e} \in \mathbb{R}^p$ represents the foreground. Hence, $\mathbf{z} = \mathbf{x} + \mathbf{e}$. Let n denote the number of frames in a sequence, t be the index of the sample/time instance of a frame and r denote the intrinsic dimension of the subspace underlying $\{\mathbf{x}_i \,|\, i = 1, 2, \ldots n\}$.

Matrices are denoted by capital letters in the following sections. In particular, $Z \in \mathbb{R}^{p \times n}$ is the matrix of a complete sequence of difference images (DIs), where its column $\mathbf{z}_i$ represents the i-th DI. Likewise, X and E are the background and the foreground matrices with $\mathbf{x}_i$ and $\mathbf{e}_i$ the vector for the i-th background and the i-th foreground. For an arbitrary real matrix M, let $||M||_1 = \sum_{i,j} |M_{i,j}|$ denote the L_1-norm of M, $||M||_F$ denotes the Frobenius norm $||M||_F = \sqrt{\sum_{i,j} |M_{i,j}|^2}$, and $||M||_* = \sum_i \sigma_i(M)$ denotes the nuclear norm, i.e., the sum of its singular values. $Tr(M)$ denotes the trace of a matrix.

3.2.3.2 Online Robust PCA

Robust PCA (RPCA) aims at estimating the subspace underlying the observed samples. Among many popular RPCA methods, Principal Component Pursuit (PCP) [23] has been proposed to solve the RPCA problem by approximating the data matrix as

the sum of a low-rank matrix and a sparse matrix. The concepts of low-rank and sparsity have been implemented using the nuclear norm and the L_1-norm of matrix respectively. This formulation is suitable for the separation of the vessel layer from the DI of an XA frame, since the background has merely minor changes, which can be modeled as a low-rank matrix. In addition, the fact that vessels and guiding catheters take up only a small portion of the complete image content fits the requirement of sparsity.

3.2.3.2.1 The OR-PCA formulation

Different from the classical formulation in [23], PCP can be reformulated as Equation (3.1) [35]:

$$\min_{X,E} \frac{1}{2}||Z - X - E||_F^2 + \lambda_1||X||_* + \lambda_2||E||_1 \tag{3.1}$$

where λ_1 and λ_2 are regularization coefficients. Through minimizing the cost function (3.1) that contains the nuclear norm of the background X and the L_1-norm (sparsity) of the foreground E, the RPCA algorithm aims at obtaining the background (X) and foreground (E) that best approximate the XA sequence (Z). Because the nuclear norm couples all samples tightly, typical methods to solve Equation (3.1), such as Augmented Lagrangian Multiplier (ALM) [63], are often implemented in a batch manner, which limits its application in scenarios that deal with streaming data, e.g. X-ray cine angiography data during coronary interventions.

To overcome this problem, Feng et al. [35] have proposed to use an equivalent form of the nuclear norm:

$$||X||_* = \underset{L,R}{inf} \left\{ \frac{1}{2}||L||_F^2 + \frac{1}{2}||R||_F^2 : X = LR^T \right\} \tag{3.2}$$

where *inf* denotes the greatest lower bound of a subset of a partially ordered set, $L \in \mathbb{R}^{p \times r}$ is the basis of the low-dimensional subspace and $R \in \mathbb{R}^{n \times r}$ can be seen as the samples' coefficient with respect to the basis. Substituting Equation (3.2) into (3.1), the RPCA problem can be reformulated as (3.3):

$$\underset{L,R,E}{min} \frac{1}{2}||Z - LR^T - E||_F^2 + \frac{\lambda_1}{2}(||L||_F^2 + ||R||_F^2) + \lambda_2||E||_1 \tag{3.3}$$

Following [35], solving Equation (3.3) is equivalent to minimizing the following empirical cost function given a sequence Z consisting of n samples $[\mathbf{z}_1 \dots \mathbf{z}_n]$:

$$f_n(L) \triangleq \frac{1}{n}\sum_{i=1}^{n} l(\mathbf{z}_i, L) + \frac{\lambda_1}{2n}||L||_F^2 \tag{3.4}$$

where the loss function $l(\mathbf{z}, L)$ for each sample is defined as:

$$l(\mathbf{z}, L) \triangleq \underset{\mathbf{r},\mathbf{e}}{min} \frac{1}{2}||\mathbf{z}_i - L\mathbf{r} - \mathbf{e}||_2^2 + \frac{\lambda_1}{2}||\mathbf{r}||_2^2 + \lambda_2||\mathbf{e}||_1 \tag{3.5}$$

Note that Equation (3.4) enables the possibility of updating the basis L based on each individual sample. To handle streaming data in practice, in [35], the estimation of basis L_t is obtained through minimizing the following surrogate function of (3.4) with respect to L for the t-th time instance:

$$g_t(L) \triangleq \frac{1}{t}\sum_{i=1}^{t}(\frac{1}{2}||\mathbf{z}_i - L\mathbf{r}_i - \mathbf{e}_i||_2^2 + \frac{\lambda_1}{2}||\mathbf{r}_i||_2^2 + \lambda_2||\mathbf{e}_i||_1) + \frac{\lambda_1}{2t}||L||_F^2 \tag{3.6}$$

Also observe that the loss function (3.5) optimizes $\mathbf{r}$ (the coefficient of $\mathbf{z}_i$ on the basis L) and $\mathbf{e}$ (the sparse component of $\mathbf{z}_i$) to minimize the cost given a fixed basis. Through an alternating optimization of $\mathbf{r}$, $\mathbf{e}$ and L, Equation (3.4) can be solved in an online manner. The complete stochastic optimization scheme for solving the OR-PCA problem is described in Algorithm 1.

Algorithm 1 Stochastic optimization for OR-PCA [35]

Require: $\{\mathbf{z}_1, \ldots, \mathbf{z}_T\}$ (sequentially revealed data samples), $\lambda_1, \lambda_2 \in \mathbb{R}$ (regularization parameters), $L_0 \in \mathbb{R}^{p\times r}, \mathbf{r}_0 \in \mathbb{R}^r, \mathbf{e}_0 \in \mathbb{R}^p, A_0 = \mathbf{0}^{r\times r}, B_0 = \mathbf{0}^{p\times r}$ (initial solution), T (number of samples).

1: **for** t = 1 **to** T **do**

2: Reveal the sample $\mathbf{z}_t$.

3: Given L_{t-1}, project the new sample:

$$\{\mathbf{r}_t, \mathbf{e}_t\} = \underset{\mathbf{r},\mathbf{e}}{argmin}\frac{1}{2}||\mathbf{z}_t - L_{t-1}\mathbf{r} - \mathbf{e}||_2^2 + \frac{\lambda_1}{2}||\mathbf{r}||_2^2 + \lambda_2||\mathbf{e}||_1 \tag{3.7}$$

4: $A_t \leftarrow A_{t-1} + \mathbf{r}_t\mathbf{r}_t^T, B_t \leftarrow B_{t-1} + (\mathbf{z}_t - \mathbf{e}_t)\mathbf{r}_t^T$

5: Update the basis L_t

$$L_t \triangleq \underset{L}{argmin}\frac{1}{2}Tr[L^TL(A_t + \lambda_1 I)] - Tr(L^TB_t) \tag{3.8}$$

6: **end for**

7: **return** $X_T = L_TR_T^T$ (the low-rank matrix), E_T (the sparse matrix).

Note that the right-hand side of Equation (3.7) in Algorithm 1 is equivalent to the loss function (3.5) for the t-th sample. To solve it, Feng et al. [35] give a closed-form solution to alternatively update $\mathbf{r}$ and $\mathbf{e}$ until a convergence criterion is met. The update of L_t in Equation (3.8) is discussed in the next section.

3.2.3.2.2 Update the subspace basis L_t

To minimize the function (3.6) with respect to L, note that the term $\frac{\lambda_1}{2}||\mathbf{r}_i||_2^2$ and $\lambda_2||\mathbf{e}_i||_1$ can be discarded, we then derived the following expression for L_t from (3.6):

$$L_t \triangleq \underset{L}{argmin} \frac{1}{2} Tr[L^T L \sum_{i=1}^{t} (\mathbf{r}_i \mathbf{r}_i^T + \frac{\lambda_1}{t} I)] - Tr(L^T \sum_{i=1}^{t} ((\mathbf{z}_i - \mathbf{e}_i)\mathbf{r}_i^T)) \tag{3.9}$$

Using the two intermediate variables A_t and B_t that accumulate information of past frames, Equation (3.9) is equivalent to (3.8) in Algorithm 1. Equation (3.8) is then solved by the block-coordinate descent method, i.e., each column of the basis L is updated sequentially while fixing the other columns (see Algorithm 2).

Algorithm 2 The basis update using block-coordinate descent [35]

Require: $L = [\mathbf{l}_1, \ldots, \mathbf{l}_r] \in \mathbb{R}^{p \times r}, A = [\mathbf{a}_1, \ldots, \mathbf{a}_r] \in \mathbb{R}^{r \times r}, B = [\mathbf{b}_1, \ldots, \mathbf{b}_r] \in \mathbb{R}^{p \times r}$. $\tilde{A} \leftarrow A + \lambda_1 I$.

1: **for** j = 1 **to** r **do**

2: Update the j-th column of L.

$$\mathbf{l}_j \leftarrow \frac{1}{\tilde{A}_{j,j}} (\mathbf{b}_j - L\tilde{\mathbf{a}}_j) + \mathbf{l}_j. \tag{3.10}$$

3: **end for**

4: **return** L.

Another way of solving Equation (3.8) is to derive a closed-form solution. Let the derivative of the right-hand side of (3.8) with respect to L be zero, we obtain

$$\frac{1}{2} L(A_t + \lambda_1 I)^T + \frac{1}{2} L(A_t + \lambda_1 I) - B_t = 0 \tag{3.11}$$

where $A_t = A_{t-1} + \mathbf{r}_t \mathbf{r}_t^T, B_t = B_{t-1} + (\mathbf{z}_t - \mathbf{e}_t)\mathbf{r}_t^T$. As $(A_t + \lambda_1 I)$ is symmetrical, a simple closed-form solution of (3.8) can be derived as

$$L_t = B_t (A_t + \lambda_1 I)^{-1}. \tag{3.12}$$

This is equivalent to the form given in [98].

3.2.3.3 Downweighting the Past Information

The previous solutions for the subspace basis update treat all samples equally, which works well for scenarios where samples are independently drawn. For stream video data, however, adjacent frames have higher correlation than "distant" frames. Thus, it may be possible to improve the basis update by treating past frames with different weights, giving close-by frames higher impact to the result than the distant frames. Inspired by the work in [70] which has reported several possibilities to handle past data in an online dictionary learning problem, we propose three approaches to downweight past information for the OR-PCA algorithm. In Algorithm 1, as A_t and B_t contain information of past frames, variations can be made to replace the following equation set on line 4 in Algorithm 1:

$$\begin{aligned} A_t &\leftarrow A_{t-1} + \mathbf{r}_t \mathbf{r}_t^T \\ B_t &\leftarrow B_{t-1} + (\mathbf{z}_t - \mathbf{e}_t)\mathbf{r}_t^T \end{aligned} \tag{3.13}$$

A logical choice is to apply an exponential decay (ED) to "forget" past information as in (3.14):

$$\begin{aligned} A_t &\leftarrow (1-\epsilon) A_{t-1} + \mathbf{r}_t \mathbf{r}_t^T \\ B_t &\leftarrow (1-\epsilon) B_{t-1} + (\mathbf{z}_t - \mathbf{e}_t)\mathbf{r}_t^T \end{aligned} \tag{3.14}$$

where ϵ is the decay rate and $0 < \epsilon < 1$. So for the t-th time instance, the weight for the i-th sample is $(1-\epsilon)^{t-i}$.

Similar to [70], as a second option we consider supra-linear decay (SLD) approach:

$$\begin{aligned} A_t &\leftarrow \left(1 - \tfrac{1}{t}\right)^{\rho} A_{t-1} + \mathbf{r}_t \mathbf{r}_t^T \\ B_t &\leftarrow \left(1 - \tfrac{1}{t}\right)^{\rho} B_{t-1} + (\mathbf{z}_t - \mathbf{e}_t)\mathbf{r}_t^T \end{aligned} \tag{3.15}$$

where ρ is a tunable decay parameter and $\rho \geqslant 0$. At the t-th time instance, the weight for the i-th sample becomes $\left(\frac{i}{t}\right)^{\rho}$. Note that: when $\rho = 0$, (3.15) turns into (3.13); when $\rho = 1$, (3.15) degrades to a linear decay.

Apart from ED and SLD that scale the past data, it is also an option to focus only on adjacent frames in a fixed-size window, so that frames within the sliding window are treated equally, whereas the frames outside the window from the earlier times are not considered for the basis update, as follows:

$$\begin{cases} \begin{aligned} A_t &\leftarrow \mathbf{r}_t \mathbf{r}_t^T \\ B_t &\leftarrow (\mathbf{z}_t - \mathbf{e}_t)\mathbf{r}_t^T \end{aligned} \quad , & t_0 = 1 \\ \begin{aligned} A_t &\leftarrow A_{t-1} + \mathbf{r}_t \mathbf{r}_t^T \\ B_t &\leftarrow B_{t-1} + (\mathbf{z}_t - \mathbf{e}_t)\mathbf{r}_t^T \end{aligned} \quad , & t_0 > 1 \text{ and } t \leqslant t_0 \\ \begin{aligned} A_t &\leftarrow A_{t-1} + \mathbf{r}_t \mathbf{r}_t^T - \mathbf{r}_{t-t_0} \mathbf{r}_{t-t_0}^T \\ B_t &\leftarrow B_{t-1} + (\mathbf{z}_t - \mathbf{e}_t)\mathbf{r}_t^T - (\mathbf{z}_{t-t_0} - \mathbf{e}_{t-t_0})\mathbf{r}_{t-t_0}^T \end{aligned} \quad , & \text{else} \end{cases} \tag{3.16}$$

where t_0 is the window size (number of frames within the window). This approach is referred to as "sliding-window (SW)".

3.2.4 Summary

The proposed online layer separation method consists of the following steps, as shown in Figure 3.1.

1. **Breathing layer separation** When a new XA frame is obtained, the breathing layer is firstly extracted by applying morphological closing on that frame, as described in Section 3.2.2. Subsequently, the breathing layer is subtracted from the original frame to obtain the DI.

2. **Quasi-static layer and vessel layer separation** Transform the DI from a matrix to vector by concatenating each column of the matrix one after another. This vector is then separated into two components by the OR-PCA method, as described in Section 3.2.3. The sparse component is reshaped to form the vessel layer, the other component is constructed as the quasi-static layer.

Finally, as pixels belonging to contrast agent always have negative value in the vessel layer, pixels with positive value in the vessel layer are heuristically set to zero to suppress artefacts.

3.3 Experiments

3.3.1 Image Data

In this work, we used three types of data for evaluation: clinical X-ray angiograms, synthetic low-contrast XA and X-ray angiographic data of pigs with variations in contrast concentration.

3.3.1.1 Clinical X-ray Angiographic Data

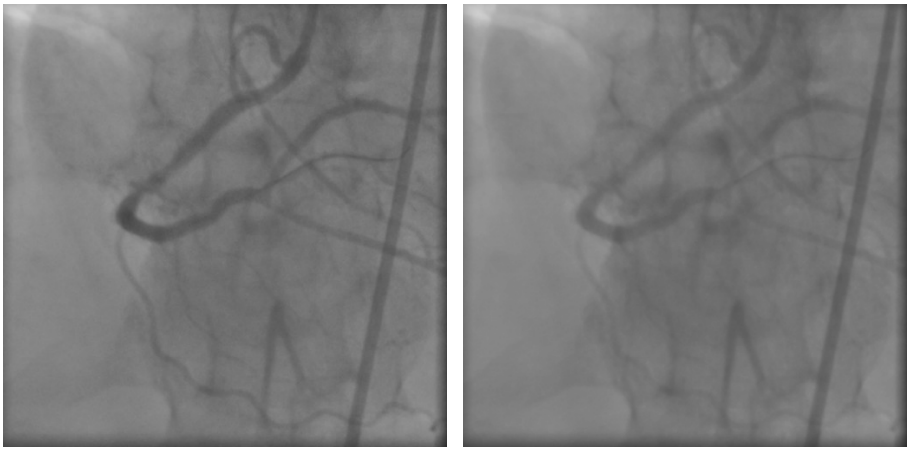
Imaging data from clinical routine that were anonymized were used for our experiments. The data were acquired under standard clinical protocol from the Department of Cardiology at Erasmus MC in Rotterdam, the Netherlands. The 42 XA sequences are from 21 patients who underwent a PCI procedure and were acquired with Siemens AXIOM-Artis biplane system. The frame rate of all sequences is 15 frames per second (fps). The number of frames per sequence varies from 46 to 244. All 42 XA sequences have in total 4886 frames. 22 sequences have 512×512 pixels, 12 have 600×600 pixels, 2 have 776×776 and 6 have 1024×1024. Their corresponding pixel sizes are 0.216×0.216 or 0.279×0.279, 0.184×0.184, 0.184×0.184 and $0.139 \times 0.139\, mm^2$, respectively. In all sequences, inflow and wash-out of contrast agent can be observed.

3.3.1.2 Synthetic Low-Contrast XA

The synthetic image data was used to simulate the condition that a reduced amount (50%) of contrast agent is administered, for the purpose of testing our online layer separation method on low-contrast XA. To create these synthetic XA sequences from the real ones, we used the off-line layer separation method in [65] (Chapter 2 of this thesis). The idea is that the real clinical XA sequence was firstly separated into three layers. The intensity of the vessel layer was then halved and added back to the other two layers to generate a new XA sequence that has half the amount of intensity compared to the original one, as shown in Equation (3.17):

$$I_{synthetic} = \alpha\, I^*_{vessel} + I^*_{static} + I^*_{breathing} \tag{3.17}$$

where $I_{synthetic}$ denotes the synthetic XA sequence, I^*_{vessel}, I^*_{static} and $I^*_{breathing}$ are the vessel layer, quasi-static layer and breathing layer separated using the method in [65] (Chapter 2 of this thesis), respectively, and $\alpha = 0.5$. The synthetic sequence



(a) Clinical XA **(b)** Synthetic XA

Figure 3.3: An example frame of real clinical XA sequences and synthetic low-contrast XA sequences: (a) the real image, (b) the synthetic XA frame with 50% vessel contrast.

has the same number of frames, same image size and resolution as its original in the clinical dataset. An example of a synthetic low-contrast XA is shown in Figure 3.3b. Note that the vessels have less contrast to the background than the original image in Figure 3.3a. We created a low-contrast XA sequence from each clinical XA described in Section 3.3.1.1, which results in 42 synthetic XA sequences in total.

3.3.1.3 X-ray Angiograms of Pigs

Additionally, *in vivo* XA data were acquired during a pig experiment performed at the Erasmus MC in Rotterdam, the Netherlands. 4 XA sequences with different contrast concentration levels were obtained from 1 FBM (familiar-hypercholesteremia Bretonchelles Meishan) pig which underwent a catheterization procedure after 14 months of high-fat diet. The XA sequences were acquired using a Siemens AXIOM-Artis monoplane system. The frame rate of all sequences is 15 frames per second. The number of frames per sequence varies from 48 to 79. The 4 XA sequences have in total 238 frames. All sequences have 776×776 pixels corresponding to a pixel size of $0.184 \times 0.184\,mm^2$. In all images, the inflow of contrast agent can be observed. The XA images were made during a manual injection of iso-osmolar X-ray contrast medium (Visipaque 320, GE Healthcare, Buckinghamshire, U.K), delivered through the guide catheter. The full-contrast images were acquired with a contrast concentration of 320 g/mL. For the 25%, 50% and 75% contrast concentration images, the contrast agent was diluted accordingly with a 0,9% sodium-chloride solution (saline). Prior to image acquisition, the guide catheter was flushed with the right concentration of the contrast agent.

In practice, the full-contrast sequence had lower visual contrast than the images

with 75% contrast. This might be due to incomplete flushing of the guiding catheter so that the contrast agent from the previous injection dilutes the current contrast agent.

3.3.2 Experiment 1: Parameter Tuning for OR-PCA

OR-PCA has three parameters: the intrinsic rank of the subspace basis r and the regularization parameters λ_1 and λ_2. In [35, 98], both λ_1 and λ_2 were set to $1/\sqrt{p}$, where p is the dimension of data. This value had been proposed in [23] as a general rule of thumb, but it can be slightly adjusted to achieve the best possible result. Javed et al. [53], for example, have empirically selected different values for λ_1 and λ_2 instead of $1/\sqrt{p}$. Unlike the rule for choosing λ_1 and λ_2, the choice for r depends more on specific applications.

In order to find the optimal parameter setting for the layer separation application on the clinical XA data, we used the following way to quantify the outcome of layer separation with a certain set of parameters.

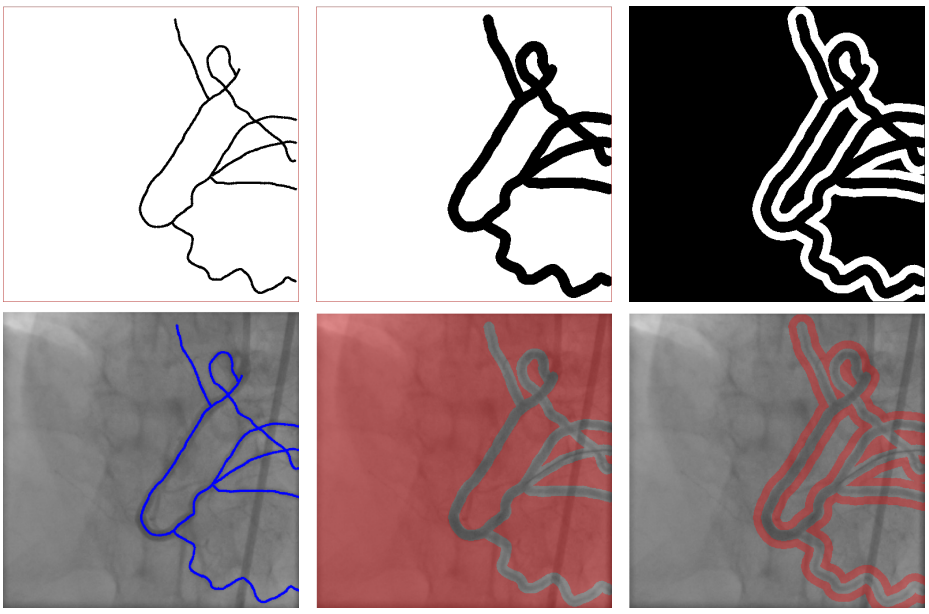
3.3.2.1 The Definition of Foreground and Background

We firstly defined the "foreground" and the "background" for the objective of optimization in Section 3.3.2.2. It is worth noticing that the foreground and the background here are merely defined for computing the vessel contrast and thus should not be confused with the foreground and background's definition coming from the layer separation scheme described in the previous sections.

We used masks to define the foreground and background. A 1 *mm* wide area around manually-labeled vessel centerlines was considered as the foreground (shown as the dark area in the mask in Figure 3.4a). This area falls entirely within the vessel, and thus is a good representative of pixels belonging to vessels. For background, we adopted two different masks for measuring "global" and "local" contrast. The first one highlights all pixels outside a 4 *mm* wide area around the vessel centerlines (the white area in the mask in Figure 3.4b). This mask can quantify the effect of the removal of diaphragm, guiding catheters, etc. and can be used in a global measurement of contrast. The local background is defined as a 3 *mm* wide neighborhood area around the dark area in the global mask (the white area in the mask in Figure 3.4c).

For each clinical XA sequence, we randomly selected 8-15 frames for mask generation and contrast evaluation. The number of selected frames depends on the sequence length. As the vessel contrast is of main interest in this paper and in practice, only the frames with contrast agent were selected. This way we also avoided choosing non-contrast frames from the beginning of a sequence where the online algorithm has not converged yet. In total, 444 frames were chosen from 42 sequences.

We also created the masks for the four pig XA sequences. From each pig XA data, we randomly chose 8-12 frames. In total, 38 frames were chosen for the mask creation. These masks are only used for evaluation of the contrast level in pig data, not for parameter optimization.



(a) (b) (c)

Figure 3.4: Examples of masks that are used for defining foreground and background in assessing the contrast of vessels. The first row shows the masks, the second row exhibits the overlay of masks on the corresponding original XA frame indicated by colors. (a) the foreground (blue), (b) the global background (red), (c) the local background (red).

3.3.2.2 The Objective for Parameter Optimization

Metrics that have previously been used to measure vessel visibility include contrast-to-noise ratio (CNR), as the work in [65] (Chapter 2 of this thesis), and the Jeffries-Matusita distance (JMD) [116, 118]. These metrics evaluate the contrast of pixels from two groups, e.g. foreground and background. However, when tuning parameters of OR-PCA using either of these two measures as the objective for optimization, the optimal parameters are those that yield a very small standard deviation of the background, thus an almost constant background, and a flawed separation of vessel layer that loses much intensity of the vessel pixels. These would result in a large CNR or JMD, but do not lead to good foreground and background separation.

To avoid the problem, the objective for OR-PCA parameter optimization should also consider the information loss in the vessel layer in addition to the vessel contrast. In this work, we integrated the difference between the original XA image and the sum of its three layers separated by OR-PCA method in the objective, such that losing too much information in the vessel layer would result in a large difference between the original XA and the sum of layer images. With this consideration, we used a corrected CNR (cCNR) as the objective to optimize the OR-PCA parameters:

$$cCNR = \frac{|\mu_F - \mu_B|}{\sqrt{\sigma_B^2 + MSE^V}} \tag{3.18}$$

where μ_F and μ_B are the mean of the pixel intensity value in the foreground and the background that were defined in Section 3.3.2.1, σ_B is the standard deviation of the pixel intensity in the background. MSE^V, the mean square error in the vessel area, which serves as a penalty term in Equation (3.18) to prevent too much information loss in the vessel layer, is defined as follows:

$$MSE^V = \frac{\sum_{x,y}(I^V_{original}(x,y) - I^V_{3-layer}(x,y))^2}{|I^V_{original}|} \tag{3.19}$$

where $I^V_{(\star)}$ denotes the operation that takes only pixels in the vessel area (defined by the dark region in the mask in Figure 3.4b) into consideration for image $I_{(\star)}$. The reason to focus only on the vessel area is that the information loss of vessel pixels only occurs in this region. $I_{original}(x,y)$ and $I_{3-layer}(x,y)$ are the pixel values of the position (x,y) in the original XA and the 3-layer sum image:

$$I_{3-layer} = I_{vessel} + I_{static} + I_{breathing} \tag{3.20}$$

where I_{vessel}, I_{static} and $I_{breathing}$ are the vessel layer, quasi-static layer and breathing layer, respectively. $|I^V_{original}|$ denotes the number of pixels in the vessel area in the frame.

According to (3.19), MSE^V indicates how well the original image can be reconstructed from the layer separation result. An undesirable reconstruction with pixel intensity loss in the vessel layer would result in a large MSE^V and, further, a small cCNR.

The cCNR for a complete sequence is defined as the average cCNR over all selected XA frames from the sequence. Global and local cCNR are computed respectively using the masks in Figure 3.4b and 3.4c.

3.3.2.3 Parameter Optimization

The parameters of OR-PCA for both subspace basis update methods were optimized by exhaustively searching the optimal parameter set that maximizes the previously defined objective cCNR within a discrete set of parameters. First, cCNR was computed for every possible parameter combination $(\lambda_1, \lambda_2, r)$ within the parameter range for each clinical XA sequence. Then, the optimal parameters were obtained by searching for the parameter set that maximizes the average local cCNR over the 42 XA sequences. This optimization was performed for the two different basis update methods in Section 3.2.3.2.2 respectively.

The range for the intrinsic rank r was chosen as the integers in [2, 20]. The regularization parameter λ_1 and λ_2 were set to the same value as $[35, 98]$ both set $\lambda_{1,2}$ to $1/\sqrt{p}$. To search for the optimal $\lambda_{1,2}$, we explored the values in $[0.1/\sqrt{p}, 10/\sqrt{p}]$ with a search step $0.1/\sqrt{p}$.

3.3.3 Experiment 2: Downweighting the Past Data in OR-PCA, Influence of the Parameters

Once the optimal parameter settings of OR-PCA had been obtained from the previous experiment, we used this setting and study the influence of history parameters that were introduced in Section 3.2.3.3 for downweighting the past data on the performance of OR-PCA. The search spaces for exponential decay (ED), supra-linear decay (SLD) and sliding-window (SW) are:

$$\epsilon \in \{0.01, 0.3, 0.6, 0.9, 0.99, 0.9968, 0.999, 0.9997, 1-10^{-4}, 1-10^{-5}, 1-10^{-6}, 1-10^{-7}\},$$

$$\rho \in \{0.5, 1, 1.5, 2, 2.5, 3, 6, 9, ..., 36, 39, 42\},$$

$$t_0 \in [1, 15] \text{ and } t_0 \in \mathbb{N}.$$

In the experiments, we combined these three approaches with the two methods that update the subspace basis L that were mentioned in Section 3.2.3.2.2, i.e., block-coordinate descent (BCD) and the closed-form solution (CF). In this paper, for example, the OR-PCA method using exponential decay to downweight past data and using closed-form solution to update L is referred to as OR-PCA (ED+CF), or (ED+CF) as a short form.

As the OR-PCA parameters $\lambda_{1,2}$ and r tuned with cCNR assure a reliable layer separation, and tuning the parameters of ED, SLD and SW does not create the previously-mentioned undesirable layer separation, therefore, we evaluated the results for this experiment using the direct measure of vessel contrast in an image CNR, which were used in [65] (Chapter 2 of this thesis), with the same masks from Section 3.3.2.1. In the evaluation, RPCA was used as a reference for the comparison purpose. Its regularization parameter λ had been optimized the same way as in Section 3.3.2.3 and was set to the optimal value $1.5/\sqrt{p}$ from the search space $[0.1/\sqrt{p}, 10/\sqrt{p}]$ with a search step $0.1/\sqrt{p}$. The experiments were carried out with the 42 clinical XA sequences.

3.3.4 Experiment 3: Comparison with Other Methods

We compared the proposed approaches to several other related methods that can be used for prospective or online layer separation. The off-line method with the batch version of robust PCA in [65] (Chapter 2 of this thesis) was used as a "benchmark" to show how close prospective or online methods can achieve to the performance of the off-line layer separation. The same way in Section 3.3.3, the regularization parameter λ of RPCA was set to $1.5/\sqrt{p}$ to achieve optimal performance. This method is referred to as **RPCA**. The following methods were tested in the experiment.

- **Median-subtraction** In [15], static background has been suppressed by subtracting the median of the first 10 frames from each frame in the sequence. This method is referred to as **MS**.

- **Morphological-closing + median-subtraction** This advanced version of median subtraction method removes the breathing layer via morphological closing and then subtracting the median of the first 10 frames. This method is referred to as **MC+MS**.

- **Robust PCA with a sliding window** As mentioned in Section 3.1.2, In [21,109], RPCA was solved within a sliding window that consists of a few frames to enable prospective foreground separation. We adopted this idea of solving RPCA for our experiment. Different from their methods using Frangi filtering to preprocess images, to adapt to our application, we applied morphological closing to remove the breathing layer and then separate the other two layers from the difference images by solving RPCA with a sliding window. We used two sets of parameters for this method. The first one was used in [21,109]: the window size was set to 4 and the regularization parameter λ was set to $1.5/\sqrt{p}$[1]. This one is referred to as **RPCA (SW)**. The second set of parameters was optimized the same way as in Section 3.3.2.3. The window size was set to 7 from the search space [2, 10] (integer two to ten) and the regularization parameter λ was set to $0.5/\sqrt{p}$ from the search space $[0.1/\sqrt{p}, 10/\sqrt{p}]$ with a search step $0.1/\sqrt{p}$. This one is referred to as **RPCA (SW)***.

These methods were compared with the OR-PCA approaches using CNR as the evaluation metric with the same masks from Section 3.3.2.1. The CNR of OR-PCA approaches were computed in a leave-one-out (LOO) manner. In each LOO loop, firstly, the OR-PCA parameters $\lambda_{1,2}$ and r were optimized on the training sequences using the method in Section 3.3.2. Next, using the trained OR-PCA parameters, the history parameters were trained on the same set of sequences with local CNR using the approach in Section 3.3.3. Lastly, the optimal parameters obtained from the previous two steps were applied to compute CNR for the left-out data. The overall CNR was then computed as the average CNR over all LOO sessions. The LOO experiment was carried out with the 42 clinical XA sequences.

In addition to CNR as an evaluation metric of vessel contrast, to gain insight into how accurate the layers obtained with each method can reconstruct the original XA image, the reconstruction error was evaluated. It was computed as follows:

$$E_{recon} = \frac{\sum_{x,y} |I_{original}(x,y) - I_{3-layer}(x,y)|}{\sum_{x,y} I_{original}(x,y)} \tag{3.21}$$

where E_{recon} denotes the reconstruction error, $I_{original}$ and $I_{3-layer}$ are defined in the same way as Equation (3.19). The layer separation parameters used for computing E_{recon} were the same that were obtained during each LOO loop in the last paragraph. For each clinical XA sequence, the reconstruction error was computed for the frames that had been selected for mask generation in Section 3.3.2.1. The average error over all selected frames from a sequence was used as the empirical reconstruction error for this sequence. This error indicates the relative absolute difference between the pixel

[1] λ was fixed to 3×10^{-3} in their papers. To adapt to different image sizes, we set λ to $1.5/\sqrt{p}$. This value is equivalent to 3×10^{-3} for images of size 512×512.

intensity of the reconstruction image and that of the original image with respect to the pixel intensity of the original image.

3.3.5 Experiment 4: Vessel Enhancement in Low-Contrast XA

One possible application of layer separation is vessel enhancement. This can be achieved through enhancing the vessel layer and adding it back to the original image. To demonstrate this concept, we conducted experiments to enhance vessels in low-contrast XA using the online layer separation approaches.

The data we used are synthetic human XA data and real XA data acquired from pigs, as introduced in Section 3.3.1.2 and 3.3.1.3. We first separated the three layers, then enhanced the vessel layer by multiplying it by an enhancement factor $\beta > 0$. Finally, the vessel-enhanced image $I_{enhanced}$ equals the enhanced vessel layer plus the original XA image, as shown in Equation (3.22):

$$I_{enhanced} = \beta\, I_{vessel} + I_{original} \tag{3.22}$$

The results were evaluated using CNR. The layer separation method we used in this Section is OR-PCA (SW+CF).

For synthetic XA data, we used the parameters obtained from the leave-one-out evaluation in Section 3.3.4 for each synthetic data. For the pig XA data, the parameter set $(\lambda_{1,2}, r, t_0) = (2.1, 5, 3)$ was used.

3.3.6 Implementation

All algorithms were implemented in MATLAB (The MathWorks, Inc.). In particular, the computation time of layer separation was recorded in MATLAB 2014a on an Intel Core i7-4800MQ 2.70 GHz computer with 16 GB RAM running Windows 7.

3.4 Results

3.4.1 Optimal Parameters for OR-PCA

The parameters of OR-PCA optimized over the whole XA dataset for the two different basis update methods are shown in Table 3.1. Here both λ_1 and λ_2 are set to the same value. Comparing the two methods, the $\lambda_{1,2}$ have similar values, while the intrinsic ranks r of the subspace are very different.

Table 3.1: The optimal parameter settings of OR-PCA for different subspace basis update methods. p is the dimension of the data, i.e. the number of pixels in a frame.

Basis Update Method	$\lambda_{1,2}$	r
Block-coordinate Descent (BCD)	$2.3/\sqrt{p}$	14
Closed-form Solution (CF)	$2.1/\sqrt{p}$	5

Using block-coordinate descent with the optimal parameter setting, an example of online layer separation of an XA sequence (512×512, 55 frames) is shown in Fig. 3.5. Note that the layer separation result for the first frame shows strong artefacts (e.g. the vertebral shape in the vessel layer) due to random initialization of the subspace basis. As time proceeds, the layer separation improves quickly. The 10^{th} frame already has a good layer separation.

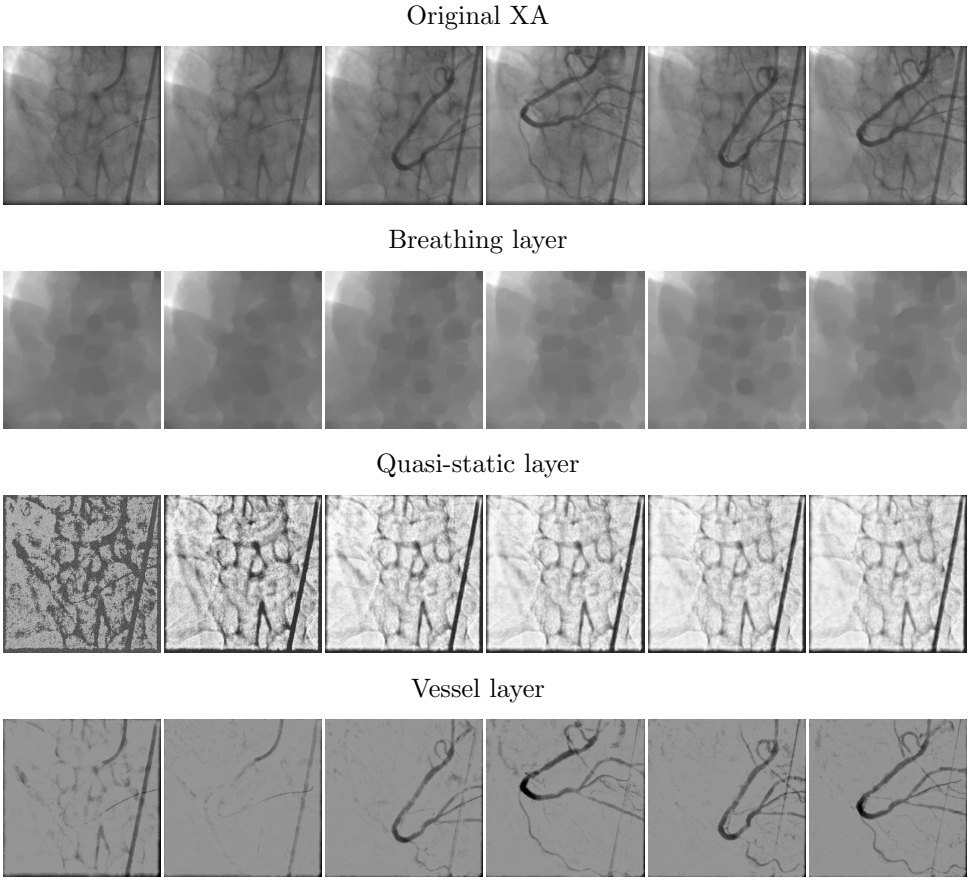


Figure 3.5: An example of online layer separation of an XA sequence using OR-PCA with the optimal parameter settings as listed in Table 3.1. The subspace basis update method used here is block-coordinate descent (BCD). Row 1-4 show the original frames, the breathing layer, the quasi-static layer and the vessel layer, respectively. Column 1-6 are 6 frames taken from the sequence in a chronological order and their layer separation outcomes. The frame ordinals from left to right are 1^{st}, 5^{th}, 10^{th}, 15^{th}, 20^{th}, 25^{th}.

3.4.2 Influence of the History Parameters

With the optimal parameter setting of OR-PCA, we quantitatively assessed how the history parameters mentioned in Section 3.2.3.3 and 3.3.3 influence the layer separation performance. Depending on the image content, the influence of the history parameters on each individual sequence may vary from sequence to sequence. The average measures over the whole dataset are shown in Figure 3.6 where the CNR values are normalized by dividing CNR by the CNR value obtained from the RPCA method (so RPCA has a constant value 1).

Compared to the original OR-PCA method (OR-PCA (BCD)) and OR-PCA (CF), the history parameters ϵ, ρ and t_0 all resulted in overall higher average local and global CNR. This improvement was more prominent in the case of global CNR compared to local CNR. It is also worth noticing that the combination of the history parameters with the CF method generally performed better than the (history parameter + BCD) option, which could be especially seen in the sliding-window case.

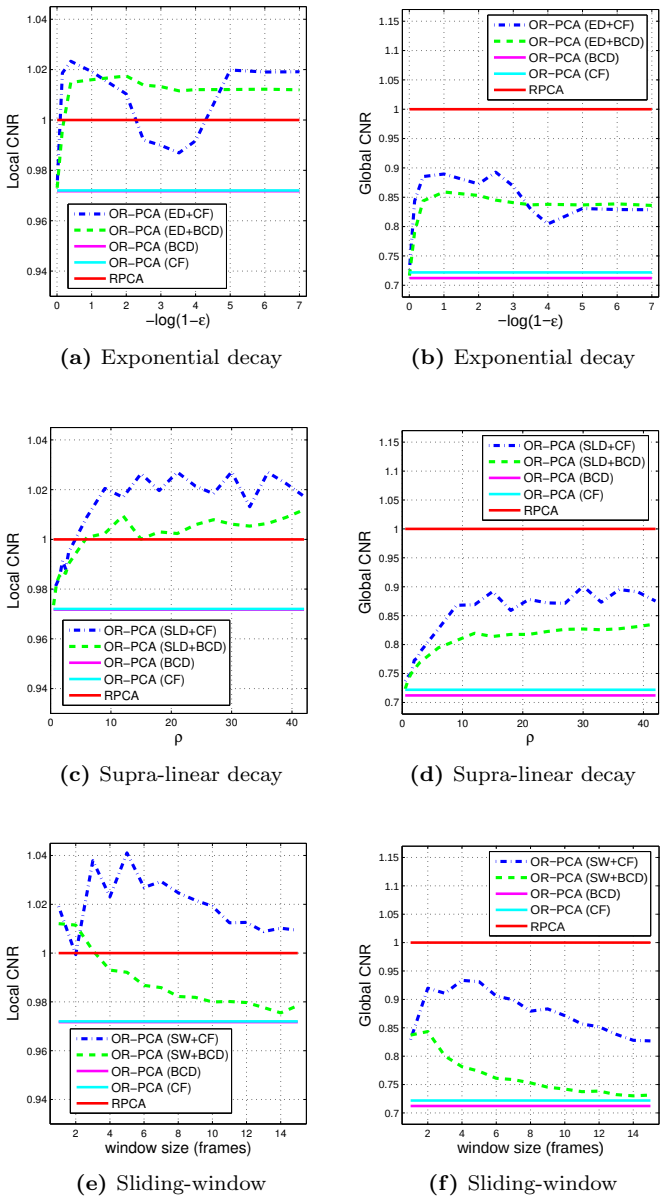
In addition to the overall comparisons of the three approaches, each of them presented a certain trend of CNR as the history parameter changed. For exponential decay (Figure 3.6a and 3.6b), as $-log(1-\epsilon)$ increases, the CNR values firstly increased fast; once they reached an optimal value when $-log(1-\epsilon)$ is between 0.5 and 2.5 (except for OR-PCA (ED+CF)), the CNRs dropped down slowly to reach a constant. In the case of supra-linear decay (Figure 3.6c and 3.6d), the CNRs increased as ρ increases, but did not change much when ρ is larger than 15. For the sliding-window approach (Figure 3.6e and 3.6f), although in general the CNRs dropped when the window size becomes larger, OR-PCA (SW+BCD) reached its optimum when t_0 equals 2, whereas OR-PCA (SW+CF) had an optimal t_0 between 3 and 5. Finally, as $-log(1-\epsilon)$ and ρ kept increasing (decreasing the weights of all past frames), the CNR curves of ED or SLD converged to where the CNR curves of SW started (only preserve information of the new frame).

3.4.3 Comparison with Other Methods

The optimal parameter sets for the methods based on OR-PCA that were obtained during the leave-one-out evaluation are listed in Table 3.2. In general, the methods that use closed-form solution (CF) had a smaller r than methods with BCD, but needed more information from the past data (lower ϵ and ρ, higher t_0) to achieve the best performance.

Table 3.2 also shows the counts of the optimal parameter sets that each LOO session generated for each method. For methods without history parameters (BCD and CF), the optimal parameter set with the largest count were identical to the results in Section 3.4.1. Most of the methods with a history parameter had a dominant optimal parameter set from LOO, except for the method (SW+CF) where the two most dominant optimal parameter sets had almost equal counts.

Table 3.3 lists the average CNR values for the original XA sequences and the vessel layers obtained with each method. Compared to the original XA, all methods achieved a substantial improvement on the CNR values in the vessel layer. Compared to method MS, MC+MS, RPCA (SW) and RPCA (SW)*, the methods that use OR-



(a) Exponential decay **(b)** Exponential decay

(c) Supra-linear decay **(d)** Supra-linear decay

(e) Sliding-window **(f)** Sliding-window

Figure 3.6: The influence of history parameters ϵ, ρ and the sliding-window size t_0 on local and global CNR. All values are normalized using RPCA method as the reference. In (a), (c) and (e), the local CNR for OR-PCA (BCD) and OR-PCA (CF) are very close that the two lines almost overlap.

Table 3.2: Counts of the optimal parameter sets obtained during the leave-one-out evaluation for each OR-PCA method. As 42 sequences are used, that yields 42 leave-one-out sessions and 42 optimal parameter sets in total.

Method	$\lambda_{1,2}$ $(1/\sqrt{p})$	r	History Parameter	Counts
BCD	2.3	14		39
	2.4	17		2
	3.3	19		1
CF	2.1	5		31
	2.3	7		11
ED+BCD	2.3	14	$\epsilon = 0.9$	39
	2.4	17	$\epsilon = 0.9$	2
	3.3	19	$\epsilon = 0.9$	1
ED+CF	2.1	5	$\epsilon = 0.6$	31
	2.3	7	$\epsilon = 0.6$	8
	2.3	7	$\epsilon = 0.3$	3
SLD+BCD	2.3	14	$\rho = 42$	39
	2.4	17	$\rho = 18$	2
	3.3	19	$\rho = 9$	1
SLD+CF	2.1	5	$\rho = 36$	29
	2.3	7	$\rho = 36$	11
	2.1	5	$\rho = 15$	2
SW+BCD	2.3	14	$t_0 = 2$	38
	2.4	17	$t_0 = 2$	2
	2.3	14	$t_0 = 1$	1
	3.3	19	$t_0 = 2$	1
SW+CF	2.1	5	$t_0 = 3$	17
	2.1	5	$t_0 = 5$	14
	2.3	7	$t_0 = 5$	11

PCA (from BCD to SW+CF in Table 3.3) had higher CNR. The CNR values of the two types of methods (using or not using OR-PCA) were statistically significantly different with the two-sided Wilcoxon signed-rank test (see Table 3.7 in Supplementary). The methods that downweight the past data (from ED+BCD to SW+CF) were able to improve the vessel contrast over the methods without history parameters (BCD and CF). The improvement was statistically significant (see Table 3.8 in Supplementary). Among all the methods that are based on OR-PCA, ED+CF, SLD+CF and SW+CF showed similar or better average local CNR than RPCA, although without statistical significance. The performance of all OR-PCA based methods was closer to the off-line benchmark RPCA than those not using OR-PCA.

Table 3.3 also shows the average reconstruction error between the original image

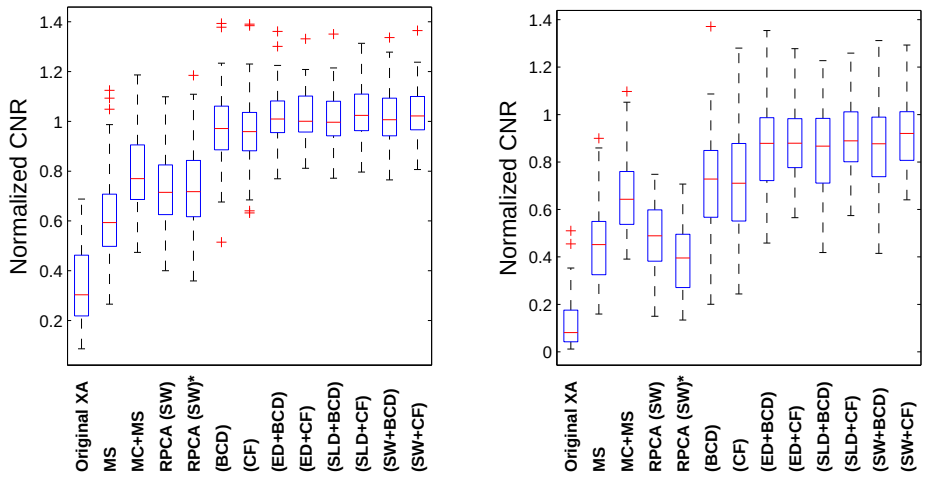
Table 3.3: The average local CNR, global CNR and reconstruction error E_{recon} (mean value ± standard deviation) for the original XA and all methods.

Method	Local CNR	Global CNR	E_{recon}
Original XA	0.991 ± 0.330	0.507 ± 0.305	
MS	1.811 ± 0.490	2.211 ± 0.648	0
MC+MS	2.396 ± 0.635	3.210 ± 0.869	0
RPCA (SW)	2.134 ± 0.511	2.278 ± 0.676	0
RPCA (SW)*	2.170 ± 0.578	1.881 ± 0.623	0
(BCD)	3.010 ± 1.065	3.453 ± 1.223	0.026 ± 0.005
(CF)	3.010 ± 1.088	3.509 ± 1.204	0.030 ± 0.007
(ED+BCD)	3.209 ± 1.192	4.257 ± 1.316	0.021 ± 0.006
(ED+CF)	3.226 ± 1.263	4.422 ± 1.466	0.023 ± 0.007
(SLD+BCD)	3.169 ± 1.172	4.085 ± 1.282	0.022 ± 0.006
(SLD+CF)	3.246 ± 1.224	4.500 ± 1.393	0.022 ± 0.006
(SW+BCD)	3.170 ± 1.161	4.150 ± 1.295	0.022 ± 0.006
(SW+CF)	3.281 ± 1.290	4.602 ± 1.465	0.022 ± 0.007
RPCA	3.227 ± 1.301	5.176 ± 2.004	0.007 ± 0.002

and the three-layer sum image for each method. For MS, MC+MS and the RPCA (SW) methods, the foreground (vessel layer) were obtained via subtraction of the background. therefore these methods, by definition, have a reconstruction error of zero. The methods based on OR-PCA made minor reconstruction errors (less than 3% of the average pixel intensity of the original images). The history downweighting techniques reduced the reconstruction errors of BCD and CF for about 17% and 26% respectively. These errors are about three times larger than the one made by RPCA.

The comparison between different methods is illustrated in Figure 3.7, where the CNR values were normalized in the same way as in Figure 3.6. Similarly, the methods that use OR-PCA outperformed the other methods on both local and global CNR. The improvement that results from downweighting history was more substantial in global CNR than local CNR. For the methods that are based on OR-PCA, those that use the closed-form solution achieved slightly better normalized CNR values than the ones with BCD.

Figure 3.8 presents examples of results of five representative prospective or on-line layer separation methods and original images on four XA sequences: MC+MS, RPCA (SW), OR-PCA (CF), OR-PCA (SW+CF) and RPCA. All methods improved the visibility of vessels in the image, but MC+MS generated strong artefacts near the guiding catheters. RPCA (SW) shows slightly better results than MC+MS, but still presents some motion artefacts near the guiding catheters. OR-PCA (CF) achieved good layer separation, while OR-PCA (SW+CF) was able to produce "cleaner" background (row 1 column 4 and 5, the left part of the images) and still maintained the vessel information (column 4 and 5). The vessel layer separated using OR-PCA (SW+CF) had more similar appearances to the ones produced by RPCA (column 5



(a) Normalized local CNR **(b)** Normalized global CNR

Figure 3.7: The boxplot that compares various methods on their performance of layer separation. In these diagrams, the CNR values of the mentioned methods are normalized by dividing their CNR by the CNR obtained with the RPCA method.

and 6) than other methods.

3.4.4 Vessel Enhancement in Low-Contrast XA

The results of vessel-enhancement on synthetic low-contrast XA and real pig XA data are shown in this section.

3.4.4.1 Synthetic Low-Contrast XA

Table 3.4 shows the average CNR values for vessel enhancement in synthetic XA sequences. With enhancement factor $\beta = 1$, the vessel-enhanced XA showed better local and global CNR than the synthetic XA with statistical significance. Compared to the original XA in Table 3.3, the CNR values of the vessel-enhanced XA was slightly lower, but these CNR values could be improved with a larger enhancement factor β.

A few examples of vessel-enhancement on synthetic XA data are shown in Figure 3.9. Compared to the original images (the first row), the synthetic XA (the second row) had poorer vessel contrast. The proposed layer separation method (SW+CF) was still able to extracted the vessel layer (the third row), while maintaining a majority of the information, and enhance the vessel contrast (the last row) to the visually similar level of the original images.

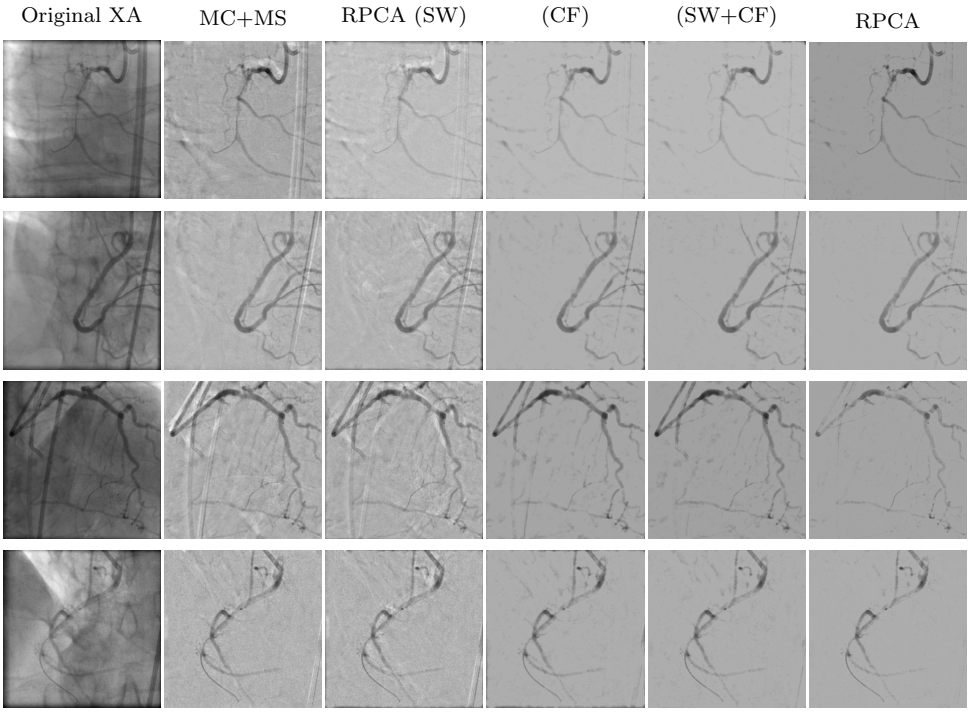


Figure 3.8: Comparison of five different layer separation methods on four example XA sequences. One representative frame is selected from each sequence to visualize the results. Row 1-4 show four sequences. Column 1 is original XA, column 2-6 are the separated foreground (vessel layer) obtained from MC+MS, RPCA (SW), OR-PCA (CF), OR-PCA (SW+CF) and RPCA.

3.4.4.2 Real XA of Pigs

We show the CNR values for vessel enhancement experiment with pig XA data in Table 3.5. In the table, since the full-contrast sequence showed lower visual contrast (see Section 3.3.1.3), the four sequences were sorted by their local CNR values in an ascending order. With enhancement factor $\beta = 2$, the vessel layers and the enhanced sequences showed an improvement on local and global CNR. This improvement increased as the local CNR of the sequence became higher.

Similar observation could be found in Figure 3.10, where representative frames from each pig XA sequence are shown. For example, the proposed method was able to increase the vessel contrast in the image of 25% contrast to the similar level as in the image of 50% contrast (see Figure 3.10i and 3.10b). The vessel contrast in the enhanced image of 50% contrast (Figure 3.10j) had better contrast than the image of full contrast (Figure 3.10c) and 75% contrast (Figure 3.10d). It is also observed that, in Figure 3.10 from left to right, the false positive enhancement of non-vessel structures (e.g. the dark spots in the right part of the images) decreased as the visual

Table 3.4: The average local and global CNR (mean value ± standard deviation) for the synthetic XA data, the vessel layers separated from the synthetic data using OR-PCA (SW+CF), and the vessel-enhanced XA sequences ($\beta = 1$). The two-sided Wilcoxon sign-rank test indicates statistically significantly difference in local and global CNR between the synthetic data and the vessel-enhanced images ($p < 0.01$).

Image types	Local CNR	Global CNR
Synthetic XA	0.592 ± 0.236	0.338 ± 0.245
Vessel Layer	3.170 ± 1.290	4.048 ± 1.592
Enhanced	0.875 ± 0.312	0.452 ± 0.285

Table 3.5: The local and global CNR of the 4 XA sequences obtained from pigs, the separated vessel layer and their vessel-enhanced sequence ($\beta = 2$). The 4 sequences are sorted by their local CNR values in an ascending order.

Contrast Concentration	25%		50%		100%		75%	
Metric (CNR)	local	global	local	global	local	global	local	global
Original	0.307	0.164	0.490	0.340	0.523	0.395	0.690	0.569
Vessel Layer	2.546	0.682	5.156	0.830	5.175	2.759	6.835	5.473
Enhanced	0.488	0.262	0.954	0.562	1.041	0.729	1.528	1.147

contrast increased.

3.4.5 Computation Time

The computation time of layer separation for each frame is shown in Figure 3.11. In this figure, the box plots of per-frame processing time for each method that is based on OR-PCA is illustrated. The processing times of these methods ranged from 0.15 to 1.60 seconds per frame. The methods that use a closed-form solution to update L were approximately two times faster than the ones that use block-coordinate descent. The methods which use ED and SLD to treat past information were slightly faster than their counterparts that do not weight past data, while the methods with SW needed slightly longer time to process one frame than their corresponding methods BCD and CF. The "outliers" shown as red marks in Figure 3.11 are from images of larger size.

The average computation time per image size is shown in Table 3.6. Generally, the table shows that images of larger size needed longer processing time per frame. On XA images of common size in clinics (512 × 512, 600 × 600 and 776 × 776), it is possible to achieve a processing rate of 3-6 frames per second (fps) with the proposed methods on our hardware.

Compared to MS and MC+MS, OR-PCA based methods showed much better layer separation performance (see Table 3.3 and Figure 3.7), while the processing time of the fastest one (ED+CF) is only about 40% slower than MC+MS.

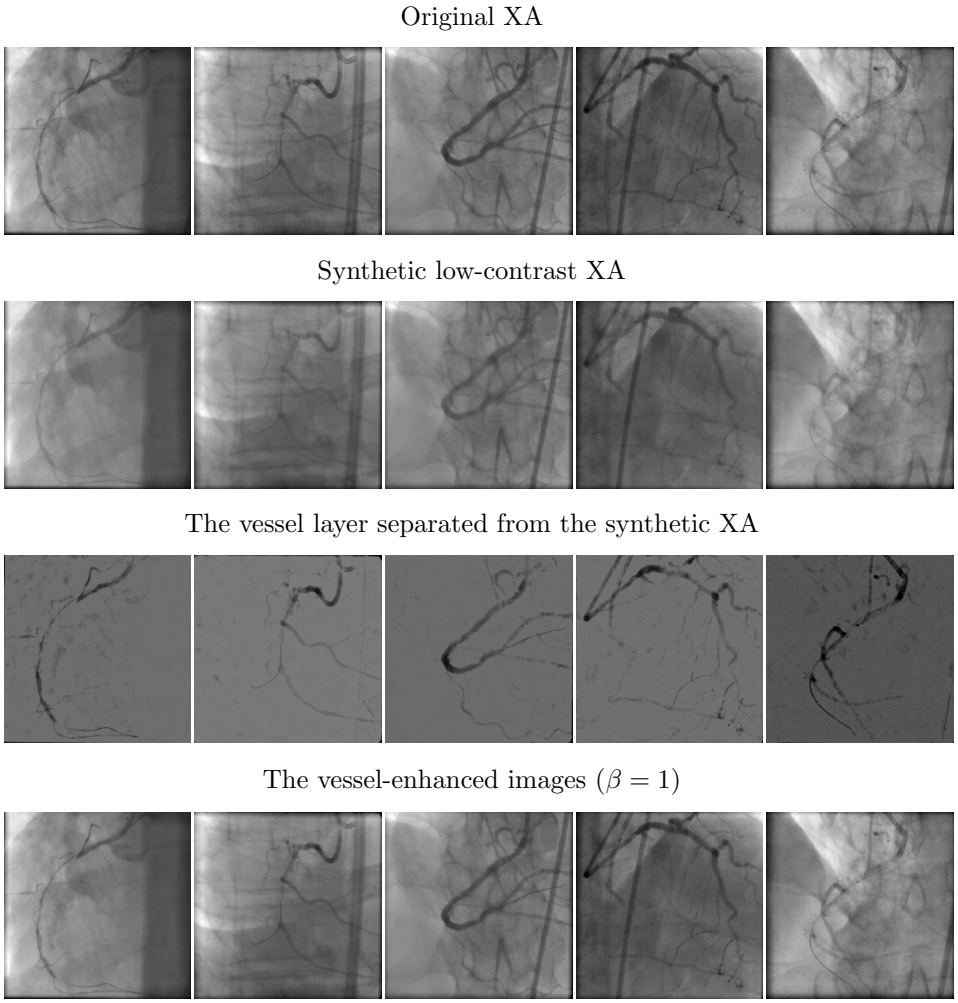


Figure 3.9: Five examples of vessel enhancement on synthetic low-contrast XA images (Column 1-5).

The processing time for RPCA (SW) is also shown in Table 3.6. In RPCA (SW), the foreground separation of a frame is delayed by the processing of its previous data block, therefore, we computed the average per frame processing time of all blocks in a sequence. Compared to RPCA (SW), the methods that use OR-PCA perform faster with a factor up to 3.

3.5 Discussion and Conclusion

We have presented a fast automatic online method to robustly separate cardiac interventional X-ray angiograms into three image layers: a breathing layer, a quasi-static

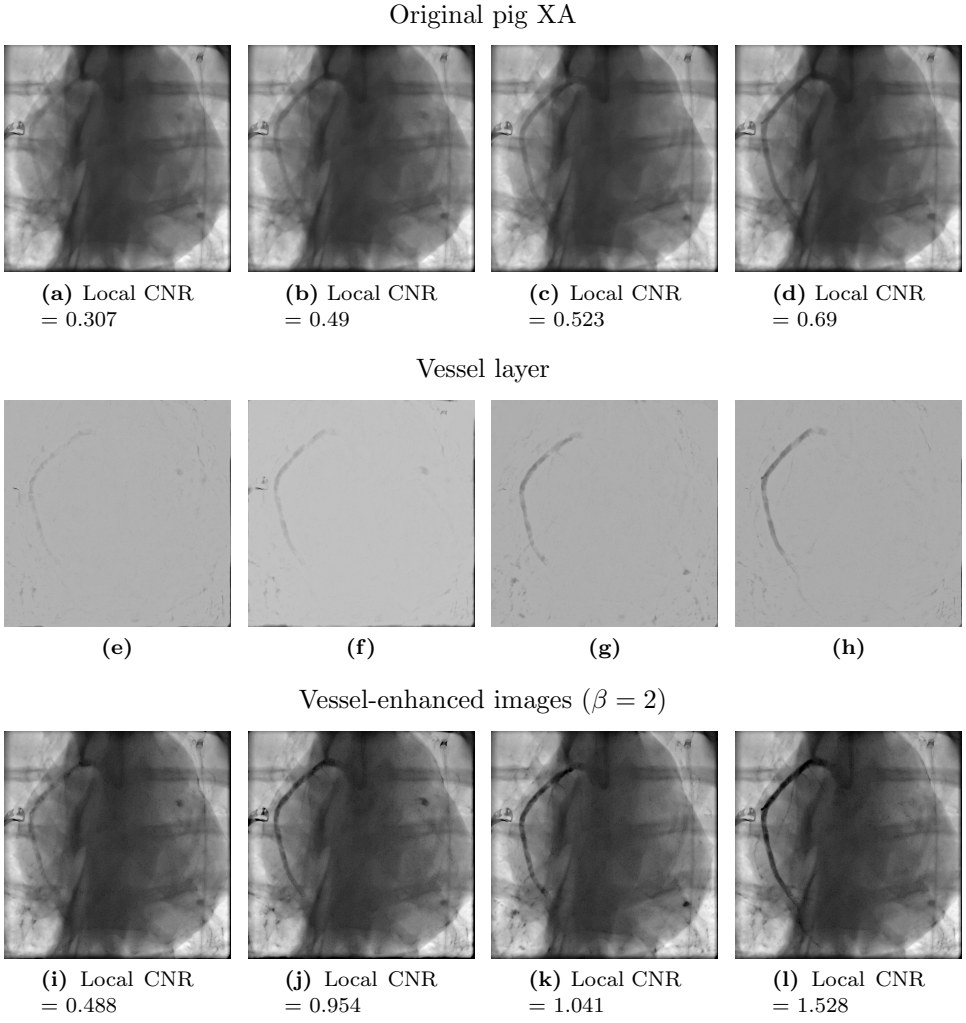


Figure 3.10: Vessel enhancement on pig XA sequences with different level of contrast agent. From left to right, the contrast concentration used for the sequence are 25%, 50%, 100% and 75%, respectively, while the local CNR increases. Row 1: the original pig XA sequences. Row 2: the separated vessel layer. Row 3: the vessel-enhanced images ($\beta = 2$).

layer and finally a vessel layer that contains information of moving thin structures, such as coronary arteries. The method relied on morphological closing and online robust PCA and we investigated different possibilities for downweighting information from previous frames for further improving layer separation. The parameters of OR-PCA were optimized on 42 clinical XA sequences. In addition, a pilot study was performed on synthetic XA sequences and pig data to show the potential of the proposed method for vessel enhancement in XA.

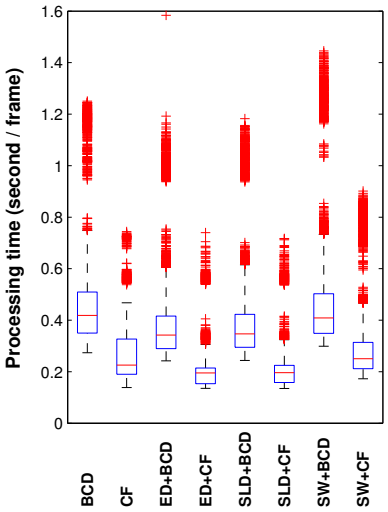


Figure 3.11: The processing time that each method that uses OR-PCA needs for layer separation. The per frame processing time (second) of every single frame in the whole dataset is shown as box plots.

The integration of OR-PCA algorithm into layer separation enables online processing XA images from the beginning of the sequence. The mechanism behind this is that OR-PCA only needs to be "fed" one frame each time, but is able to update the subspace basis of the low-rank component based on the new information. This is an important difference from the method in [65] (Chapter 2 of this thesis) which worked "off-line" and needed the complete sequence as input. The proposed approach is also different from the method in [109], where the online implementation needed several frames to solve RPCA in a mini-batch manner, and hence, resulting in a delay in processing for each following mini-batch. Although the layer separation results of our proposed method might suffer from random initialization at the beginning, the algorithm converges fast and obtain reasonable layer separation after a few frames (see an example in Figure 3.5).

In addition to the advantage of online processing, the methods that use OR-PCA show good performance on layer separation. It significantly improves the vessel visibility of the original XA images with minor reconstruction errors: the background structures were removed and the vessel contrast was visually and quantitatively enhanced. Compared to those methods that model a total static background, e.g. MS and MC+MS, the approaches that are based on OR-PCA are superior because they are able to model a dynamic scene, therefore can adapt to small dynamic changes in the background. The method that separates layers by solving RPCA with a mini-batch of data suffers from motion artefacts that remain around vessels and catheters.

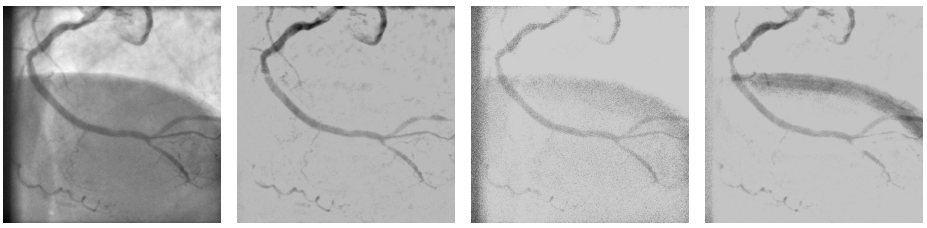
Table 3.6: The average processing time (seconds/frame) for each layer separation method on XA sequences of different frame size. The bold numbers indicate processing time that the fastest OR-PCA based method needs for a certain frame size.

Method	512 × 512	600 × 600	776 × 776	1024 × 1024
BCD	0.36	0.51	0.69	1.14
CF	0.20	0.30	0.36	0.58
ED+BCD	0.30	0.41	0.63	1.02
ED+CF	**0.16**	**0.21**	**0.34**	**0.57**
SLD+BCD	0.30	0.42	0.63	1.02
SLD+CF	0.17	0.22	**0.34**	**0.57**
SW+BCD	0.36	0.50	0.75	1.27
SW+CF	0.22	0.31	0.46	0.77
RPCA (SW)	0.37	0.53	0.85	1.50
RPCA (SW)*	0.46	0.68	1.10	1.97
MS	0.02	0.03	0.05	0.11
MC+MS	0.11	0.15	0.25	0.45

This might be because it uses the frames in the previous block to infer the background for the current block of images, which might fall behind the background change. Furthermore, it uses the same background for all images in the block to compute their vessel layers, which ignores the possible small background change within the block. OR-PCA updates L frame by frame to keep up the background change and has a "customized" background for each individual frame.

It is also worth noticing the advantage of removing breathing structures prior to the separation of the other two layers with OR-PCA. If the original XA image is directly "fed" to the OR-PCA method, the breathing structures should stay in the same layer with the static structures in order to obtain a reasonable vessel layer, because OR-PCA decomposes a source image into only two components. Due to breathing motion, this background layer will contain more variation than a layer that only contains quasi-static structures, which might require a much higher r parameter of OR-PCA to allow reasonable convergence of the algorithm. However, it is often inevitable to observe strong breathing motion artefacts in the output vessel layer. Figure 3.12 provides an example that illustrates the cases without removing breathing structures prior to the OR-PCA computation. In Figure 3.12c, the image still contains a large amount of noise and a static dark band on the left. In Figure 3.12d, OR-PCA has a better convergence with a higher r, less noise and no dark band is observed, but a stronger artefact of diaphragm remains.

The parameters of OR-PCA used in this work were optimized based on our image data, instead of being assigned the "rule of thumb" value $1/\sqrt{p}$ as [53, 98]. The optimum of $\lambda_{1,2}$ for method BCD and CF are similar and the values are close to $1/\sqrt{p}$. The optimal intrinsic rank r for BCD and CF are different. A possible explanation is that CF computes the quasi-static subspace basis L in one step, whereas BCD updates



(a) (b) (c) (d)

Figure 3.12: An example that compares the cases without removing breathing layer before the OR-PCA operation to the one resulted from the proposed method: (a) original image; (b) the vessel layer obtained with OR-PCA (BCD) using the proposed method and parameters; (c) the vessel layer obtained without the separation of breathing layer using OR-PCA (BCD) with the same parameters as the case in (b); (d) the vessel layer obtained without removing breathing structures using a higher r value ($r = 50$) for OR-PCA (BCD). Strong artefacts due to breathing motion can be observed in (c) and (d).

L column-wisely and thus needs more variations to achieve the same accuracy as CF.

The performance of layer separation using OR-PCA can be improved by down-weighting the past frames. In this work, we have proposed three different ways: exponential decay (ED) and supra-linear decay (SLD) are methods to scale all past data, and the sliding-window approach (SW) only preserve the information of the few most recent frames, which could be interpreted as "binary" scaling. The results on vessel visibility and reconstruction error showed that all three ways improve the overall layer separation by giving recent frames higher weight than earlier frames. This suggest that not all past information is necessary for best inferring the current status in the scenario of online learning. A possible explanation of this finding is that using the downweighting techniques in the online learning algorithm promotes faster convergence [70]. Although the improvement might vary between sequences, depending on specific image content, they overall present an improvement with history parameters.

The optimized history parameters show that only the most recent 2-5 frames are needed to update the subspace basis L, which seems too "few". The reasons might be two-fold. First, since L is the subspace basis of the quasi-static layer which does not contain much variation, it should not need information from a large number of frames to update L. Second, for the case of BCD, note that in Equation (3.10), the update of L still partly relies on its previous version, not solely on A and B. This means that every version of L can inherit information from its previous version and thus is a compact representation of all past information, but updating L needs only the very recent frames.

The combination of the subspace basis update methods and the past information downweighting techniques yields 8 different OR-PCA variants, while in practice one might choose one of them for a layer separation task. In terms of the performance on improving vessel visibility, (SW+CF) might be the best choice, as it gives the highest CNR value. On the other hand, if speed is of great concern, (ED+CF) is a good

option, since it runs the fastest among the eight and the performance on CNR is not much worse than (SW+CF). In addition, the implementation of (ED+CF) is easier than (SW+CF), in that it does not explicitly store a few past values of A and B, but does the scaling implicitly.

In terms of computational efficiency, the methods that use OR-PCA run fast. For 512×512 frames, OR-PCA with CF was able to achieve a 5-6 fps processing rate on standard PC; for 1024×1024 frames, the proposed methods could reach at highest about 2 fps. This is faster than RPCA (SW) based approaches either from the result of our experiment (see Table 3.6) or the literature, e.g. [109] which reported about 1 fps for frame size in the range of 824×1024 to 1024×1024, and [21] which achieved 3 fps for 512×512 images. According to [35], the computation complexity for batch RPCA is $O(np^2)$ and for OR-PCA is $O(pr^2)$. Since $p \gg r$ and normally also $p \gg r^2$, OR-PCA runs much faster than RPCA. We also notice that OR-PCA with CF ran faster than BCD, this is because the computation complexity of both methods is $O(pr^2)$, and in our experiments, the r for CF was smaller than the r for BCD. The methods with history downweighting schemes are faster than BCD and CF. This might be due to the faster convergence of OR-PCA when downweighting the past information. Finally, it is worth noticing that the timing reported in this paper were based on a MATLAB implementation that ran on a single CPU core. A parallelized version of the method may achieve real-time processing rate (about 15 fps) for clinical applications.

One of the potential direct clinical applications of the proposed method is to enhance vessels in X-ray images with low vessel contrast, which suffer from poor diagnostic quality. X-ray contrast agent used for angiography may have side effects including allergic reaction which can be life-threatening, and nephrotoxicity (contrast-induced nephropathy, CIN) which can result in chronic renal failure with all its sequelae [10, 105]. Thus, it is clinically relevant to limit the use of X-ray contrast agent during interventions. The method we proposed in this paper provides a possibility to use X-ray contrast of lower concentrations. In the experiments, we have evaluated whether it is possible to achieve better vessel visibility on low-contrast images by enhancing vessels using the methods based on OR-PCA. We have used synthetic 50%-contrast XA images and 4 real pig XA sequences with different contrast concentrations for the test. The results showed a good improvement on the vessel visibility on both kinds of images, implying a potential application for coronary interventions.

The proposed layer separation methods are based on some assumptions. First, the morphological closing operation with a circular structuring element of 8.5 *mm* in diameter worked well for small vessels, such as coronary arteries. However, for other kinds of interventions that operate on large vessels, such as aorta or pulmonary arteries, the structuring element of the proposed size is not large enough. In those cases, one might consider using a larger structuring element for morphological closing and adjusting the parameters of OR-PCA for a reasonable layer separation. Another important assumption underlying the methods that use OR-PCA is that there is dynamic change in the foreground, and it detects the dynamic change. This assumption holds true most of the time because coronary arteries always move together with heartbeat. However, in the case that the guiding catheter tip segment moves together with heartbeat, the proposed layer separation method cannot separate this

moving catheter segment from vessels. The method also requires a certain amount of contrast agent, i.e. the signal of the vessels should not be too weak. As the methods based on OR-PCA assumes a sparse foreground, when the contrast of vessel is not strong enough, the proposed methods might enhance noise or detect other non-vessel structures in the foreground, as can be seen in Figure 3.10.

In the future, it is of great interest to investigate the potential of the proposed method. One important direction would be to evaluate the clinical potential, e.g. how the proposed layer separation method would work under different contrast concentration levels on a larger dataset. One could also think of improving the visibility of instruments, such as catheters or guidewires for other cardiac applications. From a methodological point of view, it might be interesting to unify the steps in the proposed method into one optimization problem, for example, incorporating the heuristic post-processing into the OR-PCA algorithm with an additional non-positive constraint on the sparse component.

In conclusion, we have presented a fast automatic online layer separation method for robust vessel enhancement in X-ray angiograms. The method separated an XA frame into three layers: a breathing layer, a quasi-static layer and a vessel layer. We proposed three ways to improve the layer separation outcome by downweighting the past frames. The proposed method significantly improved the vessel visibility and outperformed other related prospective or online layer separation approaches. The method does not need much computation time, making it potentially applicable for clinical practice without the necessity of using advanced hardware, opening the way for relevant clinical applications, such as improving the vessel visibility under conditions of low contrast concentrations, so as to allow a reduced amount of contrast agent usage to prevent contrast-induced side effects.

Acknowledgement This work was supported by Technology Foundation STW, IMAGIC project under the iMIT program (grant number 12703).

Supplementary

In support of the results in section 3.4.3, the complete statistical test results for the differences of CNR for different methods are shown in the supplementary in Table 3.7 and Table 3.8.

Table 3.7: The p values of the two-sided Wilcoxon signed-rank test between OR-PCA and non-OR-PCA approaches.

Non-OR-PCA →	Original XA	MS	MC+MS	RPCA (SW)	RPCA (SW)*	RPCA
OR-PCA approaches ↓	p value for local CNR					
BCD	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	0.057
CF	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	0.024
ED+BCD	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	0.788
ED+CF	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	0.722
SLD+BCD	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	0.750
SLD+CF	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	0.269
SW+BCD	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	0.812
SW+CF	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	0.109
OR-PCA approaches ↓	p value for global CNR					
BCD	< 0.001	< 0.001	0.073	< 0.001	< 0.001	< 0.001
CF	< 0.001	< 0.001	0.022	< 0.001	< 0.001	< 0.001
ED+BCD	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
ED+CF	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
SLD+BCD	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
SLD+CF	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
SW+BCD	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
SW+CF	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001

Table 3.8: The p values of the two-sided Wilcoxon signed-rank test between CNR of OR-PCA approaches with history parameters and that of OR-PCA without treating the past information.

With history parameter →	ED+BCD	ED+CF	SLD+BCD	SLD+CF	SW+BCD	SW+CF
Without history parameter ↓	p value for local CNR					
BCD	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
CF	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
Without history parameter ↓	p value for global CNR					
BCD	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
CF	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001

Chapter Four

PCA-derived Respiratory Motion Surrogates From X-ray Angiograms For Percutaneous Coronary Interventions

Abstract — *Purpose* Intraoperative coronary motion modelling with motion surrogates enables prospective motion prediction in X-ray angiograms (XA) for percutaneous coronary interventions. The motion of coronary arteries is mainly affected by patients breathing and heartbeat. Purpose of our work is therefore to extract coronary motion surrogates that are related to respiratory and cardiac motion. In particular, we focus on respiratory motion surrogates extraction in this paper.

Methods We propose a fast automatic method for extracting patient-specific respiratory motion surrogate from cardiac XA. The method starts with an image preprocessing step to remove all tubular and curvilinear structures from XA images, such as vessels, guiding catheters, etc., followed by principal component analysis (PCA) on pixel intensities. The respiratory motion surrogate of an XA image is then obtained by projecting its vessel-removed image onto the first principal component.

Results This breathing motion surrogate was demonstrated to get high correlation with ground truth diaphragm motion (correlation coefficient over 0.9 on average). In comparison to other related methods, the method we developed did not show significant difference ($p>0.05$), but did improve robustness and run faster on monoplane and biplane data in retrospective and prospective scenarios.

Conclusions we developed and evaluated a method in extraction of respiratory motion surrogate from interventional X-ray images that is easy to implement and runs in real-time, and thus allows extracting respiratory motion surrogates during interventions.

Based upon: **H. Ma**, G. Dibildox, C. Schultz, E. Regar and T. van Walsum: PCA-derived Respiratory Motion Surrogates From X-ray Angiograms For Percutaneous Coronary Interventions. *International Journal of Computer Assisted Radiology and Surgery*, vol. 10 (6), pp. 695-705, 2015.

4.1 Introduction

Percutaneous coronary intervention (PCI) is a minimally invasive procedure for treating patients with advanced coronary artery disease. With this technique, a catheter system is introduced into patients' circulation system through their femoral or radial artery under local anesthesia. A preshaped guiding catheter is positioned into the ostium of the coronary artery. Through this catheter, a guide wire serving to deploy devices, such as balloon catheters and stents, is advanced into the branches of the artery. Once a stenosis is targeted, the balloon is deployed at the lesion site to fix the vessel blockag [42].

PCI is normally performed in a catheterization lab under guidance of X-ray angiography (XA) that coronary arteries are opacified with contrast agent. However, due to contrast agent toxicity, its injection times are limited, such that guide wire and device advancement into the target lesion is performed under "vessel-free" images. In this situation, interventional cardiologists have to mentally reconstruct the position of coronary arteries and stenosis based on previous images, which increases the risk of failure for the procedures, especially in difficult cases.

To address this problem, Shechter et al. proposed to model the coronary motion with surrogate signals from contrast-enhanced images and hence the guidance in "vessel-free" images becomes feasible by prospective motion correction with such a model [96]. As the motion of coronary arteries is mainly affected by patient's breathing motion and cardiac heart beat, it is reasonable to model the coronary motion with surrogates which are correlated with patients' respiratory and cardiac motion. With such an aim, we have been focusing on respiratory-induced coronary motion modeling. Therefore, the purpose of our study was to develop and evaluate a method for fast and robust extraction of respiratory motion surrogates from X-ray angiograms for PCI.

Related works on respiratory motion surrogates have been reported. Signals from external apparatus, such as navigators or bellows, have been used in many studies for respiratory motion modeling [72]. Usage of image-based surrogates have also been investigated. One commonly used surrogate is diaphragm superio-inferior (SI) motion [96] [69] [56]. This is extracted by drawing a rectangular ROI on diaphragm border followed by manual tracking the diaphragm or automatic calculating the 1-D translation. These methods involve human interaction to draw such a ROI and hence not entirely automatic. Automatic diaphragm detection and tracking were reported in [82] [28]. These methods use morphological operation to preprocess XA images followed by a 2nd-order curve fitting to the diaphragm border. Studies on other respiratory-related objects, e.g. coronary sinus catheter and tracheal bifurcation, can be seen in [69] [77]. These methods require specific objects being present in images, which is not always the case in XA images for PCI. Dimension reduction techniques have been used for studying respiratory motion as well. In [78], an automatic method based on principal component analysis (PCA) was designed for retrospective motion gating. This method first creates a mask using Hessian-based vesselness filtering and analyzes pixels inside the mask with PCA technique. In another study [18], hierarchical manifold learning was used to find correlation between image regions and

respiratory motion.

In this work, we developed a real-time, PCA-based method for extracting a respiratory surrogate from coronary XA sequences. Our contributions are three-fold: first, we develop a method that is simple to implement and runs in real-time on common PC hardware; second, we evaluate the method on several clinical datasets, comparing the results with manual ground truth and existing methods; third, we assess the usability of the method on monoplane and biplane image data in both retrospective and prospective scenarios.

4.2 Methods

Coronary motion analysis in frames of XA sequence is complicated by the existence of both respiratory and cardiac motion in images. Therefore, respiratory motion surrogate extraction could possibly benefit from elimination of the objects representing cardiac motion from XA images. In this situation, respiratory motion becomes the major source of intensity change in XA sequence and could be analyzed with methods having source decomposition capability, such as principle component analysis.

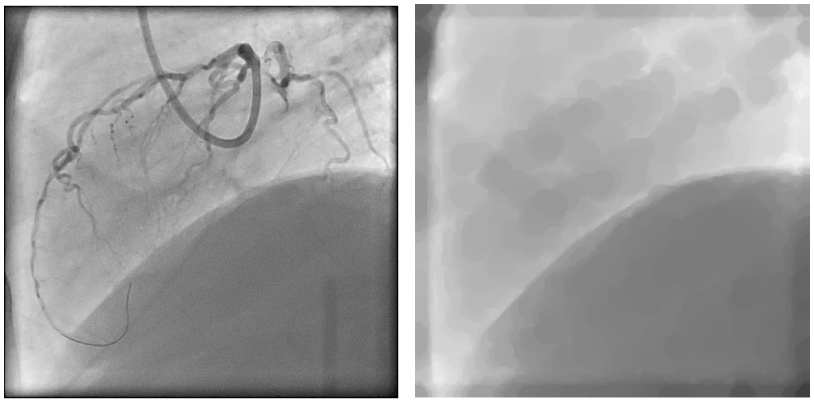
To give an overview, our proposed method consists of two major steps. First, images are downsampled and processed with morphological-closing to remove coronary arteries, guiding catheters, etc.. Next, pixel intensity changes in the "vessel-removed" images are analyzed with principal component analysis to extract respiratory motion information. Each of the steps is explained in more detail in the next sections.

4.2.1 Preprocessing of XA Images

First, each frame of the sequence is downsampled. Depending on the original image size, the downsampling factor is chosen to be 4 if original size is 512×512 or 600×600, or 8 if previous size is 1024×1024. This operation converts the original frame to an image of size 128×128 or 150×150, which already allows fast processing in later steps and still preserves enough original information that we need for subsequent analysis.

Next, as we are interested in respiratory motion only, we remove structures that show cardiac motion. To this end, similar to [82], a morphological closing is applied to the downsampled image with a circular structuring element in order to remove any tubular and curvilinear structures, such as coronary arteries, guiding catheters, guide wires and stitches. The size of the structuring element is chosen based on the maximal diameter of coronary arteries and guiding catheters. Dodge reported that the lumen diameter of the left main artery measures 4.5 ± 0.5 mm [30]. In another study using transthoracic echocardiography [81], the average wall thickness of left anterior descending artery was 1.1 ± 0.2 mm and its external elastic membrane diameter is 4.5 ± 0.9 mm. Having a maximum coronary diameter of 5-7 mm, and a maximum magnification of 1.5, we use a structuring element of around 11 mm in diameter (roughly 7-8 pixels in radius in the downsampled images) to remove the curvilinear structures. This size is shown to be adequate and guarantee a complete removal of vessels and guiding catheters from our images.

With the mentioned operations, objects representing cardiac motion are effectively



(a) (b)

Figure 4.1: Morphological closing operation on an XA image. (a) Original XA. (b) Image processed with morphological closing: guiding catheter and coronary arteries are removed.

removed from downsampled XA images while the diaphragm border and lung tissues still remain. Morphological closing might cause circular artifacts which, however, have lower contrast than the arteries, which is sufficient to prevent the subsequent analysis from being "contaminated" by cardiac motion. An example morphological-closed image is showed in Fig 4.1.

4.2.2 Principal Component Analysis

Principal component analysis (PCA) is typically used for dimension reduction. It transforms a multivariate dataset to a new orthogonal coordinate system such that most variance of this dataset could be represented by a few coordinates. Hence reducing its dimension is normally achieved by preserving only a few coordinates in the new coordinate system without losing much information [16].

Similar to [78], we first use the PCA technique on morphological-closed images to obtain principal components for each sequence. Representing a frame of an XA sequence with an $n \times n$ matrix, we concatenate each pixel in such matrix into a single column vector $\mathbf{x}^i$, whose size is $D \times 1$, where $D = n^2$. Thus an XA sequence consisting of N frames is represented as a $D \times N$ matrix $X = [\mathbf{x}^1, ..., \mathbf{x}^N]$. We then center X to obtain a zero mean matrix. Without losing generality, we still write the zero mean matrix as X. Seeking the principal components of X is equivalent to computing the eigenvectors of covariance matrix XX^T, which is a $D \times D$ matrix. As D is usually a large number and in our case $D >> N$, we adopt the approach from [16] to apply eigen analysis to the $N \times N$ matrix $X^T X$. Then we have

$$E = X\tilde{E}\Lambda^{-1}, \tag{4.1}$$

where E is the $D \times N$ matrix of eigenvectors of XX^T, $\tilde{E}$ is the $N \times N$ matrix of eigenvectors and Λ is the $N \times N$ diagonal matrix of eigenvalues of $X^T X$. With this approach, we benefit computation efficiency from computing the eigenvectors of a smaller matrix $X^T X$. Next, we project the XA sequence on the first principal component $\mathbf{e}_1$ by computing

$$\mathbf{p} = X^T \mathbf{e}_1, \tag{4.2}$$

where $\mathbf{e}_1$ is the first column of E representing the direction of the largest variance and $\mathbf{p}$ is a $N \times 1$ projection vector. So each frame in such sequence is represented by an element in vector $\mathbf{p}$. The assumption underlying our approach is that respiratory motion is the major source of variation in these sequences where cardiac motion is eliminated. Therefore, we use $\mathbf{p}$ as our breathing surrogate.

4.3 Experiments

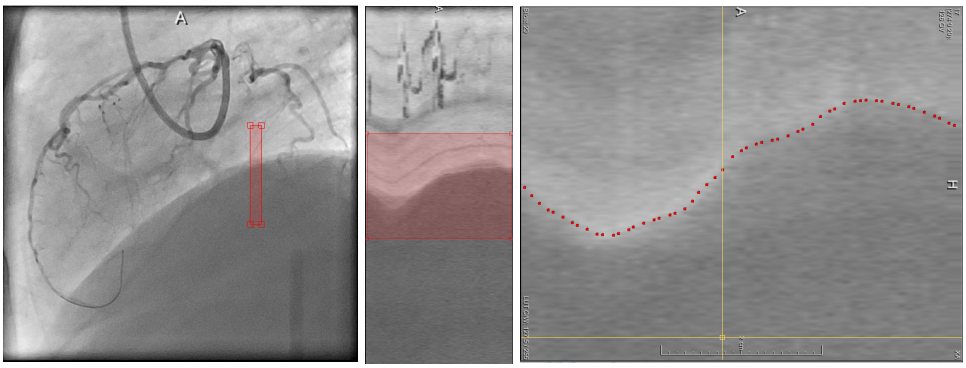
4.3.1 Image Data

For our experiments we used anonymized imaging data that were acquired from Department of Cardiology at Erasmus MC (University Medical Center Rotterdam) in Rotterdam, The Netherlands. XA images of 8 patients who underwent a PCI procedure that were acquired with Siemens AXIOM-Artis biplane system were analyzed. The frame rate of all sequences is 15 frames per second. The number of frames per series ranges from 55 to 244, corresponding to 3.7 to 16.3 seconds of imaging time. All 8 patients have in total 1898 frames. From our image data, five are 512×512 pixels, two are 600×600 and one is 1024×1024, with pixel size 0.216×0.216 or 0.279×0.279, 0.184×0.184 and 0.139×0.139 mm^2. The diaphragm can clearly be seen in 7 patients in both images of the biplane data, whereas the diaphragm border is not visible in the other one. In that case, some lung tissues motion can still be observed and served as the main indicator of respiratory motion. Contrast agent injection and fading during imaging can be seen in all sequences.

4.3.2 Ground Truth Data

Ideally, ground truth should be a direct indicator of respiratory motion. We first manually selected a rectangular ROI in original XA sequence on diaphragm border or where there is lung tissue motion, see Fig. 4.2a. Stacking all frames into a image volume and inspecting the "sagittal" view of the ROI, we observed a profile representing the change of diaphragm position, see Fig. 4.2b.

Manual labeling diaphragm or lung tissue's motion track was subsequently done on the sagittal view of ROI image (Fig. 4.2c). This labeling was performed such that there is only one marker in each frame. This labeling step resulted in a vector of real numbers representing the diaphragm position in the ROI over time and it served as the ground truth data in our study.



(a) (b) (c)

Figure 4.2: Ground truth data. (a) Drawing a rectangular ROI on diaphragm border. (b) Sagittal view of ROI. (c) Manual labeling of diaphragm border.

4.3.3 Retrospective Evaluation

For retrospective evaluation, we used all frames of an XA sequence to compute the first principal component. The resultant projection vector $\mathbf{p}$ represents respiratory motion of such sequence and was compared with the previously mentioned ground truth diaphragm labeling. This comparison was quantified by calculating the correlation coefficient between projection vector $\mathbf{p}$ and ground truth vector. In order to gain insight of the usability of our method on different system, the retrospective analysis was tested with both monoplane and biplane data. For biplane data, we combined information from both planes by simply using the concatenated matrix $X = \begin{pmatrix} X_A \\ X_B \end{pmatrix}$ in the same approach we have described in Section 4.2.2. By doing so, we could calculate one single projection vector $\mathbf{p}$ for both planes.

In addition, we compared the performance of our method, called *Vessel Removed* in later sections, with a recently published method that uses a masked-PCA approach [78]. Masked-PCA technique was designed for retrospective cardiac and respiratory motion gating on interventional cardiac x-ray images. In order to extract respiratory motion surrogate with this method, we slightly changed its implementation by directly using the projection vector on the 1st principal component without filtering it. We call it *With Mask* method in subsequent sections. We also investigated other possible variations of PCA-based methods, e.g. running PCA on the downsampled images without morphological closing (called *Downsampled Image*), and running PCA with an inverted mask of the mask created in [78] (called *Inverted Mask*). In all cases, the correlation coefficient of the resulting respiratory motion surrogate and the ground truth was calculated to quantify the performance on respiratory motion extraction.

4.3.4 Prospective Evaluation

We also evaluated whether the motion surrogate derived from our method could be used for prospective respiratory-induced coronary motion modeling in PCI. In this ex-

periment, we only utilized a part of the sequence to build the PCA-derived model and used it to make predictions on subsequent frames. Considering a scenario in coronary motion modeling, we would build a model based on frames with contrast agent and use it to improve alignment of preoperative data onto XA on frames where contrast agent has been flushed away. This makes it reasonable to choose frames before contrast agent starts fading for the PCA learning phase and use frames without contrast for the prediction phase. Similar to retrospective experiment, we used the correlation coefficient to quantitatively evaluate the results with monoplane and biplane data and also compared with performances of other previously mentioned methods.

All experiments were implemented in Matlab 2013b on an Intel Core2 2.66 GHz computer with 4 GB RAM running Windows. Computation time for each patient was recorded in all experiments.

4.4 Results

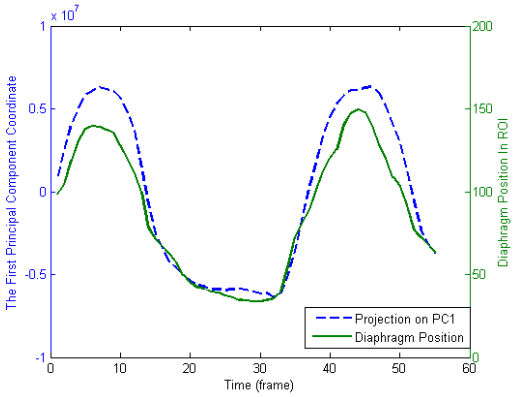
4.4.1 Retrospective Analysis

Example results of experiments on monoplane and biplane data from one patient are showed in Fig. 4.3. Fig. 4.3a, 4.3c and 4.3e present the comparison of projection vector $\mathbf{p}$ and ground truth diaphragm position. These figures show a high correlation between the two vectors. Linearly rescaling $\mathbf{p}$ to the range of ground truth data and overlaying it onto the ROI image provides another way to evaluate their correlation qualitatively. Fig. 4.3b, 4.3d and 4.3f reveal a good agreement between $\mathbf{p}$ and diaphragm border.

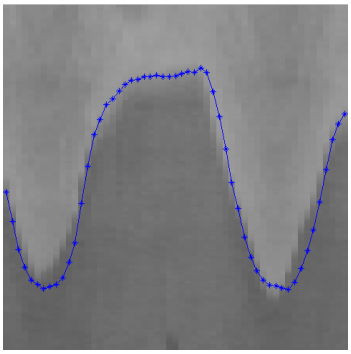
Table 4.1 provides quantitative measure results on correlation coefficients. The average correlation coefficient was calculated over all sequences for the various methods. From the table, it can be seen that all methods give high correlation coefficient (over 0.85, close to 1). The *Vessel Removed* method has slightly higher average correlation and lower standard deviation than other methods. For the patient whose XA sequences contain no diaphragm border, the correlation coefficients for *Vessel Removed* method are also high for the monoplane (0.89 and 0.88) and biplane data (0.88 and 0.95). Boxplots in Fig. 4.4 illustrate similar observations: The majority of correlation coefficients are over 0.8 for all methods; non-mask based methods (*Vessel Removed* and *Downsampled Image*) slightly outperform other methods.

To investigate whether there is a statistically significant difference between the performance of these methods on retrospective respiratory motion surrogates extraction, we used a one-tailed Wilcoxon rank sum test to check the correlation coefficients, as their distribution is not necessarily a normal distribution and most of values are close to 1. Result (see Table 4.2) shows that the p values for monoplane experiments range from 0.09 to 0.55 (in upper triangle); p values for biplane experiments range from 0.11 to 0.42 (in lower triangle). The result means these methods have similar performance on retrospective task in terms of statistical significance.

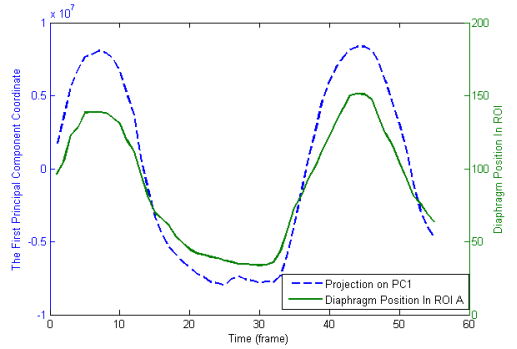
Fig. 4.5 illustrates the frequency distribution of correlation coefficients for various methods in our retrospective analysis. In both the monoplane and biplane ex-



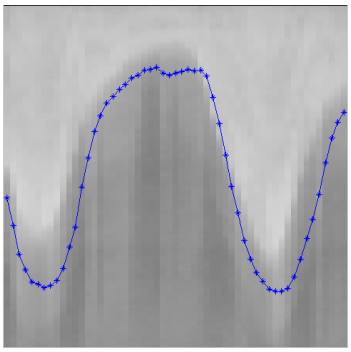
(a) Monoplane



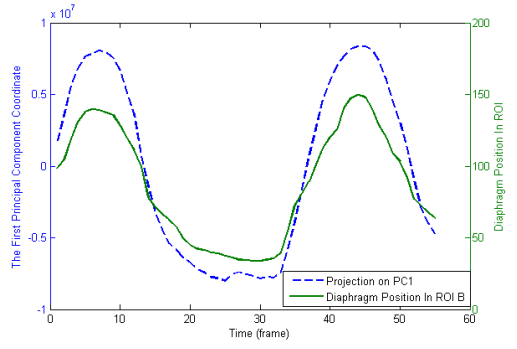
(b) Monoplane



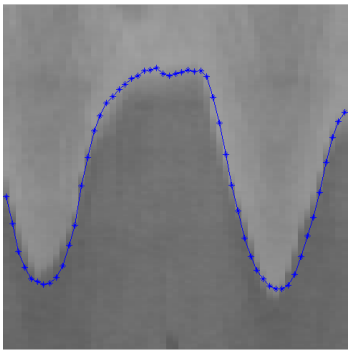
(c) Biplane A



(d) Biplane A



(e) Biplane B

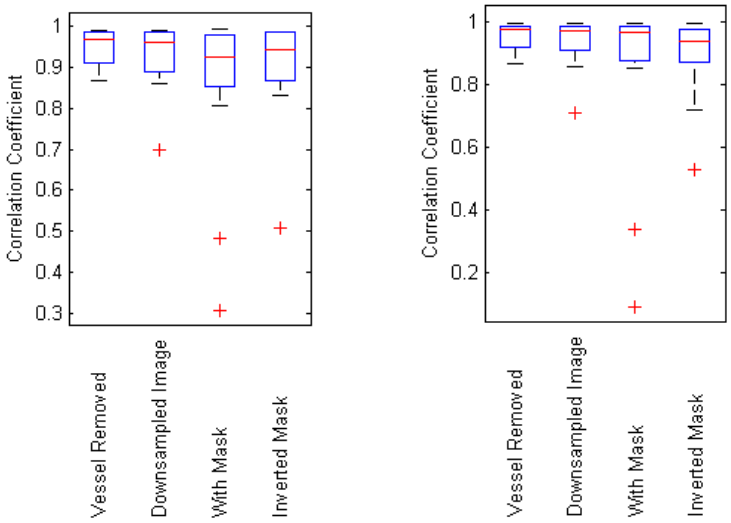


(f) Biplane B

Figure 4.3: Retrospective projection on the first principal component for one patient, in comparison with diaphragm position in ROI image. (a) and (b) PCA was done on one sequence of the biplane data, (c) -(f) principal components were derived from the concatenated sequence of both planes. (c) and (d) show the projection in comparison with plane A, (e) and (f) illustrate the comparison with plane B.

Table 4.1: Average correlation coefficient of projection vector **p** and diaphragm positions for various methods for retrospective evaluation.

Methods	Monoplane (mean ± std)	Biplane (mean ± std)
Vessel Removed	0.9490 ± 0.0446	0.9529 ± 0.0424
Downsampled Image	0.9330 ± 0.0754	0.9357 ± 0.0742
With Mask	0.8637 ± 0.1883	0.8552 ± 0.2503
Inverted Mask	0.9032 ± 0.1156	0.9007 ± 0.1197



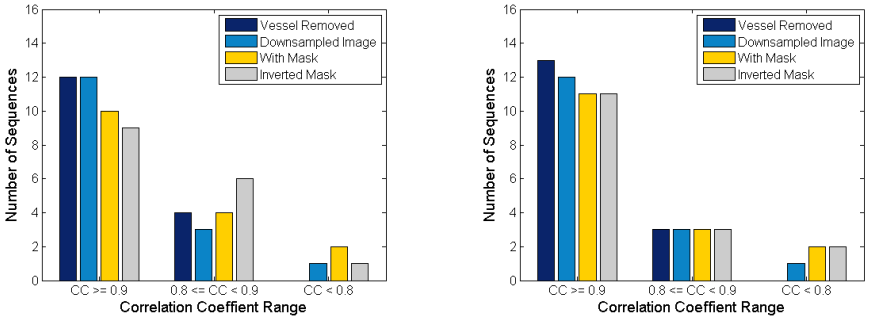
(a) Monoplane **(b)** Biplane

Figure 4.4: Boxplot of correlation coefficients calculated using various methods on monoplane and biplane data for retrospective evaluation.

Table 4.2: Statistical significance between various methods for retrospective evaluation (p values). The numbers in the upper and lower triangle in the table show the results of monoplane and biplane cases respectively.

Biplane \ Monoplane	Vessel Removed	Downsampled Image	With Mask	Inverted Mask
Vessel Removed	×	0.28	0.15	0.09
Downsampled Image	0.31	×	0.24	0.22
With Mask	0.23	0.42	×	0.55
Inverted Mask	0.11	0.24	0.36	×

periments, it can be observed that *Vessel Removed* method has the most number of correlation coefficients over 0.9 and no correlation coefficients lower than 0.8, which outperforms all other methods. This observation suggests that the *Vessel Removed* method is more robust than the other approaches.



(a) Monoplane **(b)** Biplane

Figure 4.5: Frequency distribution of correlation coefficients for various methods in retrospective analysis.

Table 4.3 compares the average per-frame computation time that is needed to compute the projection vector **p**. This includes the time for image preprocessing, building statistical model through PCA and making projection on the first principal component. The comparison reveals the advantage of non-mask based methods to mask-based method that the computation time they need is much shorter, which is favored for real clinical workflow.

4.4.2 Prospective Analysis

Example results of the prospective analysis for the same patient as in Fig. 4.3 are shown in Fig 4.6. Fig. 4.6a, 4.6c and 4.6e present retrospective projection for frames

Table 4.3: Average per frame computation time for various methods for retrospective evaluation (milliseconds).

Methods	Monoplane (ms)	Biplane (ms)
Vessel Removed	17.2	34.4
Downsampled Image	8.0	15.5
With Mask	821.6	2227.8
Inverted Mask	854.7	2920.0

used for learning the statistical model and prospective projection for frames excluded from the learning phase. It can be seen that the prospective projection vector **p** still maintains good correlations with ground truth diaphragm position.

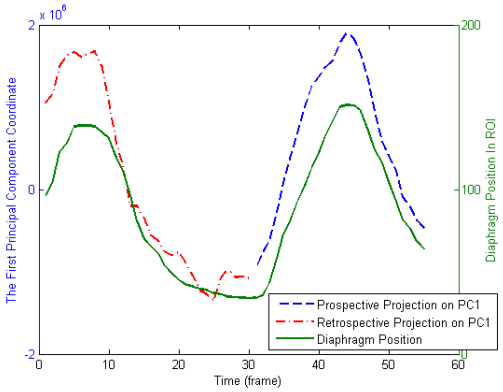
Correlation coefficients are shown in Table 4.4. In general, these numbers are lower than those in the retrospective experiments, while *Vessel Removed* method still maintains a high average correlation coefficient over 0.9. Its standard deviation is also lower than other methods. For the patient whose diaphragm cannot be seen in the XA sequences, the correlation coefficients for *Vessel Removed* method remain good for one of sequences in the monoplane data (0.96 and 0.70) and both sequences in the biplane data (0.91 and 0.87). Boxplots in Fig. 4.7 show that the medians of all methods are quite close to each other, but *Vessel Removed* method has fewer correlation coefficients lower than the median value compared to other methods. We also used Wilcoxon rank sum test to check the statistical significance, results are shown in Table 4.5. No significant difference is found among these methods ($p>0.14$), which means their performance are similar to each other in terms of statistical significance.

Table 4.4: Average correlation coefficients of the prospective projection vector and diaphragm positions for various methods.

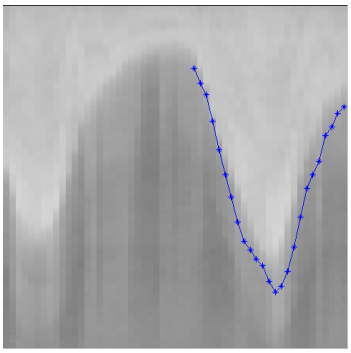
Methods	Monoplane (mean ± std)	Biplane (mean ± std)
Vessel Removed	0.9197 ± 0.0733	0.9128 ± 0.1239
Downsampled Image	0.8823 ± 0.1418	0.8815 ± 0.1413
With Mask	0.7548 ± 0.3291	0.8282 ± 0.2540
Inverted Mask	0.7772 ± 0.2649	0.8201 ± 0.2117

Frequency distribution of correlation coefficients in Fig. 4.8 reveals that for both monoplane and biplane experiments, more lower value correlation coefficients appear for all methods compared to retrospective analysis. It is also clear that *Vessel Removed* method has the most high correlation coefficients ($CC \geq 0.9$) and the least lower correlation coefficients ($CC < 0.8$) among all methods.

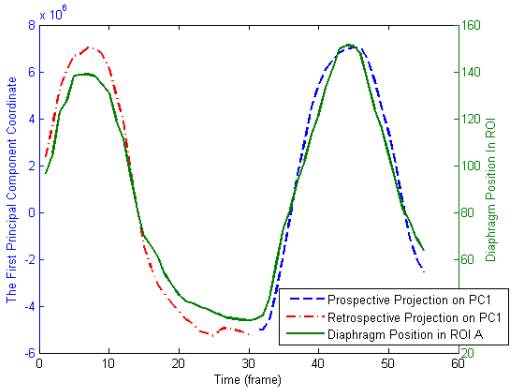
Average per frame prediction time for various methods is listed in Table 4.6. The prediction time includes the time for image preprocessing and computing projection on the first principal component. It is obvious that *Downsampled Image* method has the shortest prediction time, while *Vessel Removed* method is also quite fast. Mask



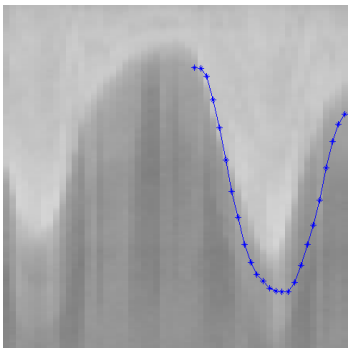
(a) Monoplane



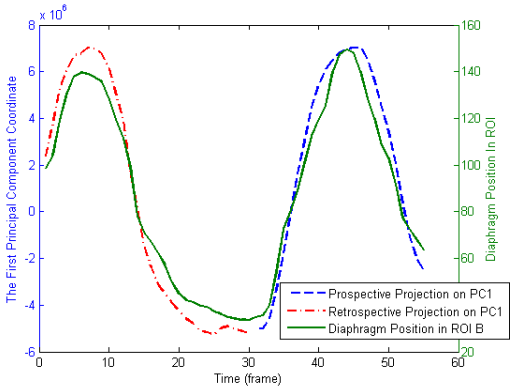
(b) Monoplane



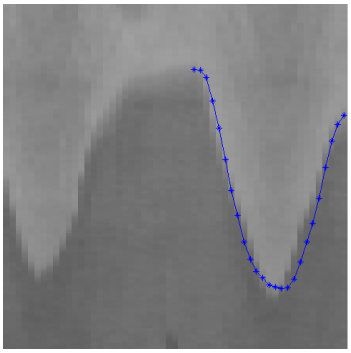
(c) Biplane A



(d) Biplane A

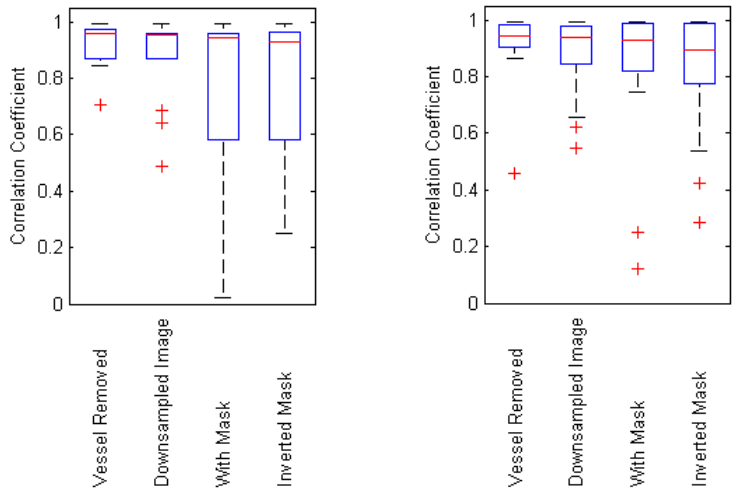


(e) Biplane B



(f) Biplane B

Figure 4.6: Prospective projection on the first principal component for one patient, in comparison with diaphragm position in ROI image. (a) and (b) PCA was implemented on monoplane data, (c) -(f) principal components were derived from the concatenated sequence of biplane. (c) and (d) show the projection in comparison with plane A, (e) and (f) illustrate the comparison with plane B.

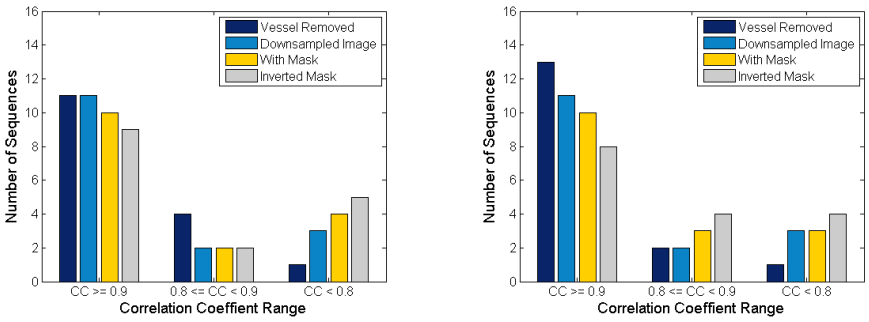


(a) Monoplane **(b)** Biplane

Figure 4.7: Boxplot of correlation coefficients calculated using various methods on monoplane and biplane data for prospective evaluation.

Table 4.5: Statistical significance between various methods in prospective evaluation (p values). The numbers in the upper and lower triangle in the table show the results of monoplane and biplane cases respectively.

Biplane \ Monoplane	Vessel Removed	Downsampled Image	With Mask	Inverted Mask
Vessel Removed	×	0.31	0.15	0.17
Downsampled Image	0.52	×	0.18	0.17
With Mask	0.31	0.36	×	0.57
Inverted Mask	0.18	0.19	0.39	×



(a) Monoplane **(b)** Biplane

Figure 4.8: Frequency distribution of correlation coefficients for various methods in prospective analysis.

based methods are slower than non-mask based methods, especially when they were run on biplane data.

Table 4.6: Average per frame prediction time for various methods (milliseconds).

Methods	Monoplane (ms)	Biplane (ms)
Vessel Removed	9.6	20.0
Downsampled Image	1.6	3.8
With Mask	16.3	282.6
Inverted Mask	31.0	792.4

4.5 Discussion

We developed an automatic method to extract patient-specific respiratory motion surrogate from cardiac interventional X-ray angiograms using principal component analysis. The method was evaluated on monoplane and biplane data in both retrospective and prospective manner.

Our experiments demonstrated that *Vessel Removed* method is able to extract breathing information having high correlation with the ground truth diaphragm or lung tissue motion. The average correlation coefficient is higher than those for other related methods in our experiments. *Vessel Removed* method is also more robust than the other three methods giving that more correlation coefficients for the method are over 0.9 and less are below 0.8.

It is also observed that the difference between the mentioned methods in this paper is yet not so profound that no statistically significant difference on the correlation coefficients was found. This might be due to the choice of similarity metric. Correlation coefficient is although one of the common ways to measure the similarity

of two time-series signals, there are other measures which could potentially improve the difference between these algorithms, such as Distance correlation [101].

The limited number of patients (only 8) may also be a reason for the lack of statistical significance. In the future, we will evaluate the method on a much larger set of patients. Despite the lack of statistical insignificance, the *Vessel Removed* method performs at least as good as the other three approaches.

From the aspect of computation efficiency, the time that *Vessel Removed* method needs for building the statistical model and making prediction on our computer are less than 67 ms, corresponding to the 15 Hz imaging rate of our data, whereas mask-based methods need longer time to accomplish the same task. This means that *Vessel Removed* method could run in real-time.

Image-based respiratory motion surrogates in interventional X-ray angiograms have been studied previously [56, 69, 96]. These works either need manually putting a rectangular ROI or require specific object being present in images. The method we have presented is fully automatic and more robust to various image content. The diaphragm is not necessarily required to be present as long as there is sufficient breathing motion observed, which is true in most of the cases in PCI procedures since the lung tissue is usually seen in the background.

The application of dimensional reduction techniques in extraction of respiratory motion information were seen in [78] and [18]. [18] presented one example case and the method in [78] was originally designed for retrospective gating. In comparison to these works, the method we have developed is simpler and needs no vessel extraction from images thus is also faster. In addition, we have evaluated the usability of our method on monoplane and biplane data in retrospective and prospective manner and achieved good correlation in both tasks.

Observations on principal components of XA images would help understanding the mechanism of our proposed method. The first four principal component images, called "eigenimages", of two example sequences with and without diaphragm are shown in Fig. 4.9. It is obvious that in both cases, the cardiac motion pattern is still present in eigenimages of the original images, but significantly suppressed in those of the morphological-closed images. Also in vessel-removed images, for the case with a diaphragm, the diaphragm border is enhanced in the eigenimages showing a white or dark stripe; while in the sequence where diaphragm is not present, white and dark pixels represent background lung tissues. It can also be observed that the first eigenimage contains strong respiratory motion signals which makes it reasonable of projecting XA sequence on the first principal component to obtain breathing motion feature.

In prospective surrogate extraction, the frames that are needed for building the statistical model are required to cover the maximal range of respiratory motion, therefore our proposed method might be limited for different breathing patterns. The method we developed is also only applied to one fixed view angle. Detector position changes during interventions require model rebuilding, which having been seen to be fast in our experiments.

The method we developed could directly be used for patient-specific coronary motion modeling. Its short modeling and prospective extraction time enables the possibility of running in real time and being used during interventions. Due to the

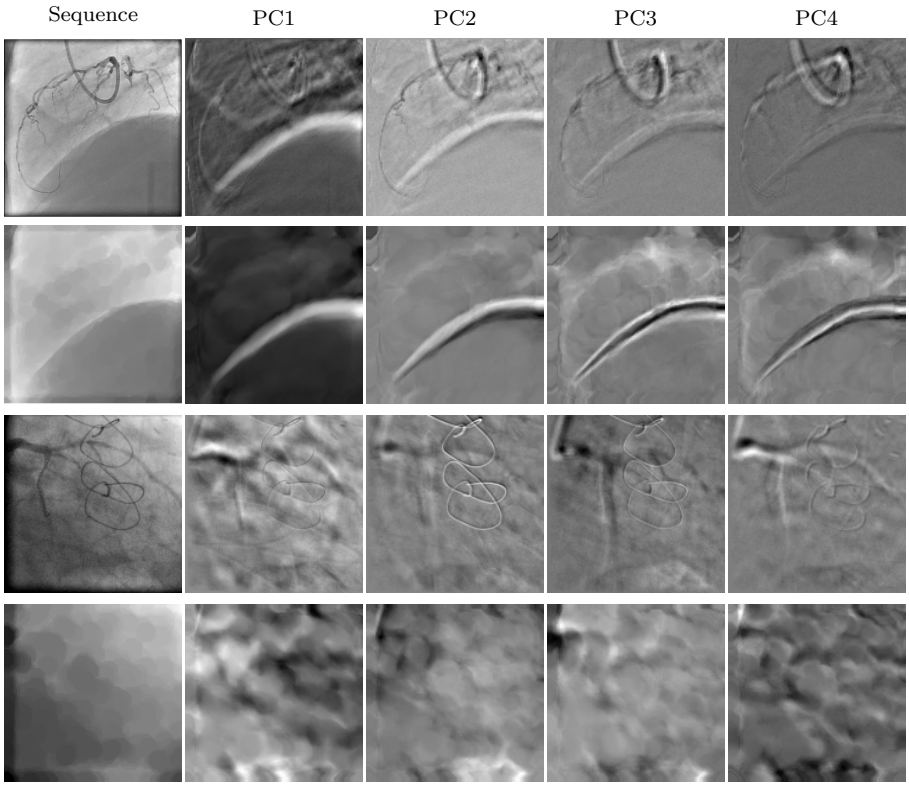


Figure 4.9: The first 4 principal components (eigenimages) of images with and without diaphragm being present. The first two rows of images have diaphragm and the last two rows do not. The 1st and 3rd row show the original images and their eigenimages, the 2nd and 4th row are the same images in the 1st and 3rd row after morphological closing operation.

robustness of the method to different image contents, it could also be potentially used for extraction of respiratory motion surrogate for other types of interventions using different imaging modalities.

In the future, we will extend the study with more patients' data. We will also investigate on cardiac motion surrogates extraction from XA sequences with similar framework and adapting the current approach to varying view angles.

4.6 Conclusion

We have presented a fast automatic method that can be used to retrospectively and prospectively extract patient-specific respiratory motion surrogate from cardiac XA sequences. Our experiments demonstrate a high correlation coefficient with manual ground truth: average correlation coefficients are over 0.9 in the retrospective and

prospective evaluations. The method is easy to implement and runs in real-time, thus allows to extract respiratory motion surrogates during interventions.

Acknowledgement The authors wish to acknowledge financial support from Technology Foundation STW IMAGIC project under iMIT program Grant No. 12703.

Chapter Five

Fast Prospective Detection of Contrast Inflow in X-ray Angiograms with Convolutional Neural Network and Recurrent Neural Network

Abstract — Automatic detection of contrast inflow in X-ray angiographic sequences can facilitate image guidance in computer-assisted cardiac interventions. In this paper, we propose two different approaches for prospective contrast inflow detection. The methods were developed and evaluated to detect contrast frames from X-ray sequences. The first approach trains a convolutional neural network (CNN) to distinguish whether a frame has contrast agent or not. The second method extracts contrast features from images with enhanced vessel structures; the contrast frames are then detected based on changes in the feature curve using long short-term memory (LSTM), a recurrent neural network architecture. Our experiments show that both approaches achieve good performance on detection of the beginning contrast frame from X-ray sequences and are more robust than a state-of-the-art method. As the proposed methods work in prospective settings and run fast, they have the potential of being used in clinical practice.

Based upon: **H. Ma**, P. Ambrosini and T. van Walsum: Fast Prospective Detection of Contrast Inflow in X-ray Angiograms with Convolutional Neural Network and Recurrent Neural Network. The 20th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI), Lecture Notes in Computer Science, vol. 10434, pp. 453-461, 2017.

5.1 Introduction

During percutaneous coronary interventions (PCI), X-ray angiography (XA) is commonly used by clinicians to identify the sites of plaque and navigate devices through the arteries of patients with advanced coronary artery disease. As X-ray imaging has poor soft tissue contrast, coronary arteries are normally visualized by injecting radio-opaque contrast agent in the vessels.

Approaches for improving image guidance in such procedures have been reported, for example fusion of coronary models from CTA [92]. Such methods can only be applied if vessels are visible in the XA, thus automated application of such methods requires detection of presence of contrast agent. Similarly, automated detection of catheter and guidewires, which can also be used for virtual roadmapping [13], is generally only possible in non-contrast enhanced frames. Therefore, an automatic way to detect contrast inflow online is relevant for further automating advanced image guidance methods for coronary interventions, reducing interactions of clinicians with computers during procedures.

Existing works for detection of contrast inflow in X-ray images fall into two categories: enhancement-based and learning-based. Enhancement-based methods [27, 61, 62, 115, 117] enhance contrasted structures, followed by a step to extract features that indicate the change of contrast throughout the sequence. The contrast-enhanced frames are then detected via analysis of the feature. Learning-based approaches [25, 50] train a classifier to detect contrast or non-contrast frames based on handcrafted image features. Among these works, [25, 62, 115] need an entire sequence to detect contrast inflow, and thus only work retrospectively. [61] does not rely on a complete sequence, but retrospectively runs on a sliding segment of a few new X-ray frames, thus there is a trade-off between the possible delay of the contrast inflow detection and the overall processing efficiency. In addition, this method was designed specifically for TAVI procedures on aorta: their contrast detection method involves aligning a predefined aorta shape model to X-ray images and a step of TEE probe detection, which is not relevant for coronary interventions. [117] uses a heuristic approach to detect the first contrast-enhanced frame from X-ray sequences of left atrium (LA) used for electrophysiology (EP) ablation procedures. [50] developed a learning-based framework on X-ray images of LA for EP procedures. The method used a SVM classifer with the heuristic features introduced in [27] and [117]. Out of these methods, [27] is the only one that may be directly used for coronary interventions and work in prospective settings.

The purpose of our work is to develop and evaluate solutions for prospective detection of contrast inflow in XA images that can fit into the clinical work-flow of coronary interventions. Specifically, we aim at prospectively detecting if a frame has contrast agent. To this end, two different approaches were developed. Due to the exceptional performances that convolutional neural networks (CNN) have in image classifications [60], and medical applications, such as tissue segmentation and surgical tools detection [41], we propose a learning-based method using CNN to classify each frame of an XA sequence into two classes: with or without contrast. Additionally, we propose a hybrid of enhancement- and learning-based. It computes a temporal

contrast feature from vessel-enhanced sequences based on which contrasted frames are detected with long short-term memory (LSTM) [49], a recurrent neural network (RNN) architecture. To the best of our knowledge, this is the first work that applies deep learning for contrast inflow detection in X-ray images. To validate the detection, the position of the beginning contrast frame (BCF) [25, 50] in a sequence (where contrast starts being visible) was used in the experiment.

5.2 Methods

5.2.1 The CNN-based Method

Let $S = \{I_1, I_2, \ldots, I_n\}$ denote a sequence of n frames in which I_c is the beginning contrast frame. All frames $I_1, \ldots, I_{c-1}$ are associated with the label "without contrast". The other frames $I_c, \ldots, I_n$ have the label "with contrast".

In order to classify the fluoroscopic frames, we used a CNN to learn the difference between the contrast frame and non-contrast frame (Fig. 5.1, top). The input of the CNN has 5 images: the current frame I_i to be classified, its 3 previous frames I_{i-1}, I_{i-2}, I_{i-3}, and the first frame I_1 (normally non-contrasted). There are 7 intermediate layers directly after the input layer, each of which has a *n-conv* block with n consecutive convolutions (Fig. 5.1, bottom). The last *n-conv* block is connected with two fully-connected layers. The final output is a softmax layer with two nodes: "with contrast" and "without contrast". The model was trained with binary cross-entropy as the loss function. In order for a faster convergence, batch normalization was used after every convolution, residual connection at every layer and the strided convolutions instead of pooling layers.

To detect the BCF of an XA sequence online using the trained model, frames of the sequence were classified one by one in a chronological order. The first frame labeled as "with contrast" in the sequence is considered as BCF.

5.2.2 The RNN-based Method

The RNN-based method consists of two major steps: vessel enhancement and contrast frame detection. An overview of this method is illustrated in Fig. 5.2.

Vessel enhancement The vessel enhancement step is crucial for accurate approximation of contrast changes in XA sequences. This step removes most non-vessel background structures using a previously developed online layer separation technique [67] (Chapter 3 of this thesis) followed by multi-scale Frangi-vesselness filtering [38].

The online layer separation method prospectively separated an XA sequence into three layers: a breathing layer, a quasi-static background layer, and a vessel layer in which vessels have better visibility. First, the breathing layer was separated via morphological closing. After this layer was removed from the original image, online robust PCA (OR-PCA) [35] was applied to separate the low-rank quasi-static layer and sparse vessel layer through alternatively projecting the new data sample (frame) to the underlying low-rank subspace basis and updating the basis using the new esti-

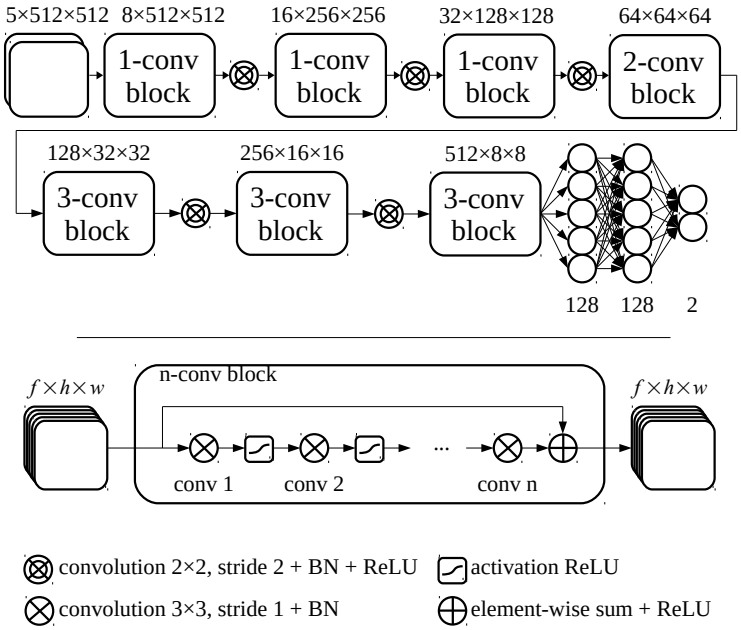


Figure 5.1: The neural networks (top) connects the 5 input images (the first, the current and its 3 previous X-ray frames) to the 2 output nodes ("with contrast" and "without contrast"). The model consists of several *n-conv* blocks. They are a succession of CNNs with a skip connection between the input and the output of the block (bottom). $f \times h \times w$ is the dimension of the data (feature number times image height times image width) .

mation of the layers. After layer separation, the structures that may cause artefacts in the next step, such as diaphragm, spine, were removed from the vessel layer.

Following the layer separation, a multi-scale vesselness filter [38] was applied on the separated vessel layer to further enhance the tubular structures. In the end, after the vessel enhancement step, for each incoming frame, a new image was created where vessel structures are enhanced.

Contrast frame detection Once the image with enhanced vessel structures is obtained, the feature that indicates the level of contrast agent was extracted from the image. In this work, we used the average pixel intensity of the complete vessel-enhanced image as the contrast feature. This results in a 1D signal for a complete sequence.

The last step is to detect contrast frames from the previously obtained 1D contrast signal. In order to fully use the temporal relation between frames, each signal point is classified as "contrast" or "non-contrast" with a recurrent neural network. The long short-term memory (LSTM) network [49] was used due to its good performance on modeling long-term temporal relations in time-series data.

Let x_k denote the feature for the kth frame I_k. The single-direction LSTM takes

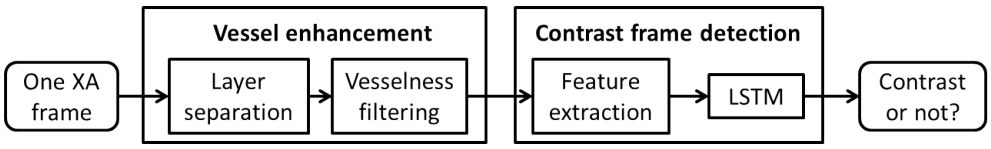


Figure 5.2: The overview of the RNN-based method.

x_k as the input. A hidden state h_k in the LSTM network is recurrently updated through nonlinear interactions between the input signal x_k, the LSTM units and its state of the last time point h_{k-1}. The output label y_k of x_k is the outcome of a nonlinear function of h_k. This process is illustrated in Fig. 5.3.

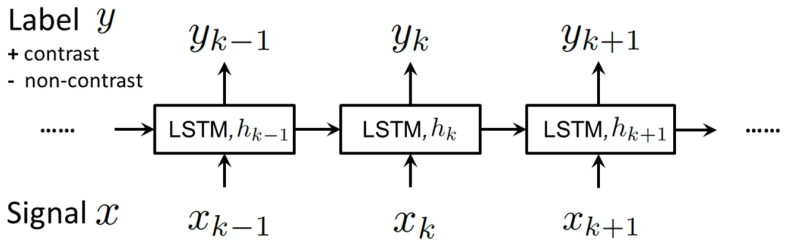


Figure 5.3: Each signal point is classified from a contrast frame or a non-contrast frame with a LSTM network.

5.3 Experiments

We retrospectively obtained anonymized data that was acquired during clinical routine with a Siemens AXIOM-Artis biplane system. The data were 120 XA sequences from 26 patients who underwent a PCI procedure. The frame rate of all sequences is 15 frames per second. The length of sequence varies from 24 to 244 frames. The size of images in our dataset are 512×512, 600×600, 776×776 and 1024×1024. In all sequences, contrast inflow can be observed. In our experiments, 40 sequences from 20 patients were used as training data, the 80 sequences from the other six patients were used for validation.

For the CNN-based method, all images were resized to 512×512 before training. The parameters of the CNN model were optimized using stochastic gradient descent with a learning rate 0.0001, a decay of 0.0005 and a momentum of 0.99. The model was trained with a batch size of 15 during 33,000 iterations. For each sequence, the six frames before and after the BCF were chosen to ensure an even number of contrast and non-contrast training images. The BCF was discarded to assist the CNN to learn more differences between contrast and non-contrast frames. As the dataset used to train the model is small, data augmentation was applied during the training to virtually create more data: translation (+/- 100 pixels), rotation (+/- 5 degrees), scaling (+/- factor 0.1), intensity shift (+/- 0.2), Gaussian noise ($\sigma_g = 0.01$) on the

normalized image between 0 and 1, and vertical flip were used to transform images.

For the RNN-based method, we manually tuned the parameters based on visual check and quantitative evaluation on the training data; the same parameters were used for testing. The images were first down-scaled 2 or 4 times to 256×256 or 300×300 or 388×388 depending on the original image size for speeding up the image processing. The parameters for layer separation were set following the approach with the sliding window option in [67] (Chapter 3 of this thesis) using the closed-form solution of OR-PCA. To improve the convergence of OR-PCA, we used a mini-batch of 5 frames (before contrast agent was injected) to get an initial estimate of the low-rank subspace basis. This was done using the layer separation method in [65] (Chapter 2 of this thesis) with fast principal component pursuit [90]. The scale of Frangi vesselness filter was set ranging from 0.6 mm to 2.8 mm according to the size of coronary arteries. The β and c parameter of the vesselness filter were 0.5 and 15. The dimension of LSTM units was set to 7 with a dropout probability being 0.2. The nonlinear activation function of the hidden layer is sigmoid function. The LSTM network was trained using RMSprop optimizer with a learning rate being 0.005 during 100 epochs. At last, the BCF was detected as the first frame in a sequence being classified as contrasted by LSTM.

In the experiments, we also compared our methods with the state-of-the-art approach of Condurache et al. [27]. For setting the parameters of the method, the first 3 feature values from non-contrast frames were modeled as a Gaussian $N_0(\mu_0, \sigma_0^2)$. The threshold T for choosing contrast frames was set to $\mu_0 + 3\sigma_0$.

The evaluation metric we used is the absolute difference between the frame index of the ground truth BCF and the frame predicted by different methods.

The image processing steps in the RNN-based method and the method of Condurache et al. were implemented in MATLAB with a single CPU core (Intel Core i7-4800MQ 2.70 GHz). LSTM and CNN were implemented in Keras with TensorFlow as backend. LSTM was running on the CPU due to its small dimension. CNN was trained and tested on an Nvidia GeForce GTX 1080 GPU.

5.4 Results and Discussion

Fig. 5.4 shows an example to illustrate steps in the RNN-based method. The statistics of the absolute errors made by the three methods are shown in Table 5.1. The results of the mean and median errors show that the two proposed approaches have smaller errors than the state-of-the-art method, especially, the RNN-based method is able to achieve a median absolute error of 2 frames. The median of non-absolute errors (prediction minus ground truth) indicates the prediction bias of each method. The method of Condurache et al. makes late predictions, while the others have a minor bias. The table also lists the number of sequences with a small prediction error (3 frames, being about 0.2 seconds) and a large error (>10 frames). The method of Condurache et al. has mis-detection on 7 sequences (the first entry in the last two columns in Table 5.1), which was also reported in [50]. While the two proposed methods both have 55 sequences with a small error ($\leqslant$ 3 frames) out of 80, the CNN-based approach has the smallest numbers of sequences with a large error ($>$ 10 frames)

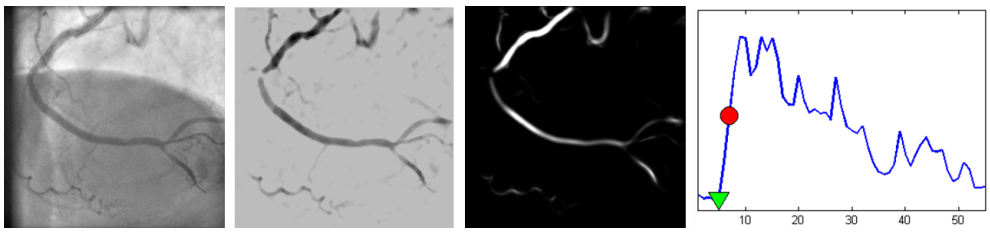


Figure 5.4: An example to illustrate the RNN-based method. From left to right are the original XA frame (left), the vessel layer after layer separation (middle left), the vesselness image (middle right), the contrast signal for the whole sequence (right). The color markers in the signal show the prediction of BCF with LSTM (red) and the ground truth (green). Note that the artefact of diaphragm does not appear in the vesselness image thanks to layer separation.

Methods	mean (std)	median (*)	#(error $\leqslant 3$)	#(error > 10)
Condurache et al. [27]	6.2 (7.1)	5 (4)	29 / 73	10 / 73
CNN-based	3.9 (4.9)	2.5 (1)	**55** / 80	**5** / 80
RNN-based	**3.6** (4.6)	**2** (-0.5)	**55** / 80	7 / 80

Table 5.1: The statistics of the absolute error for the 3 methods. The two columns in the middle show the mean, standard deviation, median of the absolute errors and the median of non-absolute errors (*) in frames. The last two columns show the number of sequences on which the method made an absolute error no larger than 3 frames or larger than 10 frames.

among the three methods.

The median error of the RNN-based method is similar to the results reported in [25]. While they achieved a mean error of less than one frame, their detection step requires the knowledge of complete sequences, hence it will not work in a prospective scenario. The learning-based method in [50] can be used for prospective detection, but some of the proposed features were heuristically designed for X-ray images of LA for EP procedure, which have different image features from the XA of coronary interventions. Compared to these methods, our approaches were designed for prospective settings and the CNN-based method is a general framework that could potentially be applied in different clinical procedures.

The RNN-based learning with a hand-crafted feature has slightly lower mean and median error than the CNN-based method, although the latter has a more complex and deeper architecture. This might contradict to what is commonly known about the performance of deep learning. The possible reasons may be two-fold. First, the size of training data was small, even with data augmentation and a reduced CNN model, some over-fitting was observed. Second, the CNN treats frames independently rather than modeling their temporal relations. Although CNNs perform excellent in many classification tasks, detecting BCF requires a classifier that has good accuracy for data on the border between two classes.

In terms of computation efficiency (test time), the method of Condurache et al. needed 111 ms to 443 ms to process a frame. While the CNN-based method ran very fast and used on average only 14 ms to process one frame. The RNN-based method ran on average 64 ms/frame on images of the original size 512×512 or 1024×1024, and 140 ms/frame on images of the original size 776×776. As the test time of the RNN-based method was based on a MATLAB implementation with a single CPU core, it has large potential to run in real-time (<66 ms) with an optimized implementation running on a modern GPU.

5.5 Conclusion

We have developed two novel approaches for prospective detection of contrast inflow in XA sequences, a CNN-based and a RNN-based approach. The proposed methods perform well in BCF detection tasks in XA sequences, and outperform a previous state-of-the-art method. Both methods work in prospective settings and run fast, therefore have the potential to be integrated in advanced image guidance systems for PCI.

Acknowledgement This work was supported by Technology Foundation STW, IMAGIC project under the iMIT program (grant number 12703).

Chapter Six

Dynamic Coronary Roadmapping via Catheter Tip Tracking in X-ray Fluoroscopy with Deep Learning Based Bayesian Filtering

Abstract — Percutaneous coronary intervention (PCI) is typically performed with image guidance using X-ray angiograms in which coronary arteries are opacified with X-ray opaque contrast agents. Interventional cardiologists typically navigate instruments using non-contrast-enhanced fluoroscopic images, since higher use of contrast agents increases the risk of kidney failure. When using fluoroscopic images, the interventional cardiologist needs to rely on a mental anatomical reconstruction. This paper reports on the development of a novel dynamic coronary roadmapping approach for improving visual feedback and reducing contrast use during PCI. The approach compensates cardiac and respiratory induced vessel motion by ECG alignment and catheter tip tracking in X-ray fluoroscopy, respectively. In particular, for accurate and robust tracking of the catheter tip, we proposed a new deep learning based Bayesian filtering method that integrates the detection outcome of a convolutional neural network and the motion estimation between frames using a particle filtering framework. The proposed roadmapping and tracking approaches were validated on clinical X-ray images, achieving accurate performance on both catheter tip tracking and dynamic coronary roadmapping experiments. In addition, our approach runs in real-time on a computer with a single GPU and can be easily integrated into the clinical workflow of PCI procedures, providing cardiologists with visual guidance during interventions without the need of extra use of contrast agent.

Based upon: **H. Ma**, I. Smal, J. Daemen and T. van Walsum: Dynamic Coronary Roadmapping via Catheter Tip Tracking in X-ray Fluoroscopy with Deep Learning Based Bayesian Filtering. *Medical Image Analysis*, in press.

6.1 Introduction

6.1.1 Clinical Background

Percutaneous coronary intervention (PCI) is a minimally invasive procedure for treating patients with coronary artery disease. During these procedures, medical instruments inserted through a guiding catheter are advanced to treat coronary stenoses. A guiding catheter is firstly positioned into the ostium of the coronary artery. Through the guiding catheter, a balloon catheter carrying a stent is introduced over a guidewire to the stenosed location. The balloon is then inflated and the stent is deployed to prevent the vessel from collapsing and restenosing.

PCI is typically performed with image-guidance using X-ray angiography (XA). Coronary arteries are visualized with X-ray opaque contrast agent. During the procedure, interventional cardiologists may repeatedly inject contrast agent to visualize the vessels, as the opacification of coronary arteries only lasts for a short period. The amount of periprocedural contrast use has been correlated to operator experience, procedural complexity, renal function and imaging setup [84]. Furthermore, the risk for contrast induced nephropathy has been associated to contrast volume [104]. Manoeuvring guidewires and material, however, typically occurs without continuous contrast injections. In these situations, the navigation of devices is guided with "vessel-free" fluoroscopic images. Cardiologists have to mentally reconstruct the position of vessels and stenosis based on previous angiograms.

6.1.2 Dynamic Coronary Roadmapping

Dynamic coronary roadmapping (DCR) is a promising solution towards improving visual feedback and reducing usage of contrast medium during PCI [32, 55, 71, 119]. This approach dynamically superimposes images or models of coronary arteries onto live X-ray fluoroscopic sequences. The dynamic overlay serves as a roadmap that provides immediate feedback to cardiologists during the intervention, so as to assist in navigating a guidewire into the appropriate coronary branch and proper placement of a stent at the stenotic site with reduced application of contrast agent. Studies with a phantom setup using research software [55] or on first cases of clinical interventions using commercially available systems [29, 102, 113] have investigated the usefulness of DCR in assisting PCI, reporting that DCR helps to reduce procedure time, radiation dose and contrast volume.

To develop a DCR system, it is important but yet a challenge to accurately overlay a roadmap of coronary arteries onto an X-ray fluoroscopic image, as limited information of vessels is present in the target fluoroscopic image for inferring the compensation of the vessel motion resulting from patient respiration and heartbeat. The methods that have been proposed on motion compensation for DCR can be generally grouped into two categories: direct roadmapping and model-based approaches.

Direct roadmapping methods use information from X-ray images and ECG signals to directly correct the motion caused by respiration and heartbeat. The first DCR system [32] used digital subtraction of a contrast sequence and a mask sequence to create a full cardiac cycle of coronary roadmaps. The roadmaps were stored and

later synchronized with the live fluoroscopic sequence by aligning the R waves of their corresponding ECG signals. This system compensates the cardiac motion of vessels, yet does not correct the respiratory motion during interventions. Two later studies [71, 119] introduced image-based respiratory motion compensation methods for DCR. Their methods assumed an affine respiratory motion model in ECG-gated fluoroscopic frames and recovered the respiratory motion from soft tissues with special handling of static structures. The limitation of these approaches is that they require relevant tissue to be sufficiently visible in the field of view for reliable motion compensation which is not always the case. In addition, they require to be run on cardiac-gated frames. In a more recent work [55], binary vessel masks were created as the roadmaps from at least one cardiac cycle of angiographic images. Temporal alignment of the roadmaps and the fluoroscopic sequence, which compensated the cardiac motion of vessels, was performed by registering ECG signals using cross-correlation. Additionally, the respiratory motion was corrected by aligning the guidewire centerline in the fluoroscopy to the contour of vessels in the angiogram where the roadmaps were created. The system has been shown useful in a phantom-based study, nevertheless no accuracy evaluation of the spatiotemporal alignment was presented. Furthermore, the spatial registration relies on robust extraction of vessels and guidewires which is often challenging for X-ray images.

Unlike direct roadmapping, the model-based approaches build a model to predict motion in fluoroscopic frames. The motion models are often functions that relate the motion of roadmaps to surrogate signals derived from images or ECG, so that once the surrogates for fluoroscopic frames are obtained, the motion can be computed by the model. For cardiac interventions including PCI, the organ motion is mainly affected by respiratory and cardiac motion. Many previous works often built a motion model parameterized by a cardiac signal derived from ECG and a respiratory signal obtained from diaphragm tracking [33, 96, 106] or automatic PCA-based surrogate [36]. Some other works model only the respiratory motion in cardiac-gated images [56, 80, 94]. For a complete review on respiratory motion modeling, we refer readers to the survey article by McClelland et al. [72]. One limitation of the model-based approaches is that the motion models are often patient-specific, which requires training the model every time for a new subject. Additionally, once the surrogate values during inference are out of the surrogate range for building the model, e.g. for abnormal motion, extrapolation is needed, which may hamper accurate motion compensation.

6.1.3 Interventional / Surgical Tool Tracking

Tracking interventional tools is relevant for motion compensation [9, 13, 22, 69, 94]. In particular for PCI, the guiding catheter tip typically remains within the coronary ostium which is in the field of view during the largest part of the intervention, making it a suitable landmark for tracking. Baka et al. [13] have shown that catheter tip motion during PCI can be modeled as a combination of cardiac and respiratory motion. As using catheter tip displacement can only correct translational motion, Baka et al. [13] further showed that, compared to a rigid motion model for the respiratory motion, modeling only the translational part of the respiratory motion deteriorated the accuracy marginally, which confirms the observations in [95] that the rotational

part of respiratory motion is small. These findings motivate motion compensation for DCR through tracking the catheter tip in X-ray fluoroscopic sequences.

Many works have been proposed to address the problem of tracking interventional or surgical tools in medical images for various applications. The tracking methods from these works can be generally categorized into two kinds of approaches: tracking by detection, and temporal tracking.

The tracking by detection approaches treat tracking as a detection problem, which rely on features only from the current image without using information from previous frames. For example in electrophysiology procedure, as the catheters present specific features in shape or intensity, *ad hoc* methods were proposed based on, e.g. blob detection, shape constrained searching and model-/template- based detection [68,69]. Chang et al. [24] modeled the catheter tracking problem by optimizing the posterior in a Bayesian framework, in which the catheter was represented as a B-spline tube model and was tracked by fitting the B-spline to measurements based on gray intensity and vesselness image. Baur et al. [17] proposed a convolutional neural network (CNN) to detect catheter electrodes in X-ray images, which treated catheter detection as a segmentation problem. The method used a weighted cross-entropy loss to cope with the class imbalancing problem due to the small size of the target. In [31, 59] surgical instruments were tracked using a deep network having an encoder-decoder architecture. Their approaches combined instrument segmentation and detection in a multi-task learning problem to make the tool detection in a cluttered background more robust.

Different from tracking by detection, which relies solely on the current image, temporal tracking also uses information from previous frames. The temporal information can reduce the search space for detection, or put additional constraints in the model, making the tracking more robust.

Temporal information has been used in various ways. Some methods mainly relied on a detection model, but incorporate temporal information in the preprocessing [22] or post-processing [39] step or in the input [8, 87]. Approaches based on background estimation have been used for catheter [114] or guidewire [83] tracking. In these approaches, the background was recursively updated for every frame, and was used for enhancing the foreground containing instruments. Apart from those, many works adopted a Bayesian framework for tracking instruments via a *maximum a posteriori* (MAP) formulation. Representations based on key points [112], B-splines [48, 51, 79, 110], or segment-like features [108] have been used to model catheters or guidewires. Markov random field (MRF) was used to model the position or deformation of the control points in the B-spline [48, 51, 79, 112]. In the work by Vandini et al. [108], temporal information was incorporated in the prior term using Kalman filter. Particularly, learning-based approaches were used in several works to obtain the likelihood for a more robust measurement using probabilistic boosting tree [110, 111] or support vector regression [79]. In addition, temporal tracking models based on Bayesian filtering were also a popular approach for instrument tracking. Ambrosini et al. [9] used a hidden Markov model (HMM) to track catheter tip in a 3D vessel tree, for which the likelihood was obtained based on the 3D-2D registration outcome. Speidel et al. [99] used particle filters to track surgical tools in medical images. They used a likelihood based on the segmentation of instruments, and a

dynamic model that incorporates samples from two previous time steps. In a later work, Speidel et al. [100] used a multi-object particle filter to track multiple instrument regions simultaneously, in which a particle is the concatenation of the states of several objects.

Despite of many existing works on inverventional or surgical tool tracking in medical images, an automatic approach for tracking the tip of guiding catheter in X-ray fluoroscopy for PCI has not been investigated yet. The challenges of this task are: (1) different from the catheters for EP that can be viewed as blobs or a circle, the guiding catheter for PCI presents a dark tubular appearance which shows no prominent features; (2) the shape of the guiding catheter tip segment varies depending on the orientation of the C-arm, making feature-/model- based detection challenging; (3) the background may contain structures that have similar appearance to a catheter tip, such as vertebral structures or residual contrast agent, which makes robust tracking difficult.

6.1.4 Contributions

We propose and evaluate a novel approach for dynamic coronary roadmapping. The approach compensates changes in vessel shapes and cardiac motion by selecting the roadmap of the same cardiac phase through ECG alignment, and corrects the respiratory induced motion via tracking the tip of the guiding catheter. Our contributions are:

1. We develop a new way to perform dynamic coronary roadmapping on free breathing, non-cardiac-gated X-ray fluoroscopic sequences. Particularly, the respiratory-induced vessel motion is robustly compensated via the displacement of catheter tip.

2. We proposed a deep learning based method within a Bayesian filtering framework for online detection and tracking of guiding catheter tip in X-ray fluoroscopic images. The method models the likelihood term of Bayesian filtering with a convolutional neural network, and integrates it with particle filtering in a comprehensive manner, leading to more robust tracking.

3. We evaluate the proposed approach visually and quantitatively on clinical X-ray sequences, achieving low errors on both tracking and roadmapping tasks.

4. The proposed DCR method runs in real-time with a modern GPU, thus can potentially be used during PCI in real clinical settings.

6.2 Scenario Setup and Method Overview

The proposed method assumes that the scenario of performing dynamic coronary roadmapping to guide a PCI procedure consists of an offline phase and an online phase. An overview of the method is shown in Fig. 6.1.

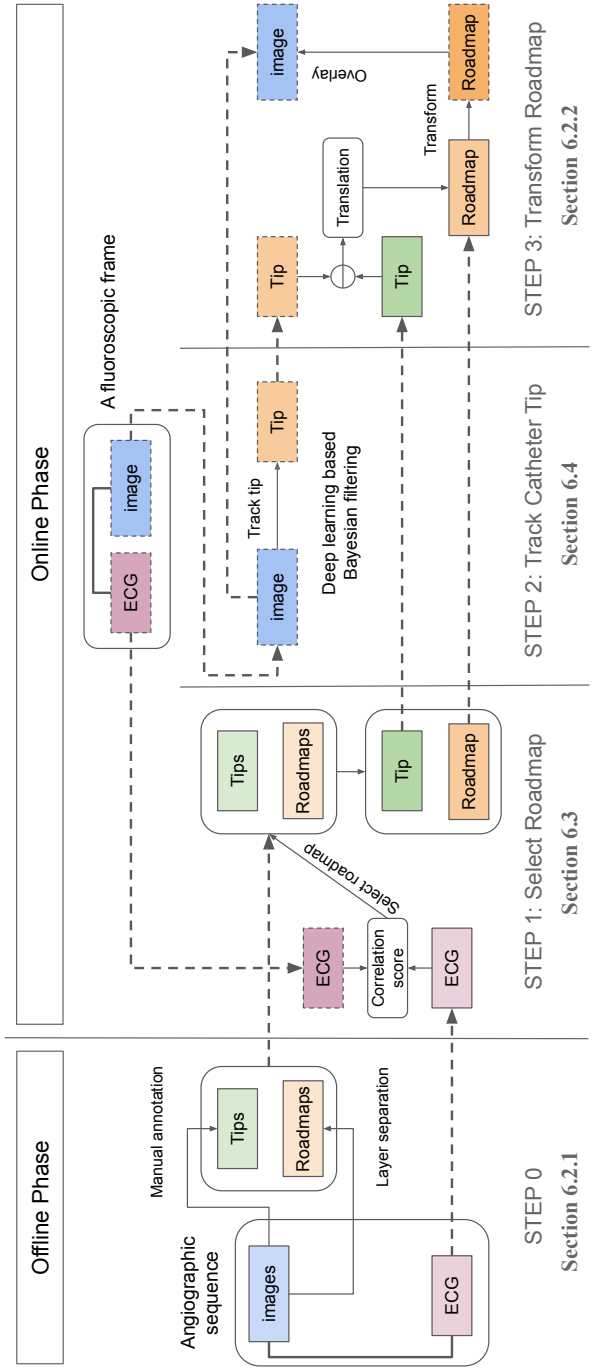


Figure 6.1: The overview of the proposed dynamic coronary roadmapping method. The colored blocks with a dash line border denote objects acquired in the online phase; the colored blocks with a solid line border are objects originated from the offline phase.

6.2.1 Offline Phase

This phase (Step 0 in Fig. 6.1) is performed off-line before the actual roadmapping is conducted. In this stage, roadmaps of coronary arteries containing multiple cardiac phases are created from an X-ray angiography sequence acquired with injection of contrast agent. A roadmap can be a vessel model in the form of centerlines, contours, masks, etc. It may also contain information of clinical interest, e.g. stenosis. Since the main focus of this paper is on accurate overlay of a roadmap, we do not investigate how to create the most suitable roadmaps, but use the images containing only vessels and catheters that are created using the layer separation method in [65] (Chapter 2 of this thesis) as the roadmaps to show the concept of dynamic coronary roadmapping. Along with the XA sequence, ECG signals are also acquired and stored for later selecting a roadmap that has similar cardiac phase to a given X-ray fluoroscopic frame in the online phase (see details in Section 6.3). Once the image sequence and ECG signals are acquired, the catheter tip location in every frame is obtained to serve as a reference point for roadmap transformation. In this work we manually annotated the catheter tip in the offline XA sequence. In real clinical scenarios, the annotation work can be done easily and efficiently by a technician who typically sits in front of monitors outside the catheterization lab to assist the procedure.

6.2.2 Online Phase

This is when the dynamic roadmapping is actually performed. In this phase, non-contrast X-ray fluoroscopic images with the same view angles as the roadmaps created during the offline phase are acquired sequentially. At the same time, ECG signals along with the roadmapping frames are also obtained and are compared with the stored ECG to select the most matched roadmap (Step 1 in Fig. 6.1; see details in Section 6.3). This is to compensate the change of vessel shape and position between frames due to cardiac motion. Simultaneously, the catheter tip location in the acquired X-ray fluoroscopic images is tracked online using the proposed deep learning based Bayesian filtering method in Section 6.4 (Step 2 in Fig. 6.1). The displacement of catheter tip between the current image and the selected roadmap image is then obtained and are applied to transform the roadmap. Finally, the transformed roadmap is overlaid on the current non-contrast frame to guide the procedure (Step 3 in Fig. 6.1).

6.3 ECG Matching for Roadmap Selection

Roadmap selection in this work is achieved by comparing the ECG signal associated with the fluoroscopic image and the ECG of the angiographic sequence, such that the most suitable candidate roadmap is selected where the best match of the ECG signals is found. The selected roadmap has the same (or very similar) cardiac phase with the X-ray fluoroscopic image, which compensates the difference of vessel shape and pose induced by cardiac motoin. An approach similar to the ECG matching method in [55] is used to accomplish this task.

To select roadmaps images based on ECG, a temporal mapping between X-ray images and ECG signal points needs to be built first. We assume that ECG signals and X-ray images are well synchronized during acqusition. In the offline phase, the beginning and the end of the image sequence are aligned with the start and end ECG signal points; the XA frames in between are then evenly distributed on the timeline of ECG. This way, a mapping between the stored sequence images and its ECG signal can be set up: for each image, the closest ECG signal point to the location of the image on the timeline can be found; for each ECG point, an image that is closest to this point on the timeline can be similarly located. Once the mapping is available, images with good vessel contrast and the ECG points that are associated to these images are selected from the XA sequence for the pool of roadmaps. In the online phase, similar to the approach in [55], for acquisition of each image, a block of N_{ECG} latest ECG signal points is constantly stored and updated in the history buffer. This is considered as the ECG signal corresponding to the fluoroscopic frame.

To compare the ECG signals associated with the angiographic sequence and the online fluoroscopic image, a temporal registration of the two signals using cross-correlation is applied [55]. The two ECG signals are first cross-correlated for every possible position on the signals, resulting in a 1D vector of correlation scores. The candidate frame for dynamic overlay is then selected as the one associated with the point on the ECG of the angiographic sequence that is corresponding to the highest correlation score.

6.4 Bayesian Filtering for Catheter Tip Tracking

Bayesian filtering is a state-space approach aiming at estimating the true state of a system that changes over time from a sequence of noisy measurement made on the system [11]. One popular application area of this approach is tracking objects in a series of images.

6.4.1 Theory of Bayesian Filtering

Bayesian filtering typically includes the following components: hidden system states, a state transition model, observations and a observation model. Let $\mathbf{x}_k \in \mathbb{R}^2$ ($k = \{0, 1, 2, ...\}$) denote the state, the location of guiding catheter tip in the k-th frame, a 2D vector representing the coordinates in the X-ray image space. The transition of the system from one state to the next state is given by the state transition model $\mathbf{x}_k = f_k(\mathbf{x}_{k-1}, \mathbf{v}_{k-1})$, where $\mathbf{v}_{k-1} \in \mathbb{R}^2$ is an independent and identically distributed (i.i.d.) process noise, $f_k : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}^2$ is a possibly nonlinear function that maps the previous state $\mathbf{x}_{k-1}$ to the current state $\mathbf{x}_k$ with noise $\mathbf{v}_{k-1}$. The observation $\mathbf{z}_k$ in this work is defined as the k-th X-ray image of a sequence, so that $\mathbf{z}_k \in \mathbb{R}^{w \times h}$, where w and h are the width and height of an X-ray image. We further define the observation model as $\mathbf{z}_k = h_k(\mathbf{x}_k, \mathbf{n}_k)$, where $\mathbf{n}_k \in \mathbb{R}^{n_k}$ is an i.i.d measurement noise (n_k is the dimension of $\mathbf{n}_k$), $h_k : \mathbb{R}^2 \times \mathbb{R}^{n_k} \to \mathbb{R}^{w \times h}$ is a highly nonlinear function that generates the observation $\mathbf{z}_k$ from the state $\mathbf{x}_k$ with noise $\mathbf{n}_k$. The state transition model f_k and the observation model h_k, respectively, can also be equivalently represented

using probabilistic forms, i.e. the state transition prior $p(\mathbf{x}_k|\mathbf{x}_{k-1})$ and the likelihood $p(\mathbf{z}_k|\mathbf{x}_k)$ from which $\mathbf{x}_k$ and $\mathbf{z}_k$ can be obtained by sampling.

With these definitions and $p(\mathbf{x}_0)$, the inital belief of $\mathbf{x}_0$, Bayesian filtering seeks an estimation of $\mathbf{x}_k$ ($k \geq 1$) based on the set of all available observations $\mathbf{z}_{0:k} = \{\mathbf{z}_i, i = 0, ..., k\}$ up to time k via recursively computing the posterior probability $p(\mathbf{x}_k|\mathbf{z}_{0:k})$ as Eq.(6.1) [11]:

$$p(\mathbf{x}_k|\mathbf{z}_{0:k}) \propto p(\mathbf{z}_k|\mathbf{x}_k) \underbrace{\int p(\mathbf{x}_k|\mathbf{x}_{k-1}) p(\mathbf{x}_{k-1}|\mathbf{z}_{0:k-1}) \mathrm{d}\mathbf{x}_{k-1}}_{p(\mathbf{x}_k|\mathbf{z}_{0:k-1})} . \tag{6.1}$$

Assuming the initial probability $p(\mathbf{x}_0|\mathbf{z}_0) = p(\mathbf{x}_0)$ is known, based on Eq.(6.1), Bayesian filtering runs in cycles of two steps: prediction and update. In the prediction step, the prior probability $p(\mathbf{x}_k|\mathbf{z}_{0:k-1})$, the initial belief of $\mathbf{x}_k$ given previous observations, is estimated by computing the integral in Eq.(6.1). In the update step, the prior probability is corrected by the current likelihood $p(\mathbf{z}_k|\mathbf{x}_k)$ to obtain the posterior $p(\mathbf{x}_k|\mathbf{z}_{0:k})$.

In Section 6.4.2, we will firstly introduce how to model the likelihood. Then in Section 6.4.3, a way to represent and efficiently approximate the posterior will be discussed. Finally in Section 6.4.4, a summary of the complete catheter tip tracking method will be given.

6.4.2 A Deep Learning based Likelihood

Directly modeling the likelihood $p(\mathbf{z}_k|\mathbf{x}_k)$ is challenging due to (1) the complexity of the generation process h_k and (2) the computational complexity of $p(\mathbf{z}_k|\mathbf{x}_k)$ for every value $\mathbf{x}_k \in \mathbb{R}^2$. In this work, we simplify the problem by only computing the likelihood $p(\mathbf{z}_k|\mathbf{x}_k)$ in the image pixel space, i.e. the integer pixel coordinate. For a subpixel $\mathbf{x}_k$, the value of $p(\mathbf{z}_k|\mathbf{x}_k)$ can possibly be approximated by interpolation. To this end, we propose to use a deep neural network $\mathcal{D}$ to approximate $p(\mathbf{z}_k|\mathbf{x}_k)$ for integer pixel locations. The network takes an image $\mathbf{z}_k$ as input and outputs a probability of observing the input $\mathbf{z}_k$ for every pixel location $\mathbf{x}_k$. Therefore, the approximated likelihood is a function of $\mathbf{x}_k$, denoted as $\mathcal{D}_{\mathbf{z}_k}(\mathbf{x}_k)$. Since $\mathbf{x}_k$ is defined within the scope of the image pixel space, $\mathcal{D}_{\mathbf{z}_k}(\mathbf{x}_k)$ is essentially a probability map having the same dimension and size with the input image $\mathbf{z}_k$, in which the entry at each location $\mathbf{x}_k^j$ ($j = 1, 2, \ldots, wh$) in the map represents the probability of observing $\mathbf{z}_k$ given $\mathbf{x}_k^j$. It is worth mentioning that the deep neural network is used for approximation of $p(\mathbf{z}_k|\mathbf{x}_k)$, which should be clearly distinguished from the generation model h_k that maps an $\mathbf{x}_k$ to $\mathbf{z}_k$. The existence of h_k is merely for the convenience of definition, its explicit form, however, is not required in the context of this work.

To obtain the training labels, we assume that there exists a mapping h_k, such that the training label can be defined as a distance-based probability map, i.e. the farther away $\mathbf{x}_k$ is from the ground truth tip location in the image $\mathbf{z}_k$, the less possible it is to observe $\mathbf{z}_k$ given $\mathbf{x}_k$ through the process h_k. This definition matches the intuition that from a location $\mathbf{x}_k$ that is far from the ground truth tip location, the probability of observing a $\mathbf{z}_k$ with the catheter tip being located at the ground truth

(a) (b) (c)

Figure 6.2: Input and ground truth labels for the deep neural network: (a) an input X-ray fluoroscopic image, (b) the binary catheter mask of (a) for catheter segmentation, (c) a 2D Gaussian PDF ($\sigma = 4$ px) for likelihood estimation for (a).

position should be low. For simplicity, a 2D Gaussian probability density function (PDF) $\mathcal{N}(\mathbf{x}_k; \mathbf{x}'_k, \sigma^2 I)$ centered at the ground truth tip location $\mathbf{x}'_k$ with variance $\sigma^2 I$ in the image space is used as the label to train the network (Fig. 6.2c). Note that this training label makes the estimation of $p(\mathbf{z}_k|\mathbf{x}_k)$ equivalent to a catheter tip detection problem such that the deep neural network learns features of catheter tip and outputs high probability at locations where the features are present. Due to this reason, we also call $p(\mathbf{z}_k|\mathbf{x}_k)$ "detection output" or "detection probability" and call the estimation of $p(\mathbf{z}_k|\mathbf{x}_k)$ "catheter tip detection" in the context of this paper.

The network that we use follows a encoder-decoder architecture with skip connections similar to U-net [91]. Additionally, similar to the work in [74], residual blocks [47] are adopted at each resolution level in the encoder and decoder to ease gradient propagation in a deep network. The encoder consists of 4 *down* blocks in which a residual block followed by a stride-2 convolution is used for extraction and down-scaling of feature maps. The number of feature maps is doubled in each downsampling step. The decoder has 4 *up* blocks where a transposed convolution of stride-2 is used for upsampling of the input feature maps. Dropout is used in the residual unit of the *up* block for regularization of the network. Between the encoder and the decoder, another residual block is used to process the feature maps extracted by the encoder. The detailed network architecture is shown in Fig. 6.3.

Due to similar appearance between a guiding catheter tip and corners of a background structure, such as vertebral bones, lung tissue, stitches or guidewires, ambiguity may exist when the network is expected to output only one blob in the probability map. To alleviate the issue, we adopt a similar strategy as [59], using a catheter mask (Fig. 6.2b) as an additional label to jointly train the network to output both the catheter segmentation heatmap and the likelihood probability map. The segmentation heatmap is obtained by applying a 1×1 convolution with ReLU activation on the feature maps of the last *up* block. To compute the likelihood probability map, a residual block is firstly applied on the feature maps of the last *up* block. The output feature maps are then concatenated with the segmentation heatmap as one additional channel, followed by a 1×1 convolution. Finally, to ensure the network detection

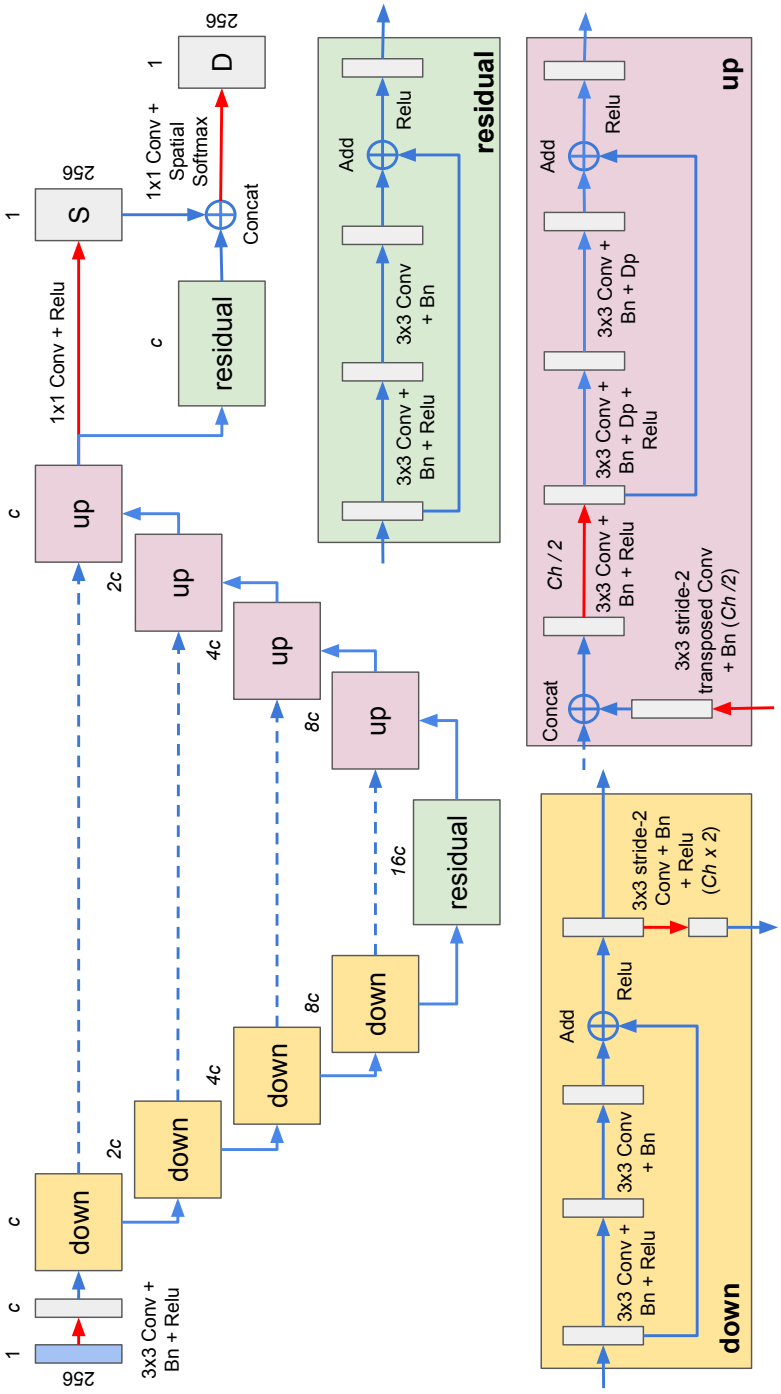


Figure 6.3: A joint segmentation and detection network for catheter tip detection. This figure shows an example network with 4 levels of depth (the number of down or up blocks). Meaning of abbreviations: *Conv*, 2D convolution; *Bn*, batch normalization; *Relu*, ReLU activation; *Dp*, dropout; *Concat*, concatenation; *Ch*, number of channels; *S*, segmentation output; *D*, detection output. The number above an image or feature maps indicates the number of channels; the number of channels in the residual network in a block is shown above the block; *c* is the basic number of channels, the channel number in the first down block. The number next to a rectangle denotes the size of the image or feature maps. Red arrows indicate a change of number of channels.

output fits the definition of a probability map on image locations, following the 1×1 convolution, a spatial softmax layer is computed as Eq.(6.2):

$$D_{k,l} = \frac{e^{A_{k,l}}}{\sum_{i,j} e^{A_{i,j}}}, \tag{6.2}$$

where A is the output feature map of the 1×1 convolution, $A_{i,j}$ denotes the value of A at location (i, j), D is the final output of the detection network, a 2D probability map representing $p(\mathbf{z}_k|\mathbf{x}_k)$. The details are shown in Fig. 6.3.

The training loss is defined as a combination of the segmentation loss and the detection loss. The segmentation loss L_s in this work is a Dice loss defined by Eq.(6.3):

$$L_s = 1 - \frac{2\sum_{i,j} M_{i,j} S_{i,j}}{\sum_{i,j} M_{i,j}^2 + \sum_{i,j} S_{i,j}^2} \tag{6.3}$$

where M denotes the ground truth binary catheter masks, S is the segmentation heatmap. The loss function for detection L_d is mean square error (MSE) given by Eq.(6.4):

$$L_d = \frac{1}{w \times h} \sum_{i \leq w, j \leq h} |T_{i,j} - D_{i,j}|^2 \tag{6.4}$$

where T denotes the ground truth PDF, w and h are the width and height of an image. The total training loss L is defined as Eq.(6.5):

$$L = L_s + \lambda L_d \tag{6.5}$$

where λ is a weight to balance L_s and L_d.

6.4.3 Approximation of the Posterior with Particle Filter

Once the deep neural network in Section 6.4.2 is trained, its weights are fixed during inference for computing the posterior $p(\mathbf{x}_k|\mathbf{z}_{0:k})$ for new data. Idealy, the network detection output $p(\mathbf{z}_k|\mathbf{x}_k)$ should be a Gaussian PDF during inference, as it is trained with labels of Gaussian PDFs. However, due to similar appearance of background structures or contrast residual, the detection output is unlikely to be a perfect Gaussian (possibly non-Gaussian or having multiple modes), which prevents the posterior $p(\mathbf{x}_k|\mathbf{z}_{0:k})$ in Eq.(6.1) being solved with an analytical method. In practice, the posterior can be approximated using a particle filter method [11].

Particle filter methods approximate the posterior PDF by a set of N_s random samples with associated weights $\{\mathbf{x}_k^i, w_k^i\}_{i=1}^{N_s}$ [11]. As N_s becomes very large, this discrete representation approaches the true posterior. According to [11], the approximation of the posterior $p(\mathbf{x}_k|\mathbf{z}_{0:k})$ is given by Eq.(6.6):

$$p(\mathbf{x}_k|\mathbf{z}_{0:k}) \approx \sum_{i=1}^{N_s} w_k^i \delta(\mathbf{x}_k - \mathbf{x}_k^i) \tag{6.6}$$

where $\delta(\cdot)$ is the Dirac delta function. The weight w_k^i can be computed in a recursive manner as Eq.(6.7) once w_{k-1}^i is known [11]:

$$w_k^i \propto w_{k-1}^i \frac{p(\mathbf{z}_k|\mathbf{x}_k^i)p(\mathbf{x}_k^i|\mathbf{x}_{k-1}^i)}{q(\mathbf{x}_k^i|\mathbf{x}_{k-1}^i, \mathbf{z}_k)} \tag{6.7}$$

where $q(\mathbf{x}_k|\mathbf{x}_{k-1}^i, \mathbf{z}_k)$ is an importance density from which it should be possible to sample $\mathbf{x}_k^i$ easily. For simplicity, a good and convenient choice of the importance density is the prior $p(\mathbf{x}_k|\mathbf{x}_{k-1}^i)$ [11], so that the weight update rule (6.7) becomes $w_k^i \propto w_{k-1}^i p(\mathbf{z}_k|\mathbf{x}_k^i)$.

A sample can be drawn from $p(\mathbf{x}_k|\mathbf{x}_{k-1}^i)$ in the following way. First, a process noise sample $\mathbf{v}_{k-1}^i$ is sampled from $p_v(\mathbf{v}_{k-1})$, the PDF of $\mathbf{v}_{k-1}$; then $\mathbf{x}_k^i$ is generated from $\mathbf{x}_{k-1}^i$ via the state transition model $\mathbf{x}_k^i = f_k(\mathbf{x}_{k-1}^i, \mathbf{v}_{k-1}^i)$. In this work, $p_v(\mathbf{v}_{k-1})$ is set to be a Gaussian $\mathcal{N}(\mathbf{0}, \sigma_v^2 I)$. The choice of motion model for f_k is important for an accurate representation of the true state transition prior $p(\mathbf{x}_k|\mathbf{x}_{k-1})$. A random motion cannot characterize well the motion of catheter tip in XA frames. In this work, we estimated the motion from adjacent frames using an optical flow method, as this approach 1) takes into account of the observation $\mathbf{z}_k$, which results in a better guess of the catheter tip motion, and 2) enables estimation of a dense motion field where the motion of a sample $\mathbf{x}_k^i$ can be efficiently obtained. Therefore, f_k is defined as Eq.(6.8):

$$\mathbf{x}_k = \mathbf{x}_{k-1} + \mathbf{u}_{k-1}(\mathbf{x}_{k-1}) + \mathbf{v}_{k-1} \tag{6.8}$$

where $\mathbf{u}_{k-1}(\cdot)$ is the motion from frame $k-1$ to frame k estimated with optical flow using the method in [34], $\mathbf{u}_{k-1}(\mathbf{x}_{k-1})$ is the motion from state $\mathbf{x}_{k-1}$.

Once samples are drawn and their weights are updated, the so-called "resampling" of the samples should be performed to prevent the degenaracy problem, where all but one sample will have negligible weight after a few iterations [11]. The resampling step resamples the existing samples according to their updated weights and then resets all sample weights to be $1/N_s$, so the number of effective samples which have actual contribution to approximate $p(\mathbf{x}_k|\mathbf{z}_{0:k})$ is maximized [11]. If the resampling is applied at every time step, the particle filter becomes a sampling importance resampling (SIR) filter, and the weight update rule follows Eq.(6.9).

$$w_k^i \propto p(\mathbf{z}_k|\mathbf{x}_k^i) \tag{6.9}$$

The final decision on catheter tip location in frame k can then be computed as the expectation of $\mathbf{x}_k$, $\hat{\mathbf{x}}_k = \int \mathbf{x}_k p(\mathbf{x}_k|\mathbf{z}_{0:k}) d\mathbf{x}_k$, which is in this case, the weighted sum of all samples:

$$\hat{\mathbf{x}}_k = \sum_{i=1}^{N_s} w_k^i \mathbf{x}_k^i. \tag{6.10}$$

6.4.4 Summary

The overall catheter tip tracking using a deep learning based Bayesian filtering method is summarized in Algorithm 3.

Algorithm 3 Deep learning based Bayesian filtering for online tracking of catheter tip in X-ray fluoroscopy

Require: $\{\mathbf{z}_0, \ldots, \mathbf{z}_T\}$ (sequentially observed frames), $\mathcal{D}$ (A trained network from Section 6.4.2), $p(\mathbf{x}_0)$ (the initial PDF), σ_v^2 (the variance of $\mathbf{v}_{k-1}$, $k = 1, \ldots, T$), T (number of frames for tracking), N_s (number of samples)

Draw $\mathbf{x}_0^i \sim p(\mathbf{x}_0)$, set $w_0^i = 1/N_s$, $\forall\, i = 1, \ldots, N_s$
for $k = 1$ **to** T **do**
  Compute $\mathbf{u}_{k-1}$ from $\mathbf{z}_{k-1}$ to $\mathbf{z}_k$ using the optical flow method in [34]
  for $i = 1$ **to** N_s **do**
    Draw $\mathbf{v}_{k-1}^i \sim \mathcal{N}(\mathbf{0}, \sigma_v^2 I)$
    Compute the motion of $\mathbf{x}_{k-1}^i$: $\mathbf{u}_{k-1}^i = \mathbf{u}_{k-1}(\mathbf{x}_{k-1}^i)$
    Draw $\mathbf{x}_k^i \sim p(\mathbf{x}_k|\mathbf{x}_{k-1}^i)$: $\mathbf{x}_k^i = \mathbf{x}_{k-1}^i + \mathbf{u}_{k-1}^i + \mathbf{v}_{k-1}^i$
    Update weight $w_k^i = p(\mathbf{z}_k|\mathbf{x}_k^i) = \mathcal{D}_{\mathbf{z}_k}(\mathbf{x}_k^i)$
  end for
  Normalize $w_k^i \leftarrow w_k^i / \sum_{i=1}^{N_s} w_k^i$, $\forall\, i = 1, \ldots, N_s$
  Prediciton in frame k: $\hat{\mathbf{x}}_k = \sum_{i=1}^{N_s} w_k^i \mathbf{x}_k^i$
  Resample $\{\mathbf{x}_k^i, w_k^i\}_{i=1}^{N_s}$ using the method in [11] (so all w_k^i are set to $1/N_s$ again)
end for

6.5 Experimental Setup

6.5.1 Data

Anonymized clinical imaging data were used for our experiments. The data were acquired with standard clinical protocol using Siemens AXIOM-Artis system, and are from 55 patients who underwent a PCI procedure at the Department of Cardiology at Erasmus MC in Rotterdam, Netherlands. Out of these data, we selected data from 37 patients which were acquired since the year 2014 to develop our method, and used the data from the other 18 patients acquired before the year 2013 for evaluation. The detailed information about the data is listed in Table 6.1.

In order to evaluate the proposed roadmapping method, for which an off-line angiographic sequence is required for roadmap preparation and an online fluoroscopic sequence taken from the same C-arm position is needed for performing the actual roadmapping (see Section 6.2), we selected the contrast frames from a real clinical sequence to simulate the off-line sequence, and chose the non-contrast frames from the same clinical sequence to simulate the online sequence. The selected contrast sequence were ensured sufficiently long to cover at least one complete cardiac cycle.

6.5.2 Data Split for Catheter Tip Detection and Tracking

To develop the catheter tip tracking method, 1086 X-ray fluoroscopic images selected from 260 non-contrast sequences of 25 patients from the development set were used for training the network from Fig. 6.3; 404 images from 94 non-contrast sequences of another 12 patients from the development set were used as validation set for the network model and hyperparameter selection. In the training and validation sets,

Table 6.1: Basic information of the acquired X-ray image data for our experiments. The number in the parenthesis next to the pixel size indicates the possible image size.

Data	Development	Evaluation
No. patients	37	18
No. sequences	354	34
Frame rate (fps)	15	15
Image size (px)	512 × 512	512 × 512
	600 × 600	600 × 600
	776 × 776	776 × 776
	960 × 960	1024 × 1024
	1024 × 1024	
Pixel size (mm)	0.108 (1024)	0.139 (1024)
	0.139 (1024)	0.184 (600)
	0.184 (600)	0.184 (776)
	0.184 (776)	0.184 (1024)
	0.184 (960)	0.216 (512)
	0.184 (1024)	0.279 (512)
	0.216 (512)	

Table 6.2: Dataset of training, validation and test for detection and tracking of catheter tip in X-ray fluoroscopic frames.

	Training (detection)	Validation (detection)	Validation (tracking)	Test (tracking)
No. patients	25	12	12	18
No. sequences	260	94	88	34
No. frames	1086	404	1583	1355
Continous frames?	No	No	Yes	Yes

4-5 frames were randomly selected from each sequence, which are not necessarily continuous. To tune the parameters for tracking, 1583 images from 88 sequences out of the 94 from the same 12 patients of the validation set were used (6 sequences were not selected for this task due to very short sequence length not more than 5 frames). Finally, to evaluate catheter tip tracking accuracy, 1355 images from 34 non-contrast sequences of 18 patients from the evaluation set were used for testing. The frames selected for tracking from each sequence must be continuous; the number of selected frames for tracking might vary, depending on the number of the non-contrast frames in the sequences. Details of the datasets for training, validation and test are listed in Table 6.2.

6.5.3 Experimental Settings for Training the Deep Network

6.5.3.1 Preprocessing

As the image data have different size ranging from 512×512 to 1024×1024, all images were resampled to a grid of 256×256 before being processed by the neural network. In addition, the image intensities were rescaled to the range from 0 to 1.

6.5.3.2 Training label

The standard deviation σ of the Gaussian PDF for the training label of the detection network was set to 4 pixels in the resampled image space (256×256). This choice corresponds to the estimation of the maximal possible catheter tip radius. An example of the Gaussian PDF is shown in Fig. 6.2c.

6.5.3.3 Data Augmentation

To increase the number of training samples and their diversity, data augmentation was used. The augmentation includes geometric transformation such as flipping (left-right, up-down), rotation of multiple of 90 degrees, random affine transformation (translation -10 to 10 px, scaling 0.9 to 1.1, rotation -5 to 5 degrees, shear -5 to 5 px), random elastic deformation (deformation range -4 to 4 px, grid size of control points 64 px). A training sample has 0.5 probability of being processed with one of the transformations. The probability for applying each transformation is: flipping left-right (1/24), flipping up-down (1/24), rotation of multiple of 90 degrees (1/12), affine transformation (1/6), elastic deformation (1/6), no transformation (1/2). To make the trained model robust to noise, in addition to the geometric transformations, we also augmented data by adding Gaussian noise to the pixel value with a zero mean and a standard deviation between 0.01 and 0.03. The probability of adding the noise is 0.5.

6.5.3.4 Evaluation Metric

To select hyperparameters and model weights in training, an evaluation metric is required. As the deep network is essentially a catheter tip detector, accurate detection of the tip location is desired. Therefore, we chose the location with the highest value in the detection output, and computed the Euclidean distance between the chosen location and the ground truth tip coordinate as the evaluation metric to tune the deep network.

6.5.3.5 Training Settings

Finally, the λ value in the training loss Eq. (6.5) was set to 10 to make the scale of the two terms similar. Adam optimizer was used to minimize the loss function with a learning rate 0.0001. The number of training samples in a batch is 4. The network was trained with 100 epochs to ensure convergence.

6.5.4 Setup for Evaluating Dynamic Coronary Roadmapping

It is in general a challenge to evaluate the roadmapping accuracy, as the structure of interest, e.g. coronary arteries in our case, is not directly visible in the target image. One possible choice introduced in [119] is to use the guidewire as a surrogate of the target vessel centerline in non-contrast images, as guidewire is always inside vessels and commonly present in image sequences during interventions. In this work, we follow a similar strategy to evaluate the accuracy of dynamic coronary roadmapping.

The first step is to select frames for roadmapping evaluation. From each non-contrast sequence in the test set for tracking in Section 6.5.2, we uniformly select 8-20 frames to annotate guidewire. The number of the selected frames from each sequence depends on the sequence length, the frame interval size and guidewire visibility. For some rare cases in our data where no guidewire is present in the image, we discarded that non-contrast frame, and chose those frames with little vessel contrast from the same sequence and annotated the vessel centerline. The selection results in 409 frames from 34 sequences in total. Once the target non-contrast frames for evaluating roadmapping are chosen, their corresponding angiographic frames were found using the ECG matching method in Section 6.3. We then annotated the centerline of the vessel corresponding to the guidewire in the non-contrast frames.

The next step is performing the transformation of the labelled vessel centerline from the angiographic frame to its corresponding target non-contrast frame via displacement of catheter tip in the two frames. This step simulates the roadmapping transformation in the last step in Fig. 6.1.

Finally, the distance between the guidewire annotation in the target frame and the transformed vessel centerline is reported as the roadmapping accuracy. In order to compute the distance between two point sets of annotations (e.g. Fig. 6.4a), point-point correspondence between the two sets is required (Fig. 6.4b). The point sets were firstly resampled with the point interval being 1 mm. We then followed the approach in [107] to find such correspondences which minimizes the sum of the Euclidean distance of all valid point-point correspondence paths. This way guarantees no cross-over connection and each point in one set is connected to at least one point in the other set. As the annotated point sets may have different size, the point correspondences to endpoints are excluded such that we only focused on the distance between corresponding sections, not the entire centerlines (Fig. 6.4c). Once the point-point correspondence is available, the distance between the two points in a pair can be used for evaluating the accuracy of DCR.

6.5.5 Implementation

The proposed method was developed in Python. The framework used for developing the deep learning approach for likelihood approximation is PyTorch. The major experiments of dynamic coronary roadmapping were performed on a computer with an Intel Xeon E5-2620 v3 2.40 GHz CPU and 16 GB RAM running Ubuntu 16.04. The deep neural network and the optical flow method were running on an NVIDIA GeForce GTX 1080 GPU. The approach for evaluating dynamic coronary roadmapping was developed and running in MeVisLab on a computer with an Intel Core i7-4800MQ

(a) (b) (c)

Figure 6.4: Correspondence between the labelled guidewire (green) and the transformed vessel centerline (red). The yellow lines connecting the two point sets illustrate the correspondence between red and green points.

2.70 GHz CPU and 16 GB RAM running Windows 7.

6.6 Experiments and Results

6.6.1 Training the Deep Neural Network

The purpose of this experiment is to train the deep neural network to output reasonable likelihood probability map. The network hyperparameters were tuned to optimize the detection performance.

The training and validation data for detection mentioned in Section 6.5.2 were used for training the deep neural network. The evaluation metric mentioned in Section 6.5.3, the mean Euclidean distance between the ground truth and the predicted tip location averaged over all validation frames, was used as the validation criteria for selecting the optimal training epoch and the network hyperparameters. When we evaluated hyperparameter settings, we firstly selected the training epoch with the lowest mean validation error for each setting, then the settings were compared using the model weights (trainable network parameters) of their chosen epochs.

The network hyperparameters we investigated in the experiments include (1) the basic channel number, i.e. the number of channels or feature maps in the first down block, (2) the network depth level, the number of down or up blocks, and (3) the dropout probability.

The validation errors for different hyperparameter settings using the experimental settings in Section 6.5.3 are shown in Table 6.3. The table shows that the hyperparameter setting with the lowest mean error, which has 4 level in depth and 64 channels in the first down block, achieves a validation error of about 2 mm. The table also shows other good choices of network architecture that have a small validation error (shown in red in Table 6.3): 32 channels in the first down block with 4 or 5 levels in depth, or 64 channels with 3 or 4 depth levels. The dropout regularization improves the accuracy of the model, compared to the ones without dropout.

Table 6.3: Validation errors (mm) for different hyperparameter settings. Red cells show the settings with the 10 smallest validation errors. *bold* number indicates the setting with the lowest error.

Basic Number of Channels	Depth Level	Dropout					
		none	0.1	0.2	0.3	0.4	0.5
8	3	5.43	4.99	5.02	5.37	4.38	4.24
	4	4.17	4.45	4.25	5.04	4.75	4.36
	5	3	4.14	3.53	4.28	3.95	4.11
16	3	3.74	4.29	3.57	4.11	3.74	3.4
	4	3.36	3.11	3.63	3.33	3.36	3.78
	5	3.38	2.89	3.16	2.52	2.71	2.74
32	3	2.99	3.02	3.26	2.82	3.26	2.56
	4	2.87	2.34	2.46	2.6	2.65	2.27
	5	3.04	2.51	2.21	2.29	2.3	2.25
64	3	2.19	2.54	2.34	2.27	2.26	2.49
	4	2.55	2.31	**2.04**	2.44	2.22	2.27
	5	2.42	2.29	2.73	2.77	2.61	2.85

The learning curves of the training process with the chosen hyperparameter setting are illustrated in Fig. 6.5. The curves indicate that both segmentation and detection reach convergence after training 100 epochs.

We did not investigate a model with more than 64 channels or 5 depth levels, because (1) it will further increase the processing time which makes online applications less feasible; (2) the results in Table 6.3 show that such a setting (64 channels, 5 depth levels) starts increasing the validation error compared to those less complex models.

The subsequent experiments will be based on the network trained with the chosen hyperparameter setting indicated in Table 6.3 (64 channels, 4 depth levels, dropout 0.2).

6.6.2 Catheter Tip Tracking

The purpose of this experiment is to assess the accuracy of catheter tip tracking with the proposed method in Section 6.4. Guiding catheter tip is tracked in X-ray fluoroscopy using Algorithm 3 based on a trained network with the optimal hyperparameter setting from Section 6.6.1. First, the parameters of the optical flow method used in Algorithm 3 and particle filter were tuned on the validation data for tracking in Section 6.5.2. We then evaluated the tracking accuray with the tuned optimal parameter setting on the test dataset, and compared the proposed tracking method with alternative approaches using only the detection network in Section 6.4.2 or using only optical flow. Finally, we investigated tracking accuracy with different ways of tip initialization in the first frame.

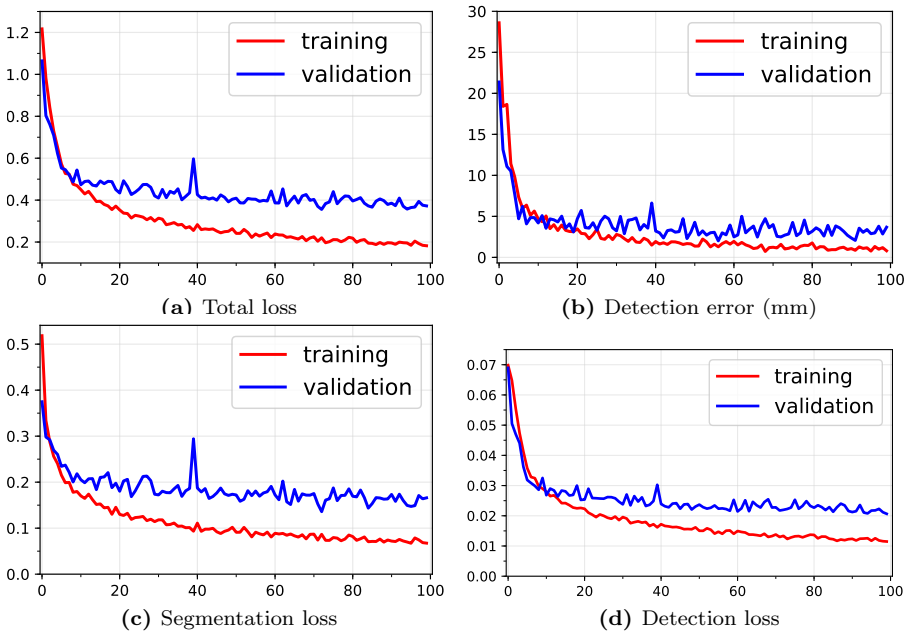


Figure 6.5: Learning curves for the chosen hyperparameter setting.

6.6.2.1 Tuning Optical Flow Parameters

The approach in [34] was used as the optical flow implementation in Algorithm 3. A grid search to find the optimal parameter setting was done on the following parameters of the method: (1) the image scale to build the pyramids, (2) the number pyramid levels, (3) the averaging window size, (4) the number of iterations, (5) the size of the pixel neighborhood used to find polynomial expansion in each pixel, and finally (6) the standard deviation of the Gaussian that is used to smooth derivatives used as a basis for the polynomial expansion.

The above parameters were tuned independently of the deep neural network, as optical flow directly estimates the catheter tip motion between two frames. To tune the parameters, we tracked the catheter tip in X-ray fluoroscopy starting from the ground truth tip position in the first frame using the motion field between two adjacent frames estimated with optical flow. The average and median distance between the tracked tip position and the ground truth were used as the evaluation criteria for the tuning.

The method in [34] was implemented by using the OpenCV function `calcOpticalFlowFarneback`. With consideration of the suggested parameter values from the documentation, the parameter setting chosen for optical flow from the grid search is `pyr_scale` = 0.5, `levels` = 3, `winsize` = 10, `iterations` = 30, `poly_n` = 5, `poly_sigma` = 1.1. Details of the parameters can be found on the function documen-

tation page[1].

6.6.2.2 Tuning Particle Filter Parameters

The parameters to tune for the particle filter are the number of samples N_s and the variance of process noise σ_v^2. When tuning them, we fixed the parameters of the trained network and the optical flow method, and used their optimal parameter settings during this experiment. Following Algorithm 3, we tracked the catheter tip from the ground truth position (probability map) in the first frame, and used the mean and median distance between the tracked and the true position as the validation metric.

The tracking results on the validation (tracking) set are shown in Table 6.4. The table shows that good choices for σ_v are 4 and 5, for N_s are 1000 and 10000. By considering the mean, the standard deviation and the median of tracking errors, the parameter setting $\sigma_v = 5$, $N_s = 1000$ was chosen for subsequent experiments.

Table 6.4: Catheter tip tracking errors (mm) on the validation (tracking) dataset of different parameter settings for particle filter. The tracked tip point was rounded to the pixel center. The error of all images (mean ± std / median) are presented. Red cells show the good choices of parameters; *bold* number indicates the chosen setting .

σ_v	N_s		
(px)	100	1000	10000
3	1.52 ± 2.19 / 0.79	1.49 ± 2.18 / 0.79	1.48 ± 2.18 / 0.79
4	1.50 ± 2.17 / 0.79	1.46 ± 2.17 / 0.79	1.47 ± 2.18 / 0.79
5	1.52 ± 2.21 / 0.79	**1.47 ± 2.17 / 0.74**	1.47 ± 2.19 / 0.74
6	1.53 ± 2.39 / 0.79	1.49 ± 2.33 / 0.79	1.48 ± 2.29 / 0.74
7	1.56 ± 2.42 / 0.79	1.50 ± 2.29 / 0.74	1.50 ± 2.39 / 0.74
8	1.58 ± 2.41 / 0.79	1.51 ± 2.40 / 0.74	1.51 ± 2.42 / 0.74
9	1.56 ± 2.22 / 0.79	1.53 ± 2.43 / 0.79	1.52 ± 2.45 / 0.61
10	2.25 ± 6.18 / 0.79	1.54 ± 2.46 / 0.79	1.53 ± 2.47 / 0.61

6.6.2.3 Tracking Methods Evaluation

In this experiment, the proposed tracking method in Algorithm 3 uses the ground truth tip probability map of the first frame as the initial PDF $p(\mathbf{x}_0)$ to draw samples. This method is referred to as "Tracking". In addition, we compared the proposed method with three alternatives. The first one tracks catheter tip using only the detection network in Section 6.4.2 with the chosen network architecture and trained parameters in Section 6.6.1, therefore, no temporal information is used. This method is referred to as "Detection (Net)". The other two methods in this experiment use

[1] `https://docs.opencv.org/2.4/modules/video/doc/motion_analysis_and_object_tracking.html?`

Table 6.5: Catheter tip tracking errors (mm) of 4 different methods on the test (tracking) dataset

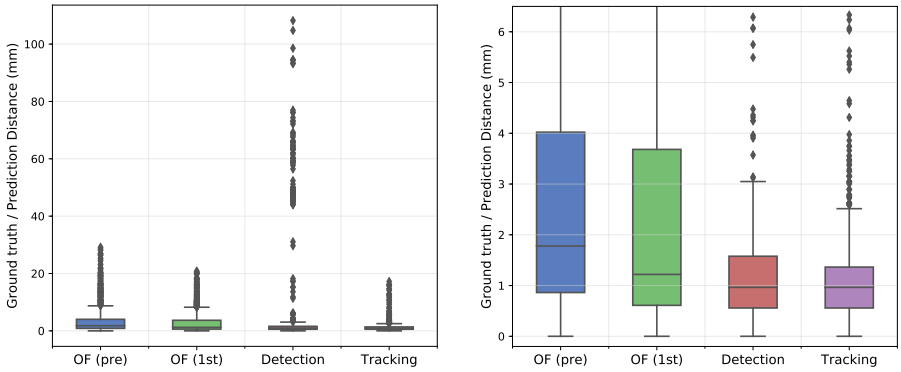
Evaluation Metrics	Optical Flow (previous)	Optical Flow (first)	Detection Net (Section 6.4.2)	Tracking
Maximal error of all images	29.16	20.83	108.20	**17.72**
Median error of all images	1.78	1.22	**0.96**	**0.96**
Mean error of all images	3.74 ± 4.93	3.05 ± 4.05	5.62 ± 15.91	**1.29 ± 1.76**
Average of sequence median error	2.35 ± 2.52	2.64 ± 3.52	6.26 ± 17.11	**1.03 ± 0.49**
Average of sequence mean error	2.59 ± 2.69	3.31 ± 2.81	6.83 ± 13.88	**1.29 ± 0.94**

only optical flow to track catheter tip starting from the ground truth tip position in the first frame. The motion field towards the current frame, estimated by the two methods, was based on the deformation from the previous frame or the first frame in the sequence, respectively. The same implementation setting as in Section 6.6.2.1 was used for these two methods. They are called "Optical Flow (previous)" and "Optical Flow (first)", or in short form, "OF (pre)" and "OF (1st)".

The tracking accuracies of all methods reported in this section were obtained on the test set from Table 6.2. The mean, the median and the maximal tracking error between the predicted and the ground truth tip position of all test images are reported in Table 6.5. In addition, as the sequences in the test set have different lengths, we also computed the mean and the median error per sequence, and report the the average of the sequence mean and median errors, so that each sequence contributes equally in these metrics. Table 6.5 shows that the results from the detection network have large average errors which are caused by some completely failed cases. The proposed tracking method has median errors of about 1 mm and mean errors of about 1.3 mm. It achieves the lowest errors compared to the other 3 methods on all listed evaluation criteria.

Fig. 6.6 illustrates the boxplots of tracking errors made by the 4 methods on all test images. It shows that the proposed tracking approach outperforms the detection method by avoiding making extremely large errors (Fig. 6.6a); meanwhile, it maintains as accurate as the detection method for cases with small errors, and is more accurate than the methods based solely on optical flow (Fig. 6.6b).

Fig. 6.7 shows longitudinal views of tracking errors of the 4 methods on 4 example sequences. Although the optical flow methods show high accuracy when the target is on the track (row 4), they present periodic error patterns in two sequences due to large cardiac motion. The detection method shows peaks of large errors, this is because temporal relation between frames is not modeled by the approach, thus the detection on different frames is independent of each other. The proposed tracking method



(a) Overall view of tracking errors **(b)** A zoom-in view of (a)

Figure 6.6: Tracking errors for the 4 methods on all test images.

overcomes the problems that other methods have and presents accurate detection on these 4 sequences. The tracking results of the 4 methods on example frames from the 4 sequences are illustrated in Fig. 6.8.

Fig. 6.9 illustrates how the proposed tracking method works on the same 4 frames in Fig. 6.8. It shows that the prior hypotheses (samples) assists to focus on the correct target location and results in reliable posterior estimation, especially when the detection produces ambiguity in cases of multiple catheters or contrast residual presented in images.

6.6.2.4 Catheter Tip Initialization

In this experiment, the initial PDF $p(\mathbf{x}_0)$ from which samples are drawn in the proposed tracking is investigated (Algorithm 3). In particular, we explored and evaluated the tracking accuracy with an automatic initialization using the probability map obtained from the trained detection network in Section 6.4.2 with the chosen setting in Section 6.6.1.

Fig. 6.10 shows the boxplot of tracking errors on all test images with automatic initialization (Auto) and manual initialization (Manual) for which the ground truth tip probability map of the first frame was used. The tracking with automatic initialization presents an accuracy similar to the one with manual initialization for small tracking errors, but has more large tracking errors which influence the mean error over all test images (Table 6.6). We, therefore, defined the tracking errors on the right side of the gap in the boxplot (> 40 mm) as outliers, and explored the statistics without those outliers.

Table 6.6 indicates that, the mean and median error of the tracking with automatic initialization excluding the outliers are only slightly higher than the tracking with manual initialization and the detection method. While the tracking with automatic initialization has 100 outliers in total from 6 sequences, the detection method that has 10 sequences containing 106 outliers.

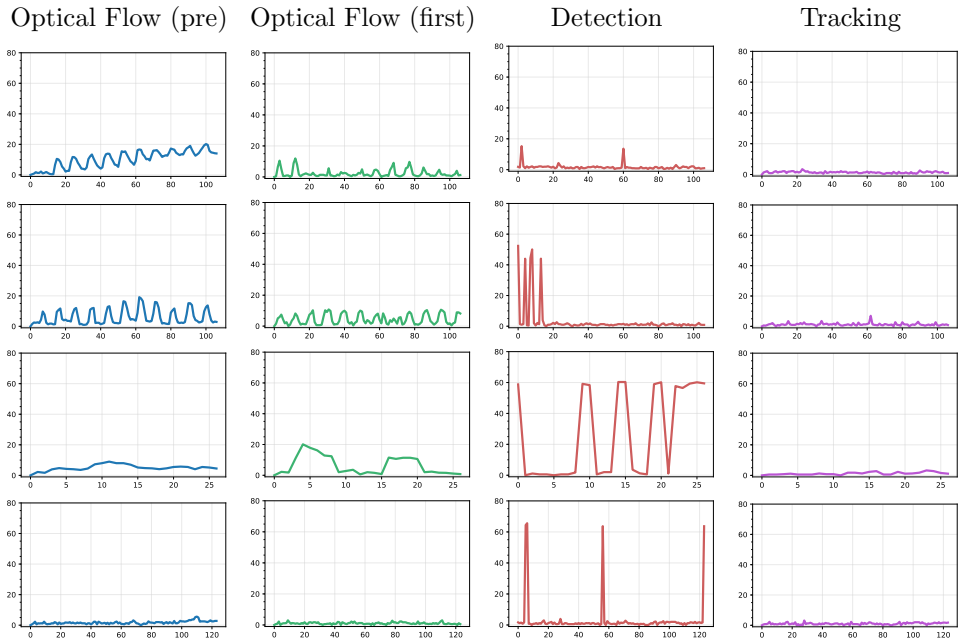


Figure 6.7: Longitudinal view of tracking errors made by the 4 methods on 4 test sequences (one sequence per row). The x-axis denotes the time steps of a sequence, the y-axis is the tracking error (mm).

Unlike the detection method for which the outliers are mainly presented as the peaks in the longitudinal views (Fig. 6.7), the outliers for the tracking with automatic initialization are more consistent over time. Fig. 6.11 shows the temporal change of tracking errors for the 6 sequences with outliers using the tracking with automatic initialization. For the 3 sequences on the top row, the tracking with automatic initialization makes large errors at the beginning, but becomes accurate very fast in a few frames; for the 3 sequences on the bottom row, however, the tracking errors remain large till the end of the sequences.

Fig. 6.12 shows example frames to give an insight of the tracking with automatic initialization on the 6 sequences in Fig. 6.11. For the 3 sequences on the top row (Fig. 6.12a), although the initialization on the first frame (frame 0) is overall not correct, the true tip positions are still covered by some samples; once the detection in subsequent frames is correct, the tracker can still converge to the right target. For the 3 sequence on the bottom row (Fig. 6.12b), the initializations of samples are ambiguous in frame 0; the detection in subsequent frames focuses on a wrong area also given by the initial samples due to residual of contrast agent or multiple catheters, the tracker then tends to find the wrong target.

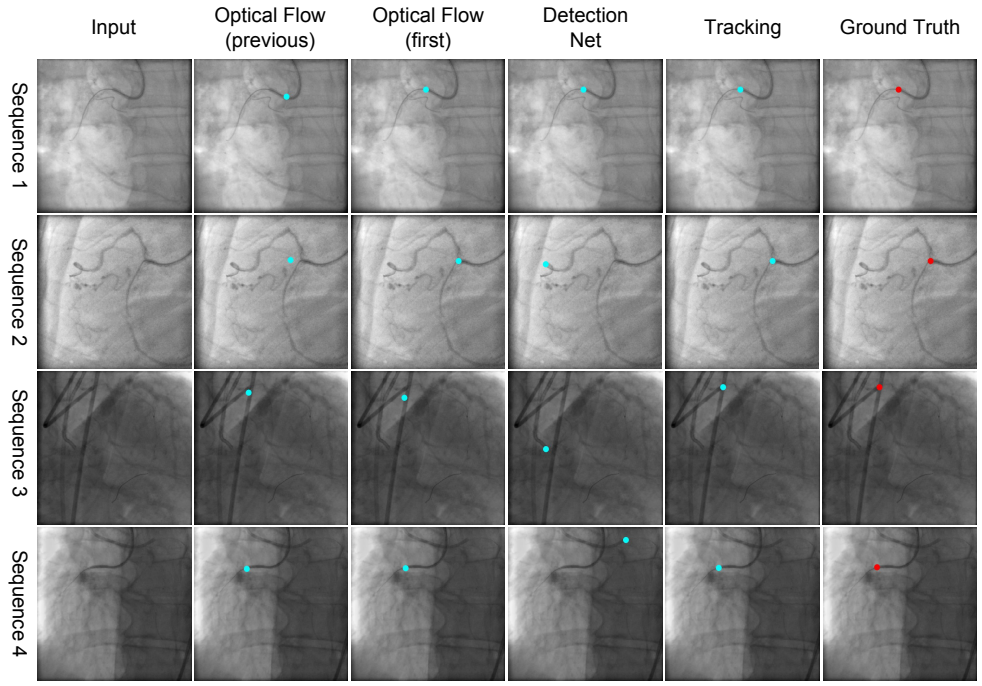


Figure 6.8: Tracking results on example frames from the same 4 sequences in Fig. 6.7. The blue point indicates the predicted catheter tip location; the red point shows the ground truth location. (Best viewed in color)

6.6.3 Dynamic Coronary Roadmapping

In this experiment, the accuracy of dynamic coronary roadmapping using the proposed method with manual tip initialization was evaluated. For roadmap selection with ECG matching (Section 6.3), the number of online ECG signal points N_{ECG} was manually determined so that the ECG signal stored in the buffer corresponding to 12 X-ray frames (0.8 second in acquisition time). Following the setup in Section 6.5.4, we used the distance between the two points in each point pair as the evaluation metric for DCR (the length of a yellow line segment in Fig. 6.4). As each frame may have different numbers of point pairs, depending on the length of the target guidewire, the average point pair distance per frame was also computed for evaluation. These distances were evaluated on 409 selected frames with manual annotation of guidewires and vessel centerlines (Section 6.5.4).

In the experiment, we compared the DCR with the proposed tracking method to those with manual tip tracking and without tracking. All three approaches were based on the same ECG matching method (Section 6.3) for selecting roadmaps. The accuracy of the DCR without tracking in Table 6.7 shows that the mean distances are reduced to less than 3 mm by compensating only cardiac motion via roadmap selection with ECG matching. Table 6.7 also shows that the DCR with the proposed

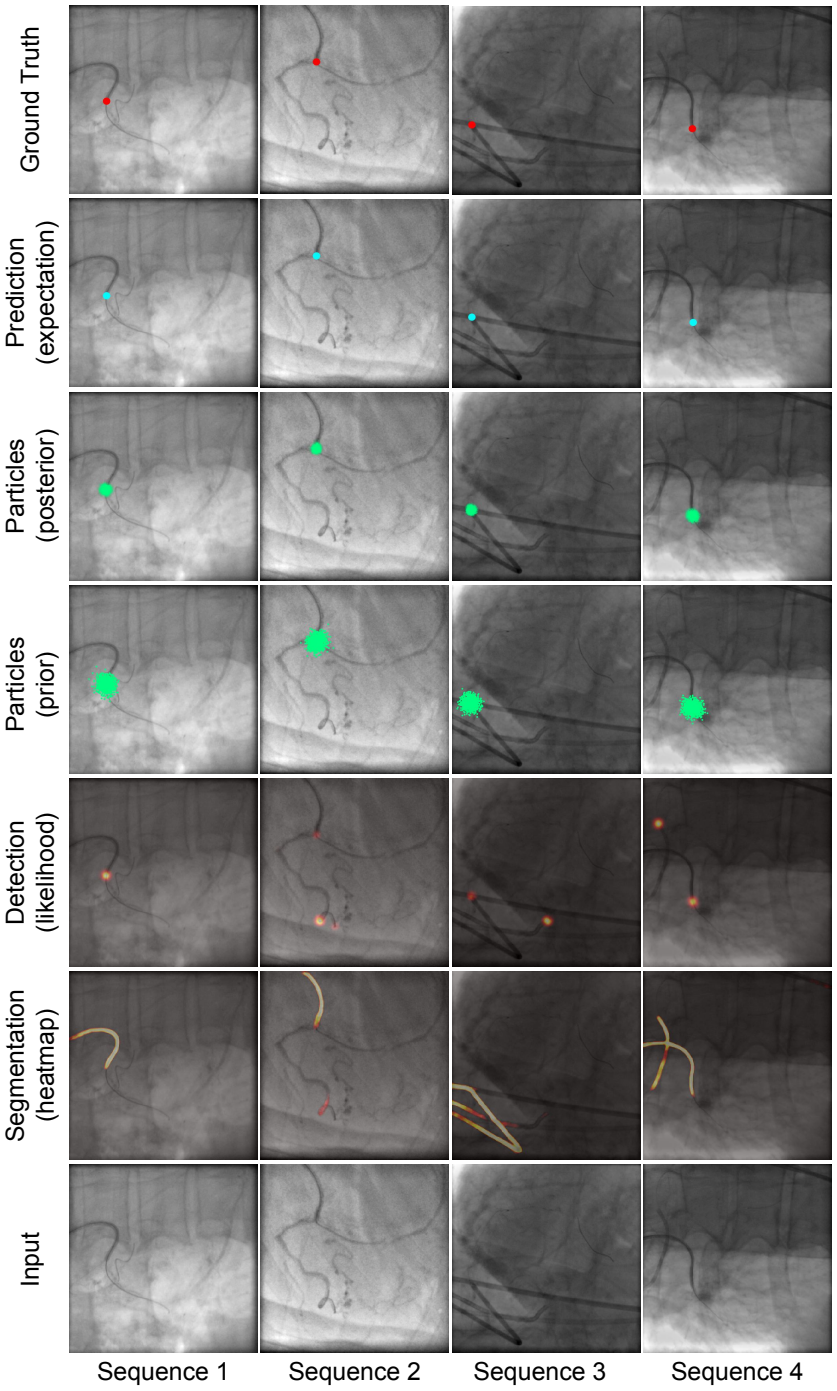


Figure 6.9: Workflow of the proposed tracking method on the same 4 frames in Fig. 6.8. The high probability is shown with bright color in the detection map. Samples or particles are presented as green dots. The blue point indicates the predicted catheter tip location; the red point shows the ground truth location. (Best viewed in color)

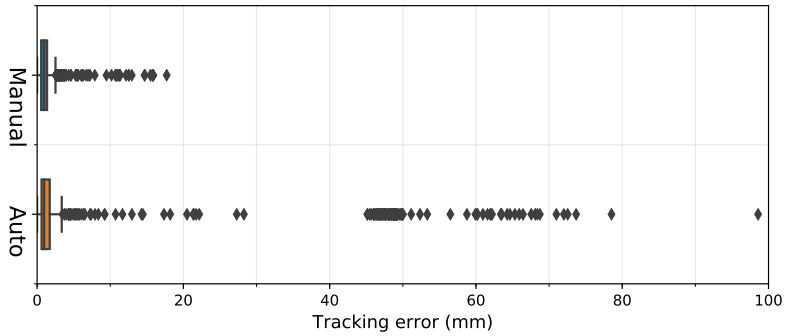


Figure 6.10: Catheter tip tracking errors (mm) with manual and automatic initialization.

Table 6.6: Catheter tip tracking errors (mm) of detection and tracking with manual and automatic initialization

	Detection	Tracking	
		Manual init.	Automatic init.
Maximal error	108.20	17.23	98.58
Median error	0.96	0.96	0.96
Mean error	5.62 ± 15.91	1.29 ± 1.76	5.16 ± 13.91
No. of outliers (> 40 mm)	106	0	100
No. of sequences with outliers	10	0	6
Maximal error of inliers	31.06	17.23	28.28
Median error of inliers	0.96	0.96	0.96
Mean error of inliers	1.17 ± 1.78	1.29 ± 1.76	1.34 ± 2.15

method achieves median distances of about 1.4 mm and mean distances of about 2 mm. The boxplots of the distances of all point pairs and the frame mean point distances of all 409 evaluation frames are illustrated in Fig. 6.13. The comparison of the three DCR approaches from Table 6.7 and Fig. 6.13 indicates that the accuracy of the proposed DCR method has shown improvement over the DCR without tracking and is only slightly less than the DCR with manual tip tracking.

Table 6.8 shows how the frame mean point distances of the 409 evaluation frames are distributed. The DCR with the proposed method has similar error distribution as the one with manual tip tracking: they both have about 1/3 of the distances less than 1 mm and 1/3 of the distances between 1 and 2 mm. The proposed method has slightly more distances larger than 5 mm than manual tip tracking. Both methods are more accurate than the DCR without tracking on intervals of small errors (< 2 mm).

Fig. 6.14 shows overlays of selected roadmaps on example frames of 4 sequences with the three DCR approaches. The DCR without tracking presents mismatch of

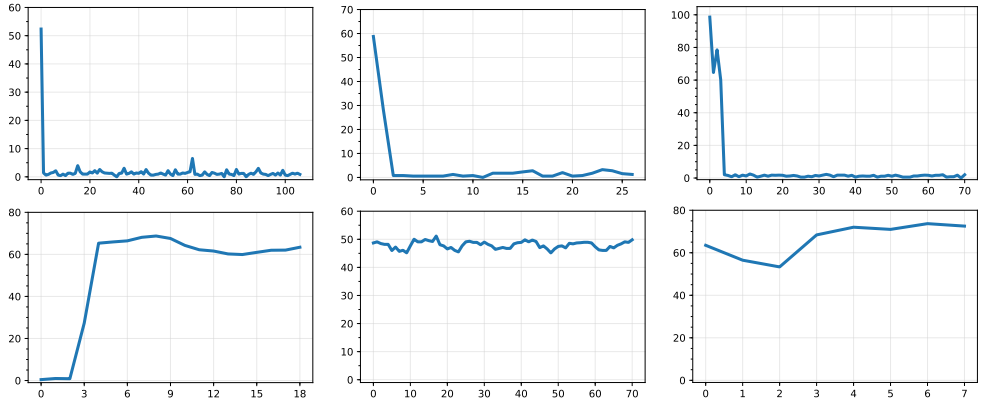


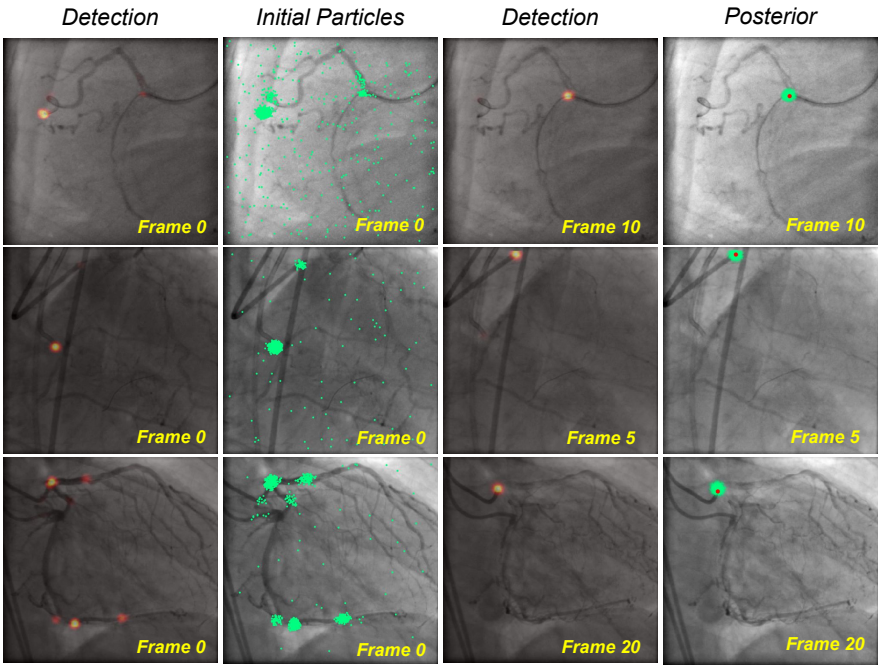
Figure 6.11: Longitudinal views of tracking errors (mm) for the 6 sequences with outliers using automatic initialization.

catheters, guidewires or residual of contrast agent in the images, whereas the other methods improve the alignment and show good match between the structures in the original X-ray image and the roadmaps. Compared to the DCR with manual tip tracking, the proposed method show similar visual alignment of the roadmaps to the original X-ray images.

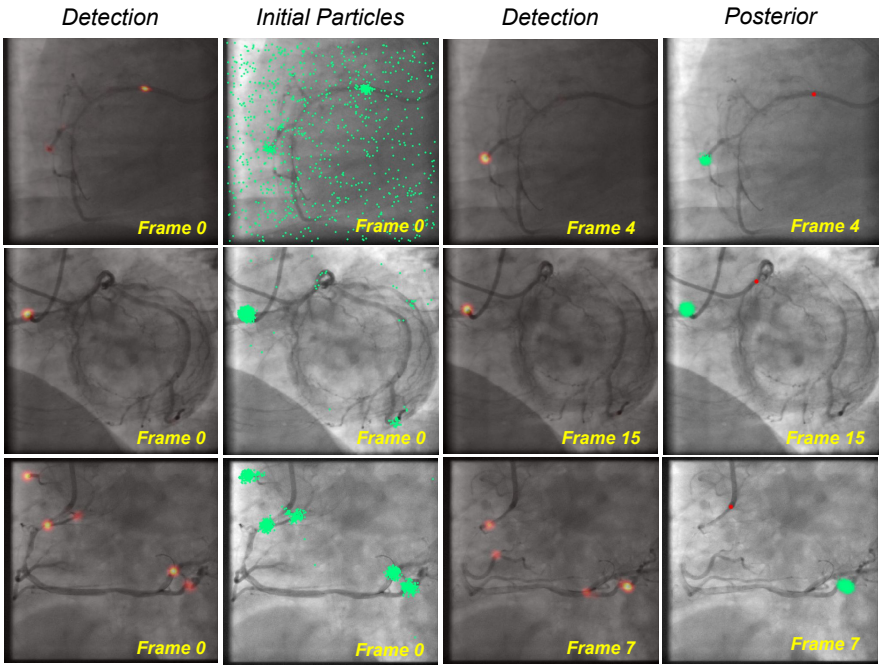
6.6.4 Processing Time

The processing time of all steps in the proposed DCR method was measured with the hardware and software setup in Section 6.5.5. The ECG matching method for roadmap selection was running in Python on the CPU of the linux machine; the deep neural network and the optical flow component of the tracking method were running on the GPU.

In the experiments, the runtimes for roadmap selection (step 1) and roadmap transformation (step 3) in Fig. 6.1 were negligible ($<$ 1 ms / frame). The runtime of the proposed catheter tip tracking method is shown in Table 6.9 and Fig. 6.15. The average time to compute the likelihood with the deep learning setup (DL) is 31.5 ms / frame. The particle filtering (PF) step, which consists of the optical flow estimation, sample propagation, sample weight update and normalization, prediction and resampling, takes on average 23 ms / frame. Therefore, the average tracking time in total is 54.5 ms / frame. The total average time of the proposed DCR including roadmap selection, catheter tip tracking and roadmap transformation is still less than the acquisition time of our data (66.7 ms / frame, 15 fps), indicating that the proposed DCR method would run in real-time with our setup.



(a) Sequence 1-3 on the top row in Fig. 6.11



(b) Sequence 4-6 on the bottom row in Fig. 6.11

Figure 6.12: Examples frames from the 6 sequences in Fig. 6.11. The high probability in the detection heatmap is highlighted as bright color. Particles are presented as green dots. The red dots in the last column indicate the ground truth tip location. (Best viewed in color)

Table 6.7: The statistics of DCR accuracy (mm) with three different tracking scenarios.

	Without Tracking	Proposed Tracking Method	Manual Tip Tracking
Maximal distance of all point pairs	27.19	20.24	13.12
Median distance of all point pairs	1.97	1.43	1.35
Mean distance of all point pairs	2.94 ± 2.83	2.07 ± 2.08	1.85 ± 1.72
Median of frame mean distance	2.11	1.42	1.38
Average of frame mean distance	2.76 ± 2.08	1.91 ± 1.52	1.75 ± 1.30

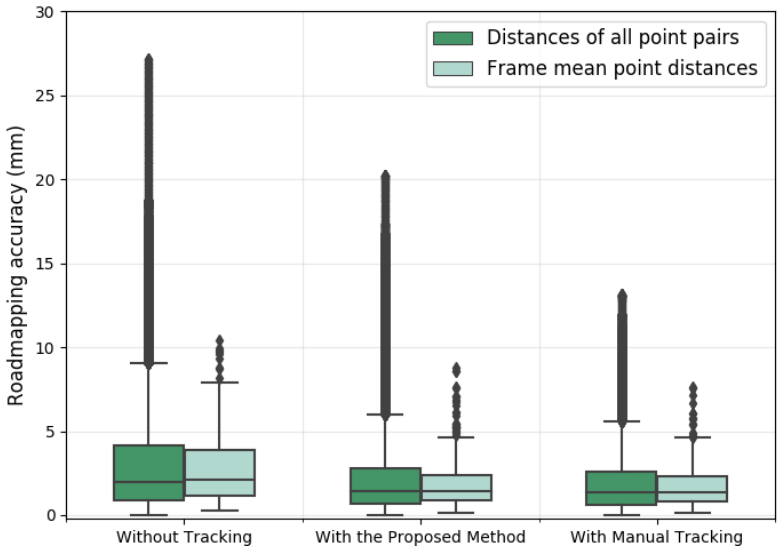


Figure 6.13: Accuracy (mm) of DCR with three different tracking scenarios.

6.7 Discussion

We have presented a new approach to perform online dynamic coronary roadmapping on X-ray fluoroscopic sequences for PCI procedures. The approach compensates the cardiac-induced vessel motion via selecting offline-stored roadmaps with appropriate cardiac phase using ECG matching, and corrects the respiratory motion of vessels by online tracking of guiding catheter tip in X-ray fluoroscopy using a proposed deep learning based Bayesian filtering. The proposed tracking method represents

Table 6.8: Distribution of frame mean point distances of the 409 evaluation frames.

Tracking Methods of DCR	Error Intervals (mm)					
	< 1	1-2	2-3	3-4	4-5	≥ 5
Without tracking	81	115	69	47	31	66
Proposed Tracking Method	131	145	61	32	17	23
Manual Tip Tracking	139	144	61	35	20	10

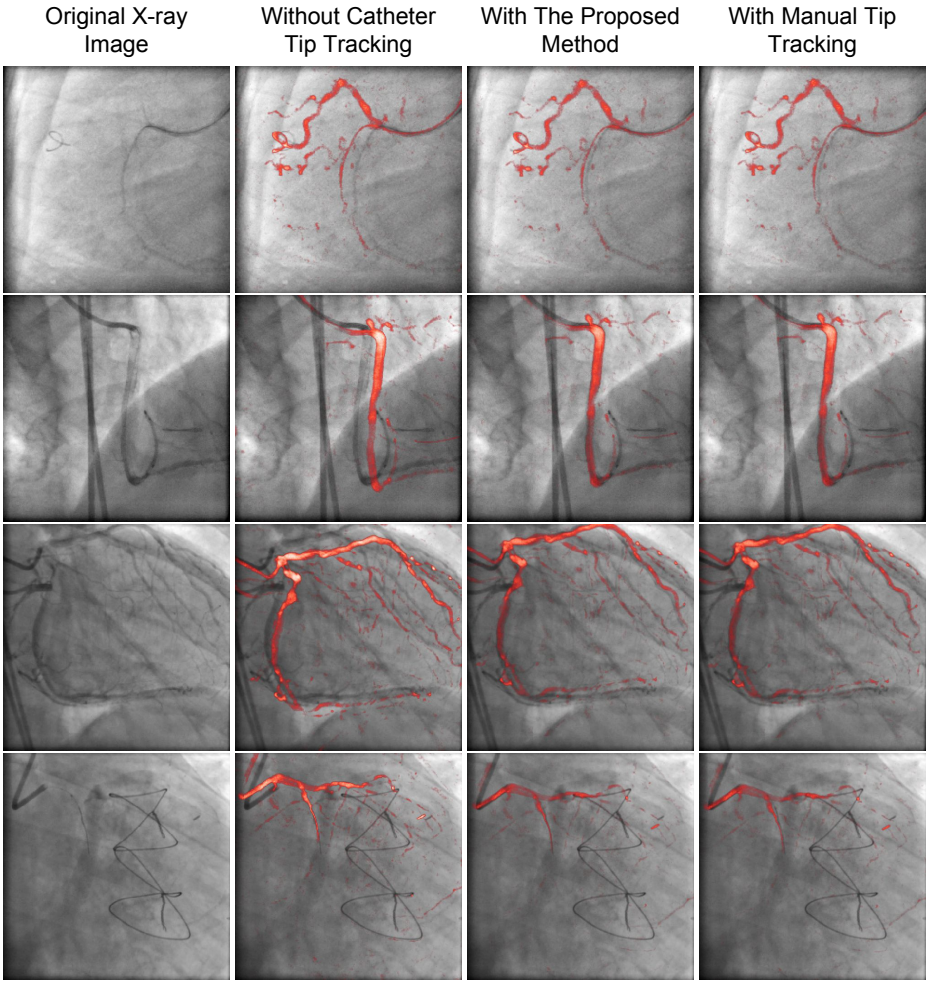


Figure 6.14: Examples of superimposition of selected roadmaps (red) on X-ray fluoroscopic frames. (Best viewed in color)

Table 6.9: Statistics of the runtime of catheter tip tracking (ms / frame) on the test (tracking) dataset.

	Deep Learning	Particle Filtering	Total Tracking Time
Mean	31.5 ± 10.3	23.0 ± 8.7	54.5 ± 12.3
Median	35.1	22.8	57.7

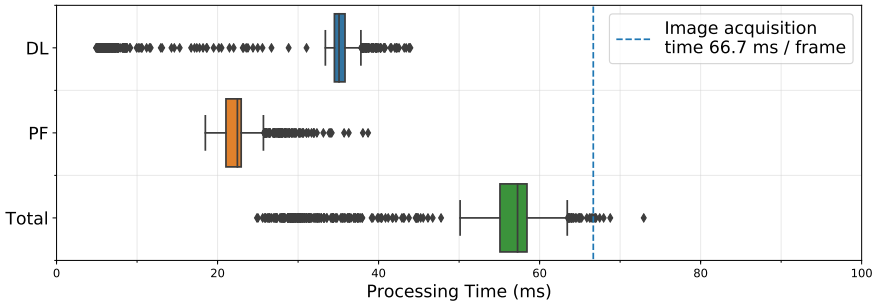


Figure 6.15: Runtime of catheter tip tracking (ms / frame) on all test frames.

and tracks the posterior of catheter tip via a particle filter, for which a likelihood probability map is computed for updating the particle weights using a convolutional neural network. In the experiments, the proposed DCR approach has been trained and evaluated on clinical X-ray sequences for both tracking and roadmapping tasks.

One prerequisite of accurate tracking with the proposed approach is to obtain a reasonably good likelihood estimation, which requires to train the deep neural network to detect catheter tip well. In this work, we have investigated the influence of three network hyperparameters on the performance of the detection network (Section 6.6.1): the basic channel number and network depth level are model capacity parameters, the dropout adds regularization to the model. The experiment showed that the detection accuracy improves when the basic channel number and the network depth level increase (Table 6.3). This observation matches the expectation that a more complex model has higher capacity to model the variation in the data, hence results in better accuracy. When the complexity reaches a certain level, e.g. 64 basic channels and 5 level of depth, the network performance does not increase much compared to those with simpler settings, implying that the model starts overfitting on our dataset.

In addition to the deep neural network, the other important component of the proposed tracking approach is the sampling in the particle filter that yields the samples for representing the prior and the posterior of catheter tip position. First, a sufficient number of samples in the whole sample space are required to well characterize the probability distributions. The experiment in Section 6.6.2.2 shows that 100 samples are suboptimal, while 1000 samples seem sufficient, as 10000 samples result in tracking accuracies similar to 1000 samples. Second, the sample dynamics also plays an important role in tracking, in particular, as indicated by Eq.(6.8), the process noise and the sample motion. The process noise has an impact on the tracking accuracy,

according to Table 6.4. It shows that the optimal choices of the standard deviation of the process noise are 4 or 5 px for the downsampled images. One possible reason for such choices may be that they are similar to the size of guiding catheters. Apart from the process noise, sample motion is another key aspect of sample dynamics. Motion estimation has previouly been incorporated in a motion-based particle filter, such as adaptive block matching [19]. In our work, optical flow was chosen for motion estimation, as its non-parametric nature allows to characterize the complexity of motion in X-ray fluoroscopy well. In addition, the advantage of such approach from a theoretical point of view is that it takes into account of the current observation, leading to a more optimal importance density [11] compared to random motion.

The tracking results in Section 6.6.2.3 show that the proposed tracking approach is able to track the catheter tip in X-ray fluoroscopy accurately with an average tracking error of about 1.3 mm. It also shows advantages over methods based only on optical flow or the detection network. The OF (pre) method relies heavily on tracking in the previous frame, hence the error could accumulate. The OF (first) method may suffer from large motion from the first frame to the current frame. The detection method uses information only from the current frame, no temporal relation between frames is utilized; therefore, it results in spikes in the longitudinal view, as shown in Fig. 6.7. The proposed tracking method has a CNN to provide an accurate observation on the current frame which improves the accuracy of optical flow tracking within the framework of Bayesian filtering. In the meantime, the optical flow based particle filter maintains and propagates the prior knowledge from the initial tip position to provide a constraint on searching for the potentially correct positions, which is useful especially when the CNN detector fails to find the correct target area. The association of knowledge from two sources together improves the tracking accuracy compared to each single source.

The initial state is a also key component of tracking approaches. In the context of Bayesian filtering, the initial state provides the prior knowledge of the tracking target. Most tracking algorithms assume a known initial state from which the target is tracked, e.g. our proposed method with manual initialization in Section 6.6.2.4. In this case, the prior knowledge is provided by human. In Section 6.6.2.4, we also investigated a scenario where the initial state is given by the detection network, so that the complete tracking process is fully automated. The results indicate that, the proposed tracking method with automatic initialization works reasonably well on most sequences even when the initialization is sometimes incorrect (Fig. 6.12a). This is because (1) the true position is covered by a few samples, and (2) the correct detection in later frames corrects the initial mistake in the first frame. The automatic initialization fails when (1) a wrong position is covered by a few samples and (2) the wrong detection in subsequent frames confirms the mistake in the initial frame (Fig. 6.12b). This happens when there is contrast agent remaining in the image or there are multiple catheters, which are the major sources causing ambiguity in detection. In practice, the automatic initialization would work well when contrast agent is washed out and only one catheter is present in the field of view, otherwise manual initialization would be needed which requires only one click to initiate tracking.

Dynamic coronary roadmapping is the direct application of the catheter tip tracking results. In our experiments, the DCR was performed with manual tip initialization

to show the potential of the proposed tracking method, and was compared with the DCR without tracking and with manual tracking. The results indicate that using catheter tip tracking can improve DCR accuracy, as the respiratory-induced vessel motion is corrected by the displacement of catheter tip in addtion to cardiac motion correction. The results also show that the proposed DCR reaches a good accuracy (mean error is about 2 mm) and performs only slightly worse than its best case, the DCR with manual tip tracking which is not applicable for intraoperative use. Additionally, according to a previous study [30], the average lumen diameters of human coronary arteries are between 1.9 mm (distal left anterior descending artery) and 4.5 mm (left main artery). This means that the accuracy achieved with the proposed approach is comparable with the size of coronary arteries.

Apart from catheter tip tracking, several other possible factors in different steps of the experiments may influence the final DCR accuracy. First, in the offline phase, the signal of contrast agent may become too strong and completely cover the catheter tip, complicating the tip visibility in some cases. In this situation, the uncertainty in the manual tip annotation may result in errors in roadmap transformation. Second, in the roadmap selection step, the offline-stored roadmaps are only discrete samples of complete cardiac cycles which might not fully characterize every possible change in the cardiac motion. This problem could possibly be addressed in the future by interpolating frames between the existing frames in the data. Additionally, variation exists between different cardiac cycles [72], therefore, choosing a roadmap from another cycle may cause inaccuracy for cardiac motion compensation. Finally, the way of DCR evaluation in Section 6.5.4 might also introduce inaccuracies in the error measurement. Since guidewires often attach to the inner wall of vessels, the small difference between the annotation of guidewire and vessel centerlines was ignored in the evaluation.

In addition to accuracy, processing speed is also critical for intraoperative applications. The results in Section 6.6.4 indicate that the total processing time of the proposed DCR approach is less than the image acquisition time on our setup, meaning that it runs in real-time. It is worth noticing that the DL and PF steps of the proposed tracking method are independent from each other. In practice, in case more than one GPU are available, the proposed DCR approach can be further accelerated by paralleling the DL and PF steps, making them running on different GPUs.

Compared to the previous works on DCR, the proposed approach in this paper shows advancement in several aspects. First, our systems works on non-cardiac-gated sequences which does not require additional setups for cardiac motion gating that were needed for some methods [71, 119]. Second, our approach compensates both respiratory- and cardiac-induced vessel motion, which is more accurate than systems that correct only cardiac motion [32]. In addition, the proposed DCR approach follows a data-driven paradigm that learns target feature from sequences acquired from different patients and various view angles, making it more robust than the method that relies on traditional vesselness filtering [55] or methods that require specific tissue being present [71, 119]. These are the major advantages of the proposed DCR over the existing direct roadmapping systems. Compared to model-based motion compensation, our approach does not require the extraction of motion surrogate signals and train a motion model for each new patient, but can be directly run with a trained

model.

The proposed deep learning based Bayesian filtering method also presents advantages over the existing instrument tracking approaches. First, the deep learning component enables a more general framework to detect instruments in medical images than methods tailored for specific tools [68, 69]. Compared to the existing detection methods based on deep learning [17, 31, 59], our approach takes into account of the information between frames; the Bayesian filtering framework allows interaction between temporal information and the detection of a convolutional neural network, making the tracking more robust. Bayesian frameworks have been used in many previous temporal instrument tracking methods. Particularly, in some works, the likelihood term was designed based on registration or segmentation outcomes [9, 99] or traditional machine learning approaches with handcrafted features [79, 110, 111]. In our method, we approximated the likelihood with a deep neural network learned from the clinical data which exempts the need of feature engineering but yet possesses more discriminative power; the network directly produces the probability map, making it more straightforward to use. Finally, compared to the existing instrument tracking approaches based on Bayesian filtering [9, 99, 100], the state transition in our method was based on the motion estimated between two adjacent frames, which is more reliable than totally random motion or artificially-designed state transition models.

From a practical point of view, the proposed DCR approach could easily fit into the clinical workflow of PCI. The offline phase of the method can be done efficiently by a technical assistant of the procedures: selecting and creating roadmaps from an angiography acquisition, annotating the catheter tip (one point) in the images. This phase is typically done before a fluoroscopy acquisition during which the guidewire advancement and stent placement are performed. In the online phase, when a fluoroscopic image is acquired, the proposed system selects the most suitable roadmap, tracks the catheter tip and transforms the roadmap to prospectively show a vessel overlay on the fluorosocpic image. The online updated coronary roadmap can provide real-time visual guidance to cardiologists to manipulate interventional tools during the procedure without the need of administering extra contrast agent.

In the future, it may be worth investigating the following directions related to this work. As the data used in this study was acquired from one hospital using a machine from a single vendor, it would be interesting to evaluate the proposed approach on multi-center data acquired with machines from different vendors. It would be also of interest to validate our approach during PCI procedures in an environment simulating the real clinical settings. Additionally, from a methodological point of view, although the proposed tracking method is invariant under different view angles, the whole DCR approach works only when the offline and online phase have the same view angle, i.e. it is a 2D roadmapping system. Therefore, one future direction would be to develop a 3D DCR system that would work with various view angles in the online phase.

6.8 Conclusion

We have developed and validated a novel approach to perform dynamic coronary roadmapping for PCI image guidance. The approach compensates cardiac motion through ECG alignment and respiratory motion by guiding catheter tip tracking during fluoroscopy with a deep learning based Bayesian filtering method. The proposed tracking and roadmapping approaches were trained and evaluated on clinical X-ray image datasets and were proved to perform accurately on both catheter tip tracking and dynamic coronary roadmapping tasks. Our approach also runs in real-time on a setup with a modern GPU and has the potential to be easily integrated into routine PCI procedures, assisting the operator with real-time visual image guidance.

Acknowledgement This work was supported by NWO-domein TTW (The Division of Applied and Engineering Sciences in The Netherlands Organisation for Scientific Research), IMAGIC project under the iMIT program (grant number 12703). Ries and Simon van Walsum are acknowledged for their contribution in the manual annotations.

Chapter Seven

Summary And Future Perspectives

7.1 Summary

In this thesis we have reported on the development and evaluation of dynamic image analysis approaches towards improving image guidance for percutaneous coronary interventions. We have proposed layer separation techniques for retrospective and prospective vessel enhancement in X-ray angiograms. In addition, we have demonstrated two applications that benefit from layer separation for extraction of respiratory motion surrogate and detection of contrast inflow in XA sequences. Finally, we have reported the development of a dynamic coronary roadmapping approach for providing real-time visual guidance on X-ray fluoroscopic images for reduced use of contrast agent during PCI procedures.

Automatic XA image processing, such as extraction of coronary arteries, may serve as a basic component for further image analysis of image guidance applications. Common vessel extraction approaches, e.g. Hessian-based vesselness filtering, are often hampered by low-contrast and background structures in XA images. In **Chapter 2**, we developed a layer separation approach for vessel enhancement to separate XA images into three layers, a breathing layer, a quasi-static layer and a vessel layer that contains coronary arteries. The method firstly separates the breathing layer using a morphological closing; the quasi-static layer and the vessel layer are then separated using robust PCA based on low-rank and sparsity assumptions. The layer separation approach was evaluated on four XA image series acquired from four patients with visual assessment and quantitative evaluation using contrast-to-noise ratio as the metric. The resulted vessel layer has shown a substantial improvement on vessel visibility compared to the original XA. The potential application of the proposed method is that it could possibly be used as a basic component for further XA image analysis.

The approach presented in Chapter 2 requires a number of frames as input to obtain reliable layer separation. It is suitable to be used as a post-processing component, but the nature of batch processing prevents its online use during interventions. In **Chapter 3**, we developed an automatic online layer separation approach that robustly separates XA images into the same kind of three layers as Chapter 2. The breathing layer is firstly separated using morphological closing; the quasi-static layer and the vessel layer are then separated with an online robust PCA (OR-PCA) algorithm. Different from the "off-line" layer separation, the online method is able to take as input a single XA image of a sequence at a time, and separates the three layers

for that image. Upon processing the new XA image, the online method also updates an internal background model using the information of the new image. This setup is a major advantage over the "off-line" approach since it makes online processing of streaming XA data possible. In addition, as the original OR-PCA treats all frames equally, we have also investigated three different ways to downweight the past frames to improve the layer separation performance. The proposed online layer separation methods run fast and were demonstrated to significantly improve the vessel visibility in 42 clinical XA sequences and showed better performance than other related online or prospective approaches. Moreover, the potential of the proposed approach was demonstrated by enhancing contrast of vessels in X-ray images with low vessel contrast using synthetic data and real XA images from four pigs. This potentially enables using a reduced amount of contrast agent to prevent contrast-induced side effects.

The layer separation approaches described in Chapter 2 and 3 enable independent analysis of the resulted layers, and can, hence, become a component in the image processing pipeline for various applications. In **Chapter 4**, we presented a method on extraction of respiratory motion surrogate signals from XA sequences. The method starts with a layer separation step to obtain the breathing layer of an XA sequence. Then principal component analysis (PCA) is applied on the pixel intensities of the breathing layer. Finally the breathing layer is projected onto the first principal component to obtain the respiratory motion surrogate. The surrogate signal was demonstrated to be highly correlated with the ground truth diaphragm displacement (average correlation coefficient over 0.9) on monoplane and biplane data from eight patients in both retrospective and prospective analysis. Although no statistically significantly difference was observed between the proposed method and other baseline approaches, our method improves the robustness and runs faster. One application direction of the proposed respiratory motion surrogate signal is that it could potentially be used as a component for patient-specific coronary motion modeling that is useful for motion compensation in image guided procedures [72].

Another application that benefits from layer separation is detection of contrast inflow in XA sequences. Automatic detection of contrast inflow in XA can assist image guidance of PCI procedures. Tasks such as registration of coronary models from CTA to XA can only be accomplished when vessels are visible in X-ray images. Likewise, automatic detection of interventional tools for dynamic roadmapping is only feasible in non-contrast X-ray images. Therefore, an automatic way to detect contrast inflow can facilitate automating the image guidance tasks, reducing the interaction of the operator and computers during the procedure. In **Chapter 5**, we developed two approaches to detect contrast inflow in XA online. More specifically, the two approaches are machine (deep) learning based classifiers to detect whether an X-ray image is with contrast agent. One method takes the advantage of layer separation, using the vessel layer followed by a Hessian-based vesselness filtering to enhance vessel structures in XA images. An LSTM classifier then takes as input the average intensity of the vessel-enhanced image to detect the contrast in that image. The second approach uses a convolutional neural network (CNN) for image classification. The network takes as input the current frame, its three previous frames, and the first frame of the sequence where no contrast agent is present. Through several

residual units with convolutional layers followed by two fully-connect layers, the network outputs the probability of the current frame being contrasted. Both methods have excellent performance on detection of the beginning contrast frame with average errors less than 4 frames and median errors about 2 frames (given the frame rate 15 fps) on 80 clinical X-ray sequences, and outperform the state-of-the-art with large margin. In terms of speed, the proposed two methods run fast, and have the potential to be integrated in advanced image guidance systems for PCI.

While contrast agents are often administered to visualize coronary arteries and assess lesions during PCI, interventional tools are typically navigated with non-contrast-enhanced X-ray images, as higher use of contrast agent increases the risk of kidney failure. When guided with only fluoroscopic X-ray images without contrast agent, the cardiologist needs to mentally reconstruct the anatomy of vessels and lesions. In **Chapter 6**, we developed a dynamic coronary roadmapping approach to address the challenge of reduced contrast use and limited visual feedback during instrument navigation in the procedure. With this approach, fluoroscopic images are augmented with a dynamic vessel layer obtained from a stored angiographic sequence acquired from the same C-arm angle. The cardiac and respiratory motion of the vessel layer are compensated by ECG alignment and catheter tip tracking, respectively. To accurately track the guiding catheter tip, we proposed a deep learning based Bayesian filtering method which unifies the probabilistic detection outcome of a convolutional neural network and the motion estimation between frames into a particle filtering framework. The proposed tracking method achieved a tracking accuracy with an average error of 1.3 mm on 34 clinical X-ray sequences and has been shown superior to detection without temporal information using the CNN and tracking with only the motion estimation using optical flow. The roadmapping with the proposed tracking algorithm achieved an average error about 2 mm on 409 frames with guidewire annotations from the 34 sequences, and is close to the accuracy (1.8 mm) of the roadmapping with manual tip tracking. In terms of computation efficiency, the proposed dynamic roadmapping approach including the catheter tip tracking step runs in real-time on a workstation with a single GPU. These experiment results suggest that the proposed dynamic coronary roadmapping approach has the potential to be integrated into routine PCI procedures, providing cardiologists with real-time visual guidance during interventions without the need of extra contrast use.

A Summary on Dynamic Analysis in This Thesis

The methodologies for dynamic X-ray angiography or fluoroscopy analysis developed in this thesis from Chapter 2 to Chapter 6 have taken advantage of the temporal, motion or multi-frame information in X-ray sequences. Depending on how the information was processed, the approaches could be categorized into two approaches: *batch processing* and *online processing*.

Batch processing takes as input more than one images or the complete sequences as a whole group for information processing, such that the temporal information is implicitly contained in the input. The approaches in Chapter 2, Chapter 4 and the CNN-based method in Chapter 5 fall into this category. The robust PCA algorithm

in Chapter 2 and PCA in Chapter 4 are unsupervised learning algorithms to discover latent structures from multiple images; the CNN-based method in Chapter 5 is a supervised deep learning approach. For example, robust PCA models the background as a low-rank component and the foreground as a sparse component from a sequence of images; PCA extract the first principle component as the dimension that well explains the respiratory motion in multiple images. These tasks can not be simply achieved with information from only one single image. The CNN-based classification in Chapter 5 may benefit from frames at multiple time points as additional input channels to provide temporal changes. In terms of image guidance applications, the batch processing approaches are well suitable as a post-processing component in retrospective image analysis pipelines. In order for prospective uses, this type of models may be trained offline in advance; during the online phase, only inference is applied with the trained model, e.g. the prospective analysis in Chapter 4 and the CNN-based approach in Chapter 5.

Online processing takes as input one image at a time, and explicitly model the relation between frames of adjacent time points. This type of models may be represented by the formula $x_t = f(x_{t-1})$, where x_t denotes the information at time t, f is the mapping function from x_{t-1} to x_t. This sequential formulation makes the models ideal for processing streaming data. The approaches in Chapter 3, Chapter 6 and the RNN-based method in Chapter 5 belong to this class. The OR-PCA algorithm in Chapter 3 solves the low-rank decomposition problem in an online manner and is able to update using information of the new coming image. Besides, the three ways to downweight information from the last frame provides variants of the mapping function. The RNN-based method in Chapter 5 is a machine learning approach based on recurrent neural network in which the relation between adjacent frames is modelled by the LSTM units. The Bayesian filtering in Chapter 6 is another type of sequential model in which the temporal relation is represented by the state transition function or the transition probability $p(\mathbf{x}_k|\mathbf{x}_{k-1})$ that involves motion estimation between adjacent frames. In terms of image guidance applications, the online processing methods well fits to the workflow of prospective use when the methods run in real-time, although an off-line model training may be needed for supervised-learning based approaches, such as the RNN-based method in Chapter 5 and the likelihood approximation of the Bayesian filtering in Chapter 6.

7.2 Future Perspectives

In this thesis, we have presented several techniques towards improving image guidance of PCI procedures from various aspects. As the image guidance task for PCI is in general challenging, especially for difficult cases, additional research may still be necessary to improve the accuracy, robustness, and to translate the techniques to clinical practice.

The layer separation approaches developed in Chapter 2 and Chapter 3 have been shown to substantially improve vessel visibility in XA images. In order to further boost the computation speed and obtain a better separation of background structures that may still remain in the vessel layer with the current proposed methods in this

thesis, a deep learning approach may be worth investigating. Preliminary studies [45, 46] (not included in this thesis, see publication list) have been conducted using the vessel layer obtained with the method in Chapter 2 as the ground truth label, and to learn the label with a fully convolutional network. These studies showed promising results. As the learning in these studies is supervised by imperfect labels obtained with another algorithm, it would be interesting to explore new deep learning approaches trained with manual labels or in unsupervised manners in future works.

The proposed layer separation approaches have also been shown useful as a basic image processing component in two applications. The approach proposed in Chapter 4 extracts surrogate signals that are highly correlated to the diaphragm displacement, which is beneficial for coronary motion modelling and exempts from the need of manual diaphragm annotation [96]. In future works, it may be interesting to model the coronary motion using the ECG signals together with the respiratory motion surrogate obtained with the method in Chapter 4, and assess motion prediction for image guidance on X-ray fluoroscopy using the motion model. The approach developed in Chapter 5 for contrast inflow detection assists automating the workflow of image guidance tasks, e.g. knowing when to apply registration of preoperative coronary model from CTA to XA images. The subsequent work in the future may be determining when to start applying dynamic roadmapping on fluoroscopic images, which requires to detect the outflow of contrast agent. In addition, from a methodological point of view, it may be also interesting to develop an unified approach that takes advantages from both the CNN- and RNN-based methods in Chapter 5, which learns features automatically and at the meantime make use of the temporal information between frames.

Image guidance and visualization are often bottlenecks in the feedback loop of operator, interventional tools and patient. The dynamic coronary roadmapping approach developed in Chapter 6 is able to accurately overlay a dynamic vessel layer on top of fluoroscopic images, which provides real-time image feedback to the cardiologist when navigating the instruments. Moreover, from a patient-specific motion modelling point of view, the catheter tip position tracked with the proposed algorithm may also serve as a surrogate signal that could be used in a motion model of coronary arteries.

Following the current research on dynamic roadmapping towards building an image guidance system for PCI and translating it to clinical practice, it might be worth investigating the following directions.

- Since the roadmap used in the study in Chapter 6 was retrospectively taken from the same sequence with the fluoroscopic images, it may be necessary to measure how accurate the dynamic roadmapping would be in a real image guidance situation, i.e. the roadmaps are acquired from a different sequence with the same C-arm configuration.

- The current proposed roadmapping approach requires the roadmaps and the fluoroscopic images are acquired with the same C-arm angle, would it be possible to develop a view-angle-independent solution, e.g. a 3D roadmapping system based on reconstructed 3D vessel and catheter models, especially for biplane X-ray angiography systems?

- For difficult clinical cases, such as chronic total occlusion (CTO) featuring as a complete obstruction of the target coronary artery, roadmaps solely based on XA might be limited for visualizing the vessel. For such cases, integration of 3D anatomical information from CTA in fluoroscopic images might be the solution to visualize the occluded vessel segment. The dynamic coronary roadmapping approach enables such an integration which was not possible with the previous 3D/2D registration approaches that need contrast agent to be present in the X-ray images.

- It should be worth exploring the possibility of building one system that combines the components developed in this thesis, i.e. the layer separation, contrast frame detection and dynamic coronary roadmapping, for an accurate, fully-automated and real-time image guidance solution. The integrated system should be optimized for the clinical workflow of PCI in a catheterization room.

- Clinical validations are needed, e.g. using multi-center data acquired from machines of different vendors, testing the system during PCI procedures with real clinical settings.

Based on the approaches developed in this thesis, and the advancement of image guidance techniques and general medical imaging area in the past years, a generalized image guidance platform for interventional support in the future might follow several trends. First, multi-modal data of various dimensions e.g. ECG (1D), XA (2D, 2D+t), CTA (3D, 3D+t), will be integrated for information processing. Second, different types of information will be merged together for better feedback and decision making, e.g. calcification and anatomical information can be introduced from CTA to XA, which is especially useful for difficult cases like chronic total occlusion. These two aspects require techniques of multi-modal registration and information fusion for image guidance, which, e.g. can be supported with the augmented vessel information by the layer separation and the dynamic roadmapping approaches developed in this thesis. Third, data-driven approaches, especially deep learning, enable learning from large datasets, resulting in higher accuracy and robustness in many image guidance tasks. These trends will aim for improved image guidance techniques that are accurate, robust, fast and also safe to patients and physicians.

In conclusion, in this thesis we have developed and evaluated novel dynamic image analysis approaches towards an improved image guidance for percutaneous coronary interventions. These approaches have the potential not only to assist information processing, but also to improve the visual feedback to clinicians and at the meantime reduce the contrast use during interventions. To translate the techniques into clinical practice, further development on integration of the approaches and validations with real interventional settings would be needed.

Bibliography

[1] ", http://www.hopkinsmedicine.org/healthlibrary/test-procedures/cardiovascular/, Accessed: 2014-04-02.

[2] "Anatomy and Function of the Coronary Arteries", https://www.hopkinsmedicine.org/health/conditions-and-diseases/anatomy-and-function-of-the-coronary-arteries, Accessed: 2019-05-23.

[3] "Artis zee - Siemens Healthineers Global", https://www.siemens-healthineers.com/angio/artis-interventional-angiography-systems/artis-zee, Accessed: 2019-06-08.

[4] "Cardiac Catheterization (Heart Cath or Coronary Angiogram)", https://www.sarh.org/Heart-and-Cardiovascular-Surgery/Cardiac-Catheterization/, Accessed: 2019-06-04.

[5] "Coronary artery disease - Diagnosis and treatment - Mayo Clinic", https://www.mayoclinic.org/diseases-conditions/coronary-artery-disease/diagnosis-treatment/drc-20350619, Accessed: 2019-05-31.

[6] "Department of Surgery - Coronary Artery Disease", https://surgery.ucsf.edu/conditions--procedures/coronary-artery-disease.aspx, Accessed: 2019-05-30.

[7] "Your Coronary Arteries, Cleveland Clinic Health Information", https://my.clevelandclinic.org/health/articles/17063-coronary-arteries, Accessed: 2019-05-23.

[8] Pierre Ambrosini, Daniel Ruijters, Wiro J Niessen, Adriaan Moelker, Theo van Walsum, "Fully automatic and real-time catheter segmentation in X-ray fluoroscopy", in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, pp. 577–585, 2017.

[9] Pierre Ambrosini, Ihor Smal, Daniel Ruijters, Wiro J Niessen, Adriaan Moelker, Theo Van Walsum, "A Hidden Markov Model for 3D Catheter Tip Tracking With 2D X-ray Catheterization Sequence and 3D Rotational Angiography.", *IEEE Trans. Med. Imaging*, vol. 36, no. 3, pp. 757–768, 2017.

[10] Michele Andreucci, Richard Solomon, Adis Tasanarong, "Side effects of radiographic contrast media: pathogenesis, risk factors, and prevention", *BioMed Research International*, vol. 2014, 2014.

[11] M Sanjeev Arulampalam, Simon Maskell, Neil Gordon, Tim Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking", *IEEE Transactions on signal processing*, vol. 50, no. 2, pp. 174–188, 2002.

[12] Vincent Auvray, Patrick Bouthemy, Jean Liénard, "Jointmotion estimation and layer segmentation in transparent image sequenc: application to noise reduction in X-ray image sequences", *EURASIP Journal on Advances in Signal Processing*, vol. 2009, p. 19, 2009.

[13] Nora Baka, BPF Lelieveldt, Carl Schultz, W Niessen, Theo van Walsum, "Respiratory motion estimation in X-ray angiography for improved guidance during coronary interventions", *Physics in Medicine & Biology*, vol. 60, no. 9, p. 3617, 2015.

[14] Nora Baka, CT Metz, Carl Schultz, Lisan Neefjes, Robert Jan van Geuns, Boudewijn PF Lelieveldt, Wiro J Niessen, Theo van Walsum, Marleen de Bruijne, "Statistical coronary motion models for 2D+ t/3D registration of X-ray coronary angiography and CTA", *Medical Image Analysis*, vol. 17, no. 6, pp. 698–709, 2013.

[15] Nora Baka, CT Metz, Carl J Schultz, R-J van Geuns, Wiro J Niessen, Theo van Walsum, "Oriented Gaussian mixture models for nonrigid 2D/3D coronary artery registration", *Medical Imaging, IEEE Transactions on*, vol. 33, no. 5, pp. 1023–1034, 2014.

[16] David Barber, *Bayesian reasoning and machine learning*, Cambridge University Press, 2012.

[17] Christoph Baur, Shadi Albarqouni, Stefanie Demirci, Nassir Navab, Pascal Fallavollita, "Cathnets: detection and single-view depth prediction of catheter electrodes", in *International Conference on Medical Imaging and Virtual Reality*, Springer, pp. 38–49, 2016.

[18] Kanwal K Bhatia, Anil Rao, Anthony N Price, Robin Wolz, Joseph V Hajnal, Daniel Rueckert, "Hierarchical manifold learning for regional image analysis", *IEEE transactions on medical imaging*, vol. 33, no. 2, pp. 444–461, 2013.

[19] Nidhal Bouaynaya & Dan Schonfeld, "On the optimality of motion-based particle filtering", *IEEE transactions on circuits and systems for video technology*, vol. 19, no. 7, pp. 1068–1072, 2009.

[20] Thierry Bouwmans & El Hadi Zahzah, "Robust PCA via principal component pursuit: a review for a comparative evaluation in video surveillance", *Computer Vision and Image Understanding*, vol. 122, pp. 22–34, 2014.

[21] Richard Brosig, Sai Gokul Hariharan, Daniele Volpi, Markus Kowarschik, Stephane Carlier, N Navab, S Demirci, "Implicit Background Subtraction for Cardiac Digital Angiography", in *Joint MICCAI workshops on Computing and Visualisation for Intravascular Imaging and Computer-Assisted Stenting*, 2015.

[22] Alexander Brost, Rui Liao, Norbert Strobel, Joachim Hornegger, "Respiratory motion compensation by model-based catheter tracking during EP procedures", *Medical Image Analysis*, vol. 14, no. 5, pp. 695–706, 2010.

[23] Emmanuel J Candès, Xiaodong Li, Yi Ma, John Wright, "Robust principal component analysis?", *Journal of the ACM*, vol. 58, no. 3, p. 11, 2011.

[24] Ping-Lin Chang, Alexander Rolls, Herbert De Praetere, Emmanuel Vander Poorten, Celia V Riga, Colin D Bicknell, Danail Stoyanov, "Robust catheter and guidewire tracking using B-spline tube model and pixel-wise posteriors", *IEEE Robotics and Automation Letters*, vol. 1, no. 1, pp. 303–308, 2016.

[25] Terrence Chen, Gareth Funka-Lea, Dorin Comaniciu, "Robust and fast contrast inflow detection for 2D X-ray fluoroscopy", in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, pp. 243–250, 2011.

[26] Robert A Close, Craig K Abbey, Craig A Morioka, James S Whiting, "Accuracy assessment of layer decomposition using simulated angiographic image sequences", *Medical Imaging, IEEE Transactions on*, vol. 20, no. 10, pp. 990–998, 2001.

[27] Alexandru Condurache, Til Aach, Kai Eck, Joerg Bredno, "Fast detection and processing of arbitrary contrast agent injections in coronary angiography and fluoroscopy", in *Bildverarbeitung fuer die Medizin 2004*, Springer, pp. 5–9, 2004.

[28] Alexandru Condurache, Til Aach, Kai Eck, Joerg Bredno, Thomas Stehle, "Fast and robust diaphragm detection and tracking in cardiac X-ray projection images", in *Medical Imaging 2005: Image Processing*, vol. 5747, International Society for Optics and Photonics, pp. 1766–1775, 2005.

[29] Lisa Dannenberg, Amin Polzin, Roland Bullens, Malte Kelm, Tobias Zeus, "On the road: First-in-man bifurcation percutaneous coronary intervention with the use of a dynamic coronary road map and StentBoost Live imaging system", *International journal of cardiology*, vol. 215, pp. 7–8, 2016.

[30] J Theodore Dodge, B Greg Brown, Edward L Bolson, Harold T Dodge, "Lumen diameter of normal human coronary arteries. Influence of age, sex, anatomic variation, and left ventricular hypertrophy or dilation.", *Circulation*, vol. 86, no. 1, pp. 232–246, 1992.

[31] Xiaofei Du, Thomas Kurmann, Ping-Lin Chang, Maximilian Allan, Sebastien Ourselin, Raphael Sznitman, John D Kelly, Danail Stoyanov, "Articulated multi-instrument 2D pose estimation using fully convolutional networks", *IEEE transactions on medical imaging*, 2018.

[32] Jonathan L Elion, "Dynamic coronary roadmapping", Oct. 31 1989, US Patent 4,878,115.

[33] Anthony Z Faranesh, Peter Kellman, Kanishka Ratnayaka, Robert J Lederman, "Integration of cardiac and respiratory motion into MRI roadmaps fused with X-ray", *Medical physics*, vol. 40, no. 3, 2013.

[34] Gunnar Farnebäck, "Two-frame motion estimation based on polynomial expansion", in *Scandinavian conference on Image analysis*, Springer, pp. 363–370, 2003.

[35] Jiashi Feng, Huan Xu, Shuicheng Yan, "Online robust PCA via stochastic optimization", in *Advances in Neural Information Processing Systems*, pp. 404–412, 2013.

[36] Peter Fischer, Anthony Faranesh, Thomas Pohl, Andreas Maier, Toby Rogers, Kanishka Ratnayaka, Robert Lederman, Joachim Hornegger, "An MR-Based Model for Cardio-Respiratory Motion Compensation of Overlays in X-Ray Fluoroscopy", *IEEE transactions on medical imaging*, vol. 37, no. 1, pp. 47–60, 2018.

[37] Peter Fischer, Thomas Pohl, Thomas Köhler, Andreas Maier, Joachim Hornegger, "A Robust Probabilistic Model for Motion Layer Separation in X-ray Fluoroscopy", in *Information Processing in Medical Imaging*, Springer, pp. 288–299, 2015.

[38] Alejandro F Frangi, Wiro J Niessen, Koen L Vincken, Max A Viergever, "Multiscale vessel enhancement filtering", in *Medical Image Computing and Computer-Assisted Intervention*, Springer, pp. 130–137, 1998.

[39] Luis C García-Peraza-Herrera, Wenqi Li, Caspar Gruijthuijsen, Alain Devreker, George Attilakos, Jan Deprest, Emmanuel Vander Poorten, Danail Stoyanov, Tom Vercauteren, Sébastien Ourselin, "Real-time segmentation of non-rigid surgical tools based on deep learning and tracking", in *International Workshop on Computer-Assisted and Robotic Endoscopy*, Springer, pp. 84–95, 2016.

[40] Jean-Louis Georges, Nicole Karam, Muriel Tafflet, Bernard Livarek, Sophie Bataille, Aurélie Loyeau, Mireille Mapouata, Hakim Benamer, Christophe Caussin, Philippe Garot, others, "Time-Course Reduction in Patient Exposure to Radiation From Coronary Interventional Procedures: The Greater Paris Area Percutaneous Coronary Intervention Registry", *Circulation: Cardiovascular Interventions*, vol. 10, no. 8, p. e005268, 2017.

[41] Hayit Greenspan, Bram Van Ginneken, Ronald M Summers, "Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique", *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1153–1159, 2016.

[42] Andreas R Gruentzig, "Transluminal dilatation of coronary artery stenosis", *Lancet*, vol. 1, p. 263, 1978.

[43] Valentin Hamy, Nikolaos Dikaios, Shonit Punwani, Andrew Melbourne, Arash Latifoltojar, Jesica Makanyanga, Manil Chouhan, Emma Helbren, Alex Menys, Stuart Taylor, others, "Respiratory motion correction in dynamic MRI using robust data decomposition registration–Application to DCE-MRI", *Medical Image Analysis*, vol. 18, no. 2, pp. 301–313, 2014.

[44] Edward L Hannan, Chuntao Wu, Gary Walford, Alfred T Culliford, Jeffrey P Gold, Craig R Smith, Robert SD Higgins, Russell E Carlson, Robert H Jones, "Drug-eluting stents vs. coronary-artery bypass grafting in multivessel coronary disease", *New England Journal of Medicine*, vol. 358, no. 4, pp. 331–341, 2008.

[45] Haidong Hao, Hua Ma, Theo van Walsum, "Layer Separation in X-ray Angiograms for Vessel Enhancement with Fully Convolutional Network", in *Intravascular Imaging and Computer Assisted Stenting and Large-Scale Annotation of Biomedical Data and Expert Label Synthesis*, Springer, pp. 36–44, 2018.

[46] Haidong Hao, Hua Ma, Theo van Walsum, "Vessel layer separation in X-ray angiograms with fully convolutional network", in *Medical Imaging 2018: Image-Guided Procedures, Robotic Interventions, and Modeling*, vol. 10576, International Society for Optics and Photonics, p. 105761V, 2018.

[47] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, "Deep residual learning for image recognition", in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[48] Hauke Heibel, Ben Glocker, Martin Groher, Marcus Pfister, Nassir Navab, "Interventional tool tracking using discrete optimization", *IEEE transactions on medical imaging*, vol. 32, no. 3, pp. 544–555, 2013.

[49] Sepp Hochreiter & Jürgen Schmidhuber, "Long short-term memory", *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[50] Matthias Hoffmann, Simone Müller, Klaus Kurzidim, Norbert Strobel, Joachim Hornegger, "Robust identification of contrasted frames in fluoroscopic images", in *Bildverarbeitung für die Medizin 2015*, Springer, pp. 23–28, 2015.

[51] Nicolas Honnorat, Régis Vaillant, Nikos Paragios, "Graph-based geometric-iconic guide-wire tracking", in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, pp. 9–16, 2011.

[52] Aapo Hyvärinen, Juha Karhunen, Erkki Oja, *Independent component analysis*, vol. 46, John Wiley & Sons, 2004.

[53] Sajid Javed, Seon Ho Oh, Andrews Sobral, Thierry Bouwmans, Soon Ki Jung, "OR-PCA with MRF for robust foreground detection in highly dynamic backgrounds", in *Asian Conference on Computer Vision*, Springer, pp. 284–299, 2014.

[54] Sajid Javed, Andrews Sobral, Thierry Bouwmans, Soon Ki Jung, "OR-PCA with dynamic feature selection for robust background subtraction", in *Proceedings of the 30th Annual ACM Symposium on Applied Computing*, ACM, pp. 86–91, 2015.

[55] Dongkue Kim, Sangsoo Park, Myung Ho Jeong, Jeha Ryu, "Registration of angiographic image on real-time fluoroscopic image for image-guided percutaneous coronary intervention", *International journal of computer assisted radiology and surgery*, vol. 13, no. 2, pp. 203–213, 2018.

[56] Andrew P King, Redha Boubertakh, Kawal S Rhode, YingLiang Ma, Phani Chinchapatnam, Gang Gao, T Tangcharoen, Matthew Ginks, Michael Cooklin, Jaswinder S Gill, others, "A subject-specific technique for respiratory motion correction in image-guided cardiac catheterisation procedures", *Medical Image Analysis*, vol. 13, no. 3, pp. 419–431, 2009.

[57] HA Kirişli, Michiel Schaap, CT Metz, AS Dharampal, WB Meijboom, SL Papadopoulou, A Dedic, K Nieman, MA De Graaf, MFL Meijs, others, "Standardized evaluation framework for evaluating coronary artery stenosis detection, stenosis quantification and lumen segmentation algorithms in computed tomography angiography", *Medical image analysis*, vol. 17, no. 8, pp. 859–876, 2013.

[58] Taisei Kobayashi & John W Hirshfeld Jr, "Radiation exposure in cardiac catheterization: operator behavior matters", 2017.

[59] Iro Laina, Nicola Rieke, Christian Rupprecht, Josué Page Vizcaíno, Abouzar Eslami, Federico Tombari, Nassir Navab, "Concurrent segmentation and localization for tracking of surgical instruments", in *International conference on medical image computing and computer-assisted intervention*, Springer, pp. 664–672, 2017.

[60] Yann LeCun, Yoshua Bengio, Geoffrey Hinton, "Deep learning", *nature*, vol. 521, no. 7553, p. 436, 2015.

[61] Rui Liao, Shun Miao, Yefeng Zheng, "Automatic and efficient contrast-based 2-D/3-D fusion for trans-catheter aortic valve implantation (TAVI)", *Computerized Medical Imaging and Graphics*, vol. 37, no. 2, pp. 150–161, 2013.

[62] Rui Liao, Wei You, Yinxiao Liu, Michelle Yan, Matthias John, Steven Shea, "Integrated spatiotemporal analysis for automatic contrast agent inflow detection on angiography and fluoroscopy during transcatheter aortic valve implantation", *Medical physics*, vol. 40, no. 4, p. 041914, 2013.

[63] Zhouchen Lin, Risheng Liu, Zhixun Su, "Linearized alternating direction method with adaptive penalty for low-rank representation", in *Advances in Neural Information Processing Systems*, pp. 612–620, 2011.

[64] Patrick J. Lynch & Mikael Häggström, "File: Coronary arteries.png - Wikimedia Commons", `https://commons.wikimedia.org/w/index.php?curid=11377867`, 2010, Accessed: 2019-05-21, CC BY-SA 3.0.

[65] Hua Ma, Gerardo Dibildox, Jyotirmoy Banerjee, Wiro Niessen, Carl Schultz, Evelyn Regar, Theo van Walsum, "Layer separation for vessel enhancement in interventional X-ray angiograms using morphological filtering and robust PCA", in *Workshop on Augmented Environments for Computer-Assisted Interventions*, Springer, pp. 104–113, 2015.

[66] Hua Ma, Gerardo Dibildox, Carl Schultz, Evelyn Regar, Theo van Walsum, "PCA-derived respiratory motion surrogates from X-ray angiograms for percutaneous coronary interventions", *International Journal of Computer Assisted Radiology and Surgery*, vol. 10, no. 6, pp. 695–705, 2015.

[67] Hua Ma, Ayla Hoogendoorn, Evelyn Regar, Wiro J Niessen, Theo van Walsum, "Automatic online layer separation for vessel enhancement in X-ray angiograms for percutaneous coronary interventions", *Medical image analysis*, vol. 39, pp. 145–161, 2017.

[68] YingLiang Ma, Nicolas Gogin, Pascal Cathier, R James Housden, Geert Gijsbers, Michael Cooklin, Mark O'Neill, Jaswinder Gill, C Aldo Rinaldi, Reza Razavi, others, "Real-time X-ray fluoroscopy-based catheter detection and tracking for cardiac electrophysiology interventions", *Medical physics*, vol. 40, no. 7, 2013.

[69] YingLiang Ma, Andy P King, Nicolas Gogin, Geert Gijsbers, C Aldo Rinaldi, Jaswinder Gill, Reza Razavi, Kawal S Rhode, "Clinical evaluation of respiratory motion compensation for anatomical roadmap guided cardiac electrophysiology procedures", *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 1, pp. 122–131, 2012.

[70] Julien Mairal, Francis Bach, Jean Ponce, Guillermo Sapiro, "Online learning for matrix factorization and sparse coding", *The Journal of Machine Learning Research*, vol. 11, pp. 19–60, 2010.

[71] Michael Manhart, Ying Zhu, Dime Vitanovski, "Self-assessing image-based respiratory motion compensation for fluoroscopic coronary roadmapping", in *Biomedical Imaging: From Nano to Macro, 2011 IEEE International Symposium on*, IEEE, pp. 1065–1069, 2011.

[72] Jamie R McClelland, David J Hawkes, Tobias Schaeffter, Andrew P King, "Respiratory motion models: a review", *Medical image analysis*, vol. 17, no. 1, pp. 19–42, 2013.

[73] Bernhard Meier, Dölf Bachmann, Thomas F Lüscher, "25 years of coronary angioplasty: almost a fairy tale", *Lancet (London, England)*, vol. 361, no. 9356, pp. 527–527, 2003.

[74] Fausto Milletari, Nassir Navab, Seyed-Ahmad Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation", in *3D Vision (3DV), 2016 Fourth International Conference on*, IEEE, pp. 565–571, 2016.

[75] Ifeanyichukwu Ogobuiro & Faiz Tuma, "Anatomy, Thorax, Heart Coronary Arteries", `https://www.ncbi.nlm.nih.gov/books/NBK534790/`, Accessed: 2019-05-28.

[76] Ricardo Otazo, Emmanuel Candès, Daniel K Sodickson, "Low-rank plus sparse matrix decomposition for accelerated dynamic MRI with separation of background and dynamic components", *Magnetic Resonance in Medicine*, vol. 73, no. 3, pp. 1125–1136, 2015.

[77] M Panayiotou, AP King, Y Ma, RJ Housden, CA Rinaldi, J Gill, M Cooklin, M ONeill, KS Rhode, "A statistical model of catheter motion from interventional X-ray images: application to image-based gating", *Physics in Medicine & Biology*, vol. 58, no. 21, p. 7543, 2013.

[78] Maria Panayiotou, Andrew P King, R James Housden, YingLiang Ma, Michael Cooklin, Mark O'Neill, Jaswinder Gill, C Aldo Rinaldi, Kawal S Rhode, "A statistical method for retrospective cardiac and respiratory motion gating of interventional cardiac X-ray images", *Medical Physics*, vol. 41, no. 7, p. 071901, 2014.

[79] Olivier Pauly, Hauke Heibel, Nassir Navab, "A machine learning approach for deformable guide-wire tracking in fluoroscopic sequences", in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, pp. 343–350, 2010.

[80] Devis Peressutti, Graeme P Penney, R James Housden, Christoph Kolbitsch, Alberto Gomez, Erik-Jan Rijkhorst, Dean C Barratt, Kawal S Rhode, Andrew P King, "A novel Bayesian respiratory motion model to estimate and resolve uncertainty in image-guided cardiac interventions", *Medical image analysis*, vol. 17, no. 4, pp. 488–502, 2013.

[81] Rebecca Perry, Majo X Joseph, Derek P Chew, Philip E Aylward, Carmine G De Pasquale, "Coronary artery wall thickness of the left anterior descending artery using high resolution transthoracic echocardiography–normal range of values", *Echocardiography*, vol. 30, no. 7, pp. 759–764, 2013.

[82] Simeon Petkov, Xavier Carrillo, Petia Radeva, Carlo Gatta, "Diaphragm border detection in coronary X-ray angiographies: new method and applications", *Computerized Medical Imaging and Graphics*, vol. 38, no. 4, pp. 296–305, 2014.

[83] Tomislav Petković & Sven Lončarić, "Guidewire tracking with projected thickness estimation", in *Biomedical Imaging: From Nano to Macro, 2010 IEEE International Symposium on*, IEEE, pp. 1253–1256, 2010.

[84] Kerstin Piayda, Laura Kleinebrecht, Shazia Afzal, Roland Bullens, Iris ter Horst, Amin Polzin, Verena Veulemans, Lisa Dannenberg, Anna Christina Wimmer, Christian Jung, others, "Dynamic coronary roadmapping during percutaneous coronary intervention: a feasibility study", *European journal of medical research*, vol. 23, no. 1, p. 36, 2018.

[85] J Samuel Preston, Caleb Rottman, Arvidas Cheryauka, Larry Anderton, Ross T Whitaker, Sarang C Joshi, "Multi-layer deformation estimation for fluoroscopic imaging.", in *Information Processing in Medical Imaging*, pp. 123–134, 2013.

[86] Lauren Reed-Guy, "What Are the Causes of Coronary Artery Disease?", `https://www.healthline.com/health/coronary-artery-disease/causes`, Accessed: 2019-05-30.

[87] Nicola Rieke, David Joseph Tan, Chiara Amat di San Filippo, Federico Tombari, Mohamed Alsheakhali, Vasileios Belagiannis, Abouzar Eslami, Nassir Navab, "Real-time localization of articulated surgical instruments in retinal microsurgery", *Medical image analysis*, vol. 34, pp. 82–100, 2016.

[88] Hannah Ritchie, "What do people die from?", `https://ourworldindata.org/what-does-the-world-die-from`, Accessed: 2019-05-30.

[89] David Rivest-Henault, Hari Sundar, Mohamed Cheriet, "Nonrigid 2D/3D registration of coronary artery models with live fluoroscopy for guidance of cardiac interventions", *Medical Imaging, IEEE Transactions on*, vol. 31, no. 8, pp. 1557–1572, 2012.

[90] Paul Rodriguez & Brendt Wohlberg, "Fast principal component pursuit via alternating minimization", in *2013 IEEE International Conference on Image Processing*, IEEE, pp. 69–73, 2013.

[91] Olaf Ronneberger, Philipp Fischer, Thomas Brox, "U-net: Convolutional networks for biomedical image segmentation", in *International Conference on Medical image computing and computer-assisted intervention*, Springer, pp. 234–241, 2015.

[92] Daniel Ruijters, Bart M ter Haar Romeny, Paul Suetens, "Vesselness-based 2D–3D registration of the coronary arteries", *International journal of computer assisted radiology and surgery*, vol. 4, no. 4, pp. 391–397, 2009.

[93] Matthias Schneider & Hari Sundar, "Automatic global vessel segmentation and catheter removal using local geometry information and vector field integration", in *2010 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, IEEE, pp. 45–48, 2010.

[94] Matthias Schneider, Hari Sundar, Rui Liao, Joachim Hornegger, Chenyang Xu, "Model-based respiratory motion compensation for image-guided cardiac interventions", in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE, pp. 2948–2954, 2010.

[95] Guy Shechter, Cengizhan Ozturk, Jon R Resar, Elliot R McVeigh, "Respiratory motion of the heart from free breathing coronary angiograms", *IEEE transactions on medical imaging*, vol. 23, no. 8, p. 1046, 2004.

[96] Guy Shechter, Barak Shechter, Jon R Resar, Rafael Beyar, "Prospective motion correction of X-ray images for coronary interventions", *IEEE Transactions on Medical Imaging*, vol. 24, no. 4, pp. 441–450, 2005.

[97] Alok Kumar Singh, "Percutaneous coronary intervention vs coronary artery bypass grafting in the management of chronic stable angina: A critical appraisal", *Journal of cardiovascular disease research*, vol. 1, no. 2, pp. 54–58, 2010.

[98] Wenjie Song, Jianke Zhu, Yang Li, Chun Chen, "Image Alignment by Online Robust PCA via Stochastic Gradient Descent", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. pp (99), 2015.

[99] Stefanie Speidel, Michael Delles, Carsten Gutt, Rüdiger Dillmann, "Tracking of instruments in minimally invasive surgery for surgical skill analysis", in *International Workshop on Medical Imaging and Virtual Reality*, Springer, pp. 148–155, 2006.

[100] Stefanie Speidel, E Kuhn, Sebastian Bodenstedt, Sebastian Röhl, Hannes Kenngott, B Müller-Stich, Rüdiger Dillmann, "Visual tracking of da Vinci instruments for laparoscopic surgery", in *Medical Imaging 2014: Image-Guided Procedures, Robotic Interventions, and Modeling*, vol. 9036, International Society for Optics and Photonics, p. 903608, 2014.

[101] Gábor J Székely, Maria L Rizzo, Nail K Bakirov, others, "Measuring and testing dependence by correlation of distances", *The annals of statistics*, vol. 35, no. 6, pp. 2769–2794, 2007.

[102] Hideyuki Takimura, Toshiya Muramatsu, Reiko Tsukahara, Masatsugu Nakano, Satoshi Nishio, Yukako Takimura, Takayuki Yabe, Mami Kawano, Tasuku Hada, "Usefulness of novel roadmap guided percutaneous coronary intervention for bifurcation lesions (the worlds first cases)", *Journal of the American College of Cardiology*, vol. 71, no. 11, p. A1365, 2018.

[103] Songyuan Tang, Yongtian Wang, Yen-Wei Chen, "Application of ICA to X-ray coronary digital subtraction angiography", *Neurocomputing*, vol. 79, pp. 168–172, 2012.

[104] Shana Tehrani, Chris Laing, Derek M Yellon, Derek J Hausenloy, "Contrast-induced acute kidney injury following PCI", *European journal of clinical investigation*, vol. 43, no. 5, pp. 483–490, 2013.

[105] Martin Tepel, Peter Aspelin, Norbert Lameire, "Contrast-induced nephropathy a clinical and evidence-based approach", *Circulation*, vol. 113, no. 14, pp. 1799–1806, 2006.

[106] Holger Timinger, Sascha Krueger, Klaus Dietmayer, Joern Borgert, "Motion compensated coronary interventional navigation by means of diaphragm tracking and elastic motion models", *Physics in Medicine & Biology*, vol. 50, no. 3, p. 491, 2005.

[107] Theo van Walsum, Michiel Schaap, Coert T Metz, Alina G van der Giessen, Wiro J Niessen, "Averaging centerlines: mean shift on paths", in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, pp. 900–907, 2008.

[108] Alessandro Vandini, Ben Glocker, Mohamad Hamady, Guang-Zhong Yang, "Robust guidewire tracking under large deformations combining segment-like features (SEGlets)", *Medical image analysis*, vol. 38, pp. 150–164, 2017.

[109] Daniele Volpi, Mhd H Sarhan, Reza Ghotbi, Nassir Navab, Diana Mateus, Stefanie Demirci, "Online tracking of interventional devices for endovascular aortic repair", *International Journal of Computer Assisted Radiology and Surgery*, vol. 10, no. 6, pp. 773–781, 2015.

[110] Peng Wang, Terrence Chen, Ying Zhu, Wei Zhang, S Kevin Zhou, Dorin Comaniciu, "Robust guidewire tracking in fluoroscopy", in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, pp. 691–698, 2009.

[111] Wen Wu, Terrence Chen, Norbert Strobel, Dorin Comaniciu, "Fast tracking of catheters in 2D fluoroscopic images using an integrated CPU-GPU framework", in *Biomedical Imaging (ISBI), 2012 9th IEEE International Symposium on*, IEEE, pp. 1184–1187, 2012.

[112] Xianliang Wu, James Housden, YingLiang Ma, Benjamin Razavi, Kawal Rhode, Daniel Rueckert, "Fast catheter segmentation from echocardiographic sequences based on segmentation from corresponding X-ray fluoroscopy for cardiac catheterization interventions", *IEEE transactions on medical imaging*, vol. 34, no. 4, pp. 861–876, 2015.

[113] Takayuki Yabe, Toshiya Muramatsu, Reiko Tsukahara, Masatsugu Nakano, Yukako Takimura, Hideyuki Takimura, Mami Kawano, Tasuku Hada, "The impact of percutaneous coronary intervention using the novel dynamic coronary roadmap system", *Journal of the American College of Cardiology*, vol. 71, no. 11, p. A1103, 2018.

[114] Liron Yatziv, Mathieu Chartouni, Saurabh Datta, Guillermo Sapiro, "Toward multiple catheters detection in fluoroscopic image guided interventions", *IEEE Transactions on Information Technology in Biomedicine*, vol. 16, no. 4, pp. 770–781, 2012.

[115] Wei You, Rui Liao, Michelle Yan, Matthias John, "Spatio-temporal analysis for automatic contrast injection detection on angiography during trans-catheter aortic valve implantation", in *2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, IEEE, pp. 702–706, 2011.

[116] Wei Zhang, Haibin Ling, Simone Prummer, Kevin Shaohua Zhou, Martin Ostermeier, Dorin Comaniciu, "Coronary tree extraction using motion layer separation", in *Medical Image Computing and Computer-Assisted Intervention*, Springer, pp. 116–123, 2009.

[117] Xuan Zhao, Shun Miao, Liang Du, Rui Liao, "Robust 2-D/3-D registration of CT volumes with contrast-enhanced X-ray sequences in electrophysiology based on a weighted similarity measure and sequential subspace optimization", in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, IEEE, pp. 934–938, 2013.

[118] Ying Zhu, Simone Prummer, Peng Wang, Terrence Chen, Dorin Comaniciu, Martin Ostermeier, "Dynamic layer separation for coronary DSA and enhancement in fluoroscopic sequences", *Medical Image Computing and Computer-Assisted Intervention*, pp. 877–884, 2009.

[119] Ying Zhu, Yanghai Tsin, Hari Sundar, Frank Sauer, "Image-based respiratory motion compensation for fluoroscopic coronary roadmapping", in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, pp. 287–294, 2010.

Samenvatting

Samenvatting

In dit proefschrift beschrijven we de ontwikkeling en evaluatie van methoden voor de analyse van dynamische Rntgen beeldreeksen om de beeldgeleiding bij percutane ingrepen aan de kransslagader te verbeteren. We hebben een methode ontwikkeld om de verschillende 'lagen' in een Röntgen afbeelding van de vaten te onderscheiden, om zowel retrospectief als prospectief de vaten in zulke afbeeldingen beter zichtbaar te maken. Daarnaast laten we twee andere toepassingen zien die gebruik kunnen maken van deze techniek: de extractie van ademhalingsinformatie uit beeldreeksen, en het bepalen van het tijdspunt waarop het contrastmiddel het beeld binnenstroomt in angiografische beeldenreeksen. Tenslotte beschrijven we de ontwikkeling van een methode die de kransslagaders kan projecteren op doorlichtingsbeelden die gemaakt worden tijdens interventies aan de kransslagaders, om zo de beeldgeleiding te verbeteren.

Automatische analyse van angiografische beeldreeksen, zoals de extractie van de bloedvaten, is een basis voor verdere analyse en verbetering van beeldgeleiding. Gebruikelijke technieken voor vaatanalyse, zoals die gebaseerd op de Hessiaan van het beeld, werken vaak niet goed bij laag contrast, en worden ook gehinderd door de aanwezigheid van andere structuren. In Hoofdstuk 2 ontwikkelden we een manier om op basis van de beweging onderscheid te maken tussen verschillende structuren in beeldreeksen, nl. statische structuren, structuren die met de ademhaling mee bewegen, en structuren die met het hart bewegen. Deze methode onderscheidt eerst de ademhalingsstructuren door een morfologische operatie, waarna vervolgens statische en bewegende structuren onderscheiden worden via de toepassing van een robuuste PCA. Deze methode is zowel visueel als kwantitatief geëvalueerd op vier beeldreeksen van vier patiënten. Het beeld met de geëxtraheerde vaatstructuren geeft een substantieel betere visualisatie van de vaten, in vergelijking met de originele beelden.

De methode uit Hoofdstuk 2 gebruikt de volledige beeldreeks om de verschillende structuren te onderscheiden, en kan dus alleen als nabewerking worden toegepast, en niet tijdens een daadwerkelijke ingreep. In Hoofdstuk 3 wordt de methode daarom verder ontwikkeld, zodat deze ook in een 'on-line' scenario, tijdens een ingreep, kan worden toegepast. De structuren die bewegen met de ademhaling worden nog steeds via een morfologisch filter bepaald, en de vaten worden dan van de statische structuren onderscheiden door een real-time versie van de robuuste PCA methode. Tijdens deze bewerking wordt continu een intern model van de statische achtergrond bepaald en

aangepast. Deze methode heeft als voordeel dat deze ook tijdens ingrepen gebruikt kan worden. Daarnaast geeft de methode, vanwege de mathematische formulering, de mogelijkheid om de bijdrage van vorige beelden aan te passen (minder belangrijk te maken). Verschillende methoden om dit te bereiken zijn hiervoor geëvalueerd. Tijdens testen op 42 beeldreeksen bleek dat de methode voldoende snel werkt, en de zichtbaarheid van vaten in de beelden kan vergroten. Daarnaast is in vier beeldreeksen van ingrepen bij varkens aangetoond dat de methode daadwerkelijk gebruikt kan worden om het contrast te verbeteren in beelden met vaten met weinig contrastmiddel. De methode heeft dus de potentie om het contrastmiddel gebruik tijdens ingrepen te verminderen.

De methode uit de hoofdstukken 2 en 3 maakt een aparte analyse van de verschillende structuren in het beeld mogelijk, en kan dus gebruikt worden als component in meer geavanceerde methoden. In Hoofdstuk 4 laten we een methode zien die een surrogaat ademhalingssignaal uit een beeldreeks kan halen. De methode begint met het bepalen van de laag met ademhalingsstructuren. Vervolgens wordt een PCA methode toegepast op de intensiteiten in de resulterende beelden. Vervolgens worden nieuwe beelden geprojecteerd op de eerste component van de PCA analyse om het ademhalingssignaal te bepalen. In beeldreeksen van 8 patiënten hebben we, zowel retrospectief als prospectief, laten zien dat dit signaal sterk gecorreleerd is met de positie van het diafragma. Hoewel er geen statistisch significant verschil was tussen de voorgestelde methode en eerder gebruikte methoden, is de voorgestelde methode robuuster, en ook sneller. Dit ademhalingssignaal zou bijvoorbeeld gebruikt kunnen worden voor bewegingscorrectie bij beeldgeleide ingrepen.

Een andere toepassing voor het onderscheiden van de verschillende structuren is het detecteren van de instroom van contrastmiddel. Het automatisch detecteren van contrastmiddelinstroom in een beeldreeks kan van belang zijn voor het verbeteren van beeldgeleiding. Sommige taken, zoals de registratie van een kransslagadermodel op basis van CTA, kunnen alleen gedaan worden als er contrastmiddel in de vaten aanwezig is. Andere taken, zoals de detectie van instrumenten in het beeld, zijn alleen goed mogelijk als er geen contrastmiddel aanwezig is. Een automatische methode om instroom te detecteren kan dus van belang zijn voor verdere automatisering van de beeldgeleiding. In Hoofdstuk 5 ontwikkelden we twee manieren hiervoor: één manier gebruikt de scheiding in verschillende structuren, gevolgd door een vaatdetectie om de vaten duidelijker te maken, en een LSTM algoritme bepaalt dan aan de hand van de gemiddelde intensiteitswaarde van de vaatafbeelding het moment van contrastmiddelinstroom. De tweede manier gebruikt een neuraal netwerk met convoluties voor dezelfde detectie. Het netwerk krijgt als invoer het eerste beeld van de beeldreeks, het huidige beeld, en de drie voorafgaande beelden. Het netwerk geeft vervolgens de waarschijnlijkheid dat het beeld contrastmiddel bevat. Beide methoden waren succesvol in het detecteren van het eerste beeld in een serie waar contrastmiddel aanwezig is. In een experiment met 80 klinische beeldreeks kregen we een gemiddelde fout die kleiner is dan 4 beelden, en een mediane fout van 2 beelden (voor een beeldfrequentie van 15 beelden per seconde). Hiermee zijn de methoden veel beter dan bestaande methoden. Voor wat betreft de rekensnelheid zijn beide methoden snel genoeg om geïntegreerd te kunnen worden in systemen voor beeldgeleiding.

Contrastmiddel wordt tijdens beeldgeleide ingrepen aan de kransslagaders vaak

toegediend om de vaten zichtbaar te maken. Maar voor het navigeren van de instrumenten door de vaten wordt vaak geen gebruik gemaakt van contrastmiddel, m.n. omdat een te grote hoeveelheid contrastmiddel schadelijk is voor de nieren. Daarom moet de cardioloog tijdens het inbrengen en navigeren van de instrumenten gebruik maken van een mentale reconstructie van de vaten. In Hoofdstuk 6 ontwikkelen we een methode om continu de vaten te projecteren op beelden zonder contrastmiddel, om zo de beeldgeleiding voor de cardioloog te verbeteren. In deze methode wordt informatie van de vaten uit eerdere beeldreeksen gebruikt om vaten af te beelden in beelden zonder contrastmiddel. Beweging veroorzaakt door het kloppen van het hart, en door de ademhaling, worden gecompenseerd door gebruik te maken van het ECG, en het volgen van de tip van de katheter. Om de kathetertip nauwkeurig te kunnen volgen, ontwikkelden we een Bayesiaanse filter methode die gebruik maakt van een diep neuraal netwerk. Op deze manier combineren we de probabilistische detectie van de tip met een particle filtering methode om de tip te volgen. De voorgestelde methode heeft een gemiddelde fout van 1.3 mm (op 34 klinische beeldreeksen), en is nauwkeuriger dan methoden die alleen het neurale netwerk, of alleen het Bayesiaanse filter gebruiken. De overprojectie van vaten, die gebruikt maakt van het volgen van de tip, heeft een gemiddelde fout van 2 mm op 409 beelden, waarbij de voerdraad gebruikt is om de nauwkeurigheid te meten. Dit is vergelijkbaar met de fout van overprojectie op basis van het handmatig volgen van de tip (1.8 mm). Voor wat betreft de rekentijd is deze methode van overprojectie (inclusief het volgen van de tip), uitgevoerd op een PC met GPU, snel genoeg om de snelheid van gegenereerde beelden bij te houden. Dit geeft aan dat deze methode mogelijk geïntegreerd kan worden in systemen voor beeldgeleiding voor ingrepen aan de kransslagaders, en dan cardiologen de vaten kan laten zien zonder gebruik van extra contrastmiddel.

Een samenvatting van de analyse van beeldreeksen in dit proefschrift

De methoden voor de analyse van beeldreeksen in dit proefschrift maken gebruik van bewegings- en/of tijdsinformatie, of van meerdere beelden, uit angiografische beeldreeksen. Afhankelijk van hoe deze informatie wordt gebruikt, kunnen de methoden gecategoriseerd worden in twee klasses: een aanpak waarbij de hele beeldreeks in een keer verwerkt wordt, en een aanpak waarbij de beelden een voor een verwerkt worden.

De methoden in Hoofdstuk 2, 4 en de neurale netwerk methode uit Hoofdstuk 5 vallen in de eerste categorie: de volledige beeldreeks wordt gebruikt. Het robuuste PCA algoritme en het PCA algoritme in Hoofdstuk 4 brengen gegevens in de structuur aan direct gebaseerd op de data, terwijl in Hoofdstuk 5 een diep netwerk methode gebruikt wordt om op basis van voorbeeldgegevens te leren. De robuuste PCA methode scheidt de (statische) achtergrond van de (bewegende) voorgrond op basis van de beeldinformatie; de PCA methode gebruikt de eerste eigenvector als de vector die sterk gecorreleerd is aan de ademhalingsbeweging. Om deze taken uit te voeren, is meer dan één beeld nodig. Het neurale netwerk uit Hoofdstuk 5 maakt gebruik van beelden op meerdere tijdspunten, zodat gebruik gemaakt kan worden van verschillen over de tijd. Voor gebruik tijdens beeldgeleide ingrepen zijn dit soort methoden eigenlijk alleen geschikt om achteraf te gebruiken. Voor gebruik tijdens een

ingreep moet dit soort modellen vantevoren getraind worden; tijdens de ingreep wordt dan het getrainde model gebruikt, zoals in het geval van de prospectieve analyse in Hoofdstuk 4, en het neurale netwerk in Hoofdstuk 5.

Methoden die tijdens de ingreep gebruikt worden zullen steeds n additioneel beeld als invoer krijgen. Modellen die hierbij gebruikt worden kunnen gerepresenteerd worden met de formule $x_t = f(x_{t-1})$, waarbij x_t de informatie op tijdstip t is, en f een functie die aangeeft wat de relatie tussen x_{t-1} en x_t is. Deze manier van formuleren maakt deze modellen geschikt voor het verwerken van beeldreeksen tijdens een ingreep. De methoden in Hoofdstuk 3, Hoofdstuk 6 en de methode gebaseerd op RNNs uit Hoofdstuk 5 behoren tot deze categorie. Het OR PCA algoritme in Hoofdstuk 3 splits het beeld in een statische achtergrond en bewegende voorgrond op deze 'on-line' manier, en past het model aan bij elk nieuw beeld dat verwerkt wordt. De methode gebaseerd op een RNN in Hoofdstuk 5 gebruikt een 'terugkerend' neuraal netwerk, waarbij de relatie tussen opeenvolgende beelden gemodelleerd wordt met LSTM eenheden. Het Bayesiaanse filter uit Hoofdstuk 6 is ook zon sequentieel model, waarbij de verandering over de tijd gerepresenteerd wordt met een transitiemodel (of transitie-waarschijnlijkheid) wat de beweging tussen opeenvolgende frames modelleert.

Voor beeldgeleide interventies is de tweede categorie methoden, waarbij de beelden één-voor-één verwerkt worden, uitermate geschikt, omdat het goed aansluit bij de situatie tijdens een interventie. Sommige van deze modellen, zoals de RNN gebaseerde methode uit Hoofdstuk 5 en het bepalen van de tip positie voor het Bayesiaanse filter uit Hoofdstuk 6, zullen eerst getraind moeten worden op eerder verkregen beeldreeksen.

Conclusie en toekomstvisie

Gebaseerd op de methoden ontwikkeld in dit proefschrift, en de vooruitgang in beeldgeleiding en medische beeldvorming in het algemeen over de afgelopen jaren, zullen methode voor beeldgeleiding bij interventies in de toekomst aan aantal trends volgen. Allereerst zullen data van verschillende formaten en herkomst, zoals ECG (1D), röntgenbeelden en angiografie(2D, 2D+t), CTA (3D, 3D+t), geïntegreerd worden. Ten tweede zullen verschillende soorten informatie bij elkaar gebracht worden, om de beeldgeleiding en de besluitvorming daarover te verbeteren. De informatie over anatomie en aderverkalking, die beschikbaar is vanuit CTA beelden, zou bijvoorbeeld geïntegreerd kunnen worden in de beeldgeleiding, wat vooral van belang kan zijn voor ingewikkelde ingrepen zoals de behandeling van chronische totale occlusies. Deze twee aspecten vereisen methoden om beelden van verschillende herkomst bij elkaar te brengen, en gezamenlijk te presenteren, zoals de methoden om vaten uit de beelden te halen en die te projecteren op de doorlichtingsbeelden, zoals beschreven in dit proefschrift. Ten derde zullen data-gedreven methoden, in het bijzonder neurale netwerk gebaseerde methoden, het mogelijk maken te leren van grote hoeveelheden data, wat kan leiden tot betere nauwkeurigheid en robuustheid in veel beeldverwerkingstaken. Deze trends zullen leiden tot een verbetering van de beeldgeleiding: deze zou nauwkeuriger, robuuster, sneller, en ook veiliger kunnen worden.

Concluderend, in dit proefschrift hebben we nieuwe methoden voor de analyse van angiografische/Röntgen beeldreeksen ontwikkeld en geëvalueerd. Deze methoden zijn vooral gericht op het verbeteren van beeldgeleiding bij beeldgeleide ingrepen aan de kransslagaders, en hebben de mogelijkheid om bijv. via overprojectie van de vaten op beelden tijdens de ingreep de beeldgeleiding te verbeteren, en ook mogelijk de hoeveelheid gebruikt contrastmiddel te verminderen. Voor translatie van deze technieken naar de klinische praktijk is verdere integratie in bestaande systemen en verdere validatie in een katheterisatiekamer nodig.

PhD Portfolio

Research

- PhD period: 2014 -2019
- Biomedical Imaging Group Rotterdam (BIGR), Department of Radiology & Nuclear Medicine, Erasmus MC, Rotterdam, The Netherlands
- Advanced School for Computing and Imaging (ASCI), The Netherlands.
- The Cardiovascular Research School Erasmus University Rotterdam (COEUR), Rotterdam, The Netherlands

In-Depth Courses:

- A25 Computer Vision by Learning, ASCI, 2014.
- A22 Knowledge driven Image Segmentation, ASCI, 2014.
- A1 Advanced Pattern Recognition, ASCI, 2015.
- Presentation course, Erasmus MC, 2015.
- Course on R, Erasmus MC, 2015.
- Biomedical English Writing, Erasmus MC, 2016.
- Scientific Integrity, Erasmus MC, 2016.

Summer Schools:

- Medical Imaging Summer School (Medical Imaging meets Computer Vision, MISS 2014), Sicily, Italy, 2014.
- NFBIA Summer School, Nijmegen, The Netherlands, 2015 (oral presentation)
- UCL Medical Image Computing Summer School (MedICSS), London, UK, 2016 (poster).
- International Computer Vision Summer School (ICVSS), Sicily, Italy, 2017.

- Medical Imaging Summer School (Medical Imaging meets Deep Learning, MISS 2018), Sicily, Italy, 2018.

International Conferences (attendance/poster/oral presentation):

- The International Conference on Information Processing in Computer-Assisted Interventions (IPCAI), Barcelona, Spain, 2015 (poster and short oral presentation).
- The MICCAI Workshop on Augmented Environments for Computer-Assisted Interventions (AE-CAI), Munich, Germany, 2015 (poster and short oral presentation).
- The Conference of the International Society for Medical Innovation and Technology (SMIT), Delft, The Netherlands, 2016 (oral presentation).
- The Conference on Neural Information Processing Systems (NIPS), Barcelona, Spain, 2016 (attendance).
- The International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI), Quebec City, Canada, 2017 (poster).
- Design of Medical Devices Conference (DMD Europe 2017), Eindhoven, The Netherlands, 2017 (oral presentation).
- SPIE Medical Imaging (SPIE MI), Houston, United States, 2018 (poster).
- International Conference on Medical Imaging with Deep Learning (MIDL), Amsterdam, The Netherlands, 2018 (attendance).
- Joint MICCAI Workshops on Computing and Visualization for Intravascular Imaging and Computer Assisted Stenting (CVII-STENT), Granada, Spain, 2018 (oral presentation).

Other:

- COEUR PhD Day, Rotterdam, 2016 (oral presentation).
- The Medical Imaging Symposium for PhD students (MISP), The Netherlands, 2014 and 2015 (attendance).
- BIGR seminar presentations (4x), Erasmus MC, 2014-2018.
- Medical Informatics Research Lunch presentations (4x), Erasmus MC, 2014-2018.
- Image Guided Intervention Meetings presentations (18x), Erasmus MC, 2014-2018.
- Attendance: BIGR seminar, Medical Informatics Research Lunch, Image Guided Intervention Meetings, literature meetings, Erasmus MC, 2014-2018.

Publications

Journal Papers

- **H. Ma**, G. Dibildox, C. Schultz, E. Regar and T. van Walsum: PCA-derived Respiratory Motion Surrogates From X-ray Angiograms For Percutaneous Coronary Interventions. *International Journal of Computer Assisted Radiology and Surgery*, vol. 10 (6), pp. 695-705, 2015.
- **H. Ma**, A. Hoogendoorn, E. Regar, W.J. Niessen and T. van Walsum: Automatic Online Layer Separation for Vessel Enhancement in X-ray Angiograms for Percutaneous Coronary Interventions. *Medical Image Analysis*, vol. 39, pp. 145-161, 2017.
- **H. Ma**, I. Smal, J. Daemen and T. van Walsum: Dynamic Coronary Roadmapping via Catheter Tip Tracking in X-ray Fluoroscopy with Deep Learning Based Bayesian Filtering. *Medical Image Analysis*, in press.

Conference Papers

- **H. Ma**, G. Dibildox, J. Banerjee, W.J. Niessen, C. Schultz, E. Regar and T. van Walsum: Layer Separation for Vessel Enhancement in Interventional X-ray Angiograms Using Morphological Filtering and Robust PCA. Workshop on Augmented Environments for Computer-Assisted Interventions (AE-CAI 2015), Lecture Notes in Computer Science, vol. 9365, pp. 104-113, 2015.
- **H. Ma**, P. Ambrosini and T. van Walsum: Fast Prospective Detection of Contrast Inflow in X-ray Angiograms with Convolutional Neural Network and Recurrent Neural Network. The 20th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI), Lecture Notes in Computer Science, vol. 10434, pp. 453-461, 2017.
- H. Hao, **H. Ma** and T. van Walsum: Vessel Layer Separation in X-ray Angiograms with Fully Convolutional Network. Proc. SPIE 10576, Medical Imaging 2018: Image-Guided Procedures, Robotic Interventions, and Modeling, 1057-61V, 2018.
- H. Hao, **H. Ma** and T. van Walsum: Layer Separation in X-ray Angiograms for Vessel Enhancement with Fully Convolutional Network. Intravascular Imaging and Computer Assisted Stenting and Large-Scale Annotation of Biomedical

Data and Expert Label Synthesis (LABELS 2018, CVII 2018, STENT 2018), Lecture Notes in Computer Science, vol 11043, pp. 36-44, 2018.

Conference Abstracts

- **H. Ma**, T. Coradi, G. Szkely, B, Haas, O. Göksel: Supervised Learning with Global Features for Image Retrieval in Atlas-Based Segmentation of Thoracic CT. The 27th International Congress and Exhibition on Computer Assisted Radiology and Surgery (CARS), Heidelberg, Germany, June, 2013.
- **H. Ma**, G. Dibildox, J. Banerjee, W.J. Niessen, C. Schultz, E. Regar and T. van Walsum: Layer Separation for Vessel Enhancement in Interventional X-ray Angiograms Using Morphological Closing and Robust PCA. The 28th Conference of the International Society for Medical Innovation and Technology (SMIT), Delft, The Netherlands, October, 2016.
- **H. Ma**, A. Hoogendoorn, E. Regar, W.J. Niessen and T. van Walsum: Automatic Online Layer Separation for Vessel Enhancement in X-ray Angiograms for Percutaneous Coronary Interventions. Design of Medical Devices Conference (DMD Europe 2017), Eindhoven, The Netherlands, November, 2017.

Acknowledgment

The path that leads to the final destination of the PhD journey is never an easy way. The journey puts us at the frontier of human knowledge and requires us to push the boundary towards the directions that are still unknown. Because of this, walking through the journey most of the time means facing challenges, dealing with uncertainties, loneliness, pressure, and those require wisdom, courage, a strong will and a bit of luck to survive. During my PhD journey, I have always felt lucky to get help, support, inspiration and encouragement from many people who made me no longer walk alone. They are gratefully acknowledged!

I would like to first express my sincere gratitude to my supervisors Prof. Wiro J. Niessen and Dr. Theo van walsum for offering me the opportunity to pursue a PhD in the BIGR group at Erasmus MC. Wiro, I would like to thank you for being a friendly and easy-going supervisor, and inspiring me from a perspective of a large picture on my project. I also enjoyed a lot the conversations we had during the dinners at several MICCAI conferences, and was inspired by the stories you shared from your early academic career. Theo, thank you a lot for being my mentor. You helped me so much on my PhD as well as my life outside work, from providing feedbacks on my presentations and papers to translating Dutch letters for me, and even teaching me how to fix a boiler. You always encouraged me and had full trust on my works. I have learned a lot from you during these years and I have my high regards for you.

I would also like to thank the innter committee of my thesis: Prof. Ton van der Steen, Dr. Jouke Dijkstra and Dr. Evelyn Regar for reviewing my thesis and giving feedbacks. Especially thank Dr. Regar for letting me watch a PCI procedure she conducted in the cathlab when she worked at Erasmus MC. This experience helped me to better understand the procedures in the real clinical environment.

I feel very lucky that I was in departments with many smart and nice colleagues during the PhD time who have helped me and filled my life with lots of fun moments. Pierre, we shared some unforgettable conference experiences and many food events together, thank you for always allowing me to choose the proper restaurants :) Jean-Marie, you helped me so much on my work and my life in NL; I admire your decent way of doing research and I will never forgot the nice trips to Cologne and Budapest. Roman, thank you for showing me your amazing grilling skills at your home and inviting me to your wedding; I wish you and Julia living a wonderful life in Prague! Marco, I enjoy very much the Sinterklaas nights playing Tombola with you; it's good that we are working in the same place again! Gerardo, thank you for helping me a lot on XA data and Mevislab, and telling so many jokes in the office! Nora, thanks for so many

valuable DL discussions and helping me on processing XA and ECG data. Jyotirmoy and Ihor, I always felt inspired by the discussions with you on my works; you both are the walking cyclopedia! Carolyn and Andres, my first summer school experience in Sicily was full of fun because of you. Gennady, you always think the stories I told are funny and I don't know why :) My dear Chinese community in BIGR, Wei, Yao, Chaoping, Yuanyuan, Shuai, Jiahang, Bo, I enjoy many great cooking, eating, chatting, gaming and traveling experiences with you together, you made me feel like home here. I also thank the following people for often enjoying the lunch at the cantine of Erasmus MC together: Gijs, Annegreet, Zahra, Florian, Gerda, Sebastian, Martijn, Karin. I would also like to thank the IGI group members for giving many presentations that taught me the knowledge of various clinical applications: Gerardo, Pierre, Jyotirmoy, Luu, Guillaume, Valerio, Erwin, Marcel T., Reinoud, Roman, Yuanyuan, Jiahang, Luisa, Danilo, Arif, Taygun, Mohamed, Tom, Dirk, Jannis. Thanks also go to the people for nice time spent during other various meetings, outings, discussions, drinks, chattings: Stefan, Dirk, Wyke, Gokhan, Kasper, Akshay, Diego, Emillie, Esther, Marleen, Marcel K., Hakim, Ghassan, Katja, Nawid, Jifke, Henri, Erik, Jose, Antonio, Johnny, Kim, Raghav, Vikram, Riwaj, Arno, Willem, Ewoud, Rebecca. Additionally, I'd like to thank the secretaries, Petra, Desiree, Tineke, Marise, Annelise and the HR staff Andreas for taking care of all the administration works; and thank Mart for the IT support.

My works could not be accomplished without the collaborations with colleagues from other departments at Erasmus MC. I would like to thank people from the BME group: Hans, Alex, Gerard and Raja for valuable discussions during the iMAGIC project UC meetings; and Ayla for providing the image data of pigs. Thank Dr. Joost Daemen from Cardiology for providing data and feedback for my last paper.

In the last few months of my PhD when I wrote my thesis, I have been working at Medis. Hence I would also like to give thanks to my Medis colleagues. I would like to thank Prof. Hans Reiber for offering me the opportunity to work at Medis. Also thank Pieter, Jasper, Viet, Marco, Merih, Evan and Hessam for making a great team to work with in AR. Thank Eelco, Mel and Paulo for arranging the Thai Friday.

I also feel grateful to many friends outside work. Chenming, we've been friends since Switzerland, thank you for the helps and spending many fun time together in these many years. Bob, thank you for several nice hiking trips. Yuan and Ruisheng, thank you for nice foods and chattings together. I would also like to thank my Chinese friends in Leiden for nice gatherings, foods: Yuchuan, Yingguang, Shengnan, Qing, Zhuo, Zhiwei, Ningning, Ling, Xinpei. And Tianfan, thanks for many discussions on CV/ML/DL, you inspired me a lot on my PhD. It was good to meet you again in Amsterdam and NIPS since many years ago. Yihua, thanks for sharing some of your important, fun life moments with me.

Thank my violin teachers Pablo and Alba for taking me into the world of music.

Finally, I want to thank some special people in my life.

Dear Baba, Mama and Nainai, thank you for your unconditional love to me for all the years. You are my reasons of being strong. You are always in my heart.

Dear Qiutan, thank you for sharing so many unforgettable, enjoyable and wonderful moments with me. Thank you for your patience and support. Thank you for encouraging me to go forward. Thank you for making me a better person.

Curriculum Vitae

Hua Ma obtained a Bsc degree in Biological Sciences at Sun Yat-Sen University in 2009 in Guangzhou, China. In 2010, he moved to Switzerland to start his master study at ETH Zurich where he developed interests in image processing for (bio)medical images and machine learning. In 2013, he obtained a Msc degree in Biomedical Engineering with a master thesis on the topic of content-based optimal CT image retrieval for atlas-based segmentation for the treatment planning in radiotherapy, a joint project of the Computer Vision Laboratory at ETH Zurich and Varian Medical Systems Imaging Laboratory GmbH.

In 2014, Hua started his PhD in Biomedical Imaging Group Rotterdam headed by Prof. Wiro Niessen. Hua has been involved in the STW IMAGIC project which aims at developing intelligent image guidance technologies for cardiac interventions, under supervision by Dr.Theo van Walsum. During his PhD time, he focused on techniques in medical imaging, computer vision and machine learning, and used them to develop novel solutions for the challenges in image guidance for percutaneous coronary interventions. His current research interests include methods in computer vision, machine learning and deep learning for medical imaging problems.